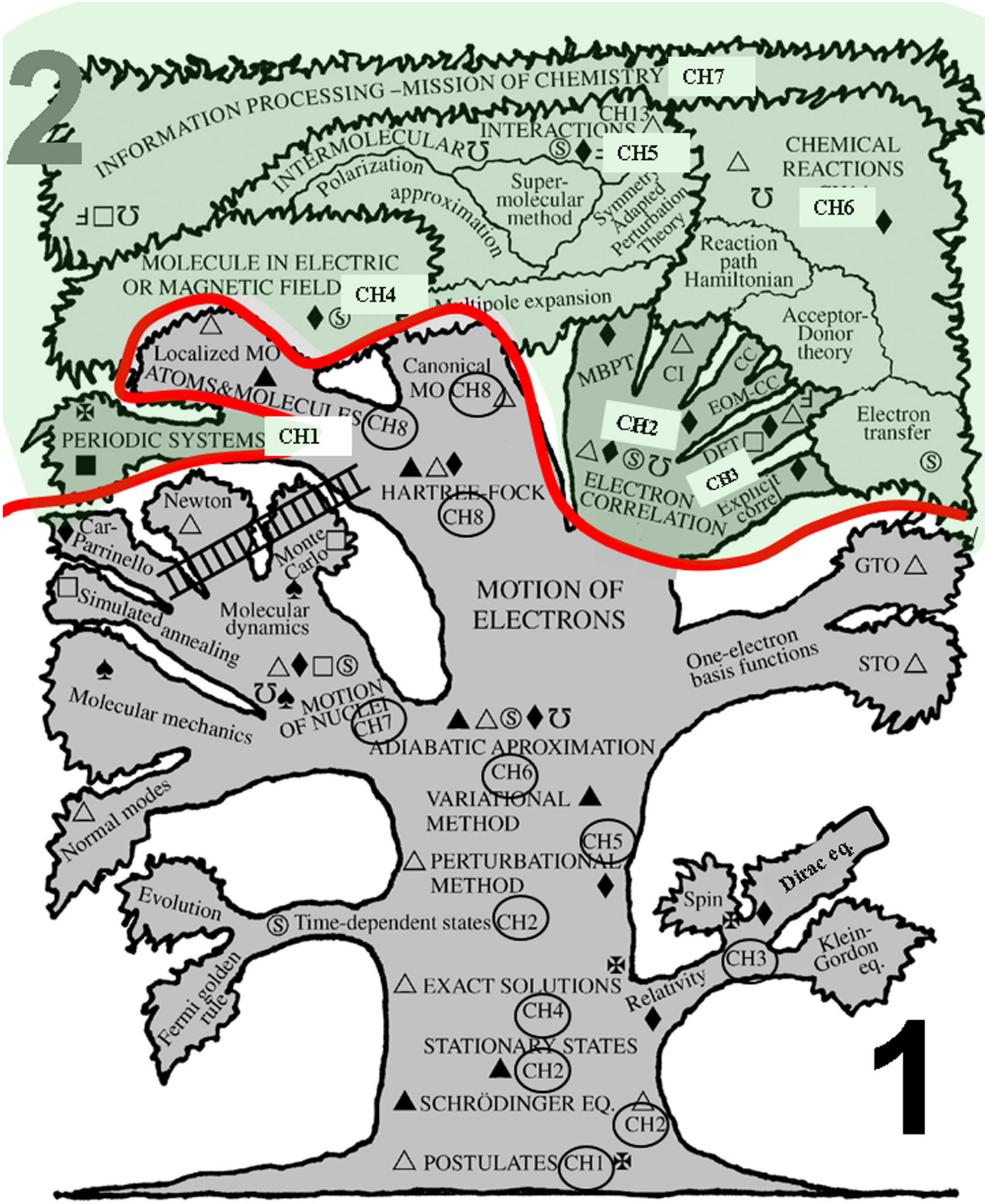
Third Edition

IDEAS OF QUANTUM CHEMISTRY

Volume 1: From Quantum Physics to Chemistry

Lucjan Piela

ELSEVIER


2
INFORMATION PROCESSING –MISSION OF CHEMISTRY
CH7
CH13
INTERMOLECULAR INTERACTIONS
CH5
Polarization approximation
Super-molecular method
Symmetry Adapted Perturbation Theory
CHEMICAL REACTIONS
CH6
Reaction path Hamiltonian
Acceptor-Donor theory
Electron transfer
MOLECULE IN ELECTRIC OR MAGNETIC FIELD
CH4
Multipole expansion
Localized MO
ATOMS&MOLECULES
CH8
Canonical MO
CH8
PERIODIC SYSTEMS
CH1
MBPT
CI
CC
EOM-CC
CH2
DFT
CH3
Explicit correl.
ELECTRON CORRELATION
HARTREE-FOCK
CH8
Newton
Car-Parrinello
Monte Carlo
Simulated annealing
Molecular dynamics
Molecular mechanics
MOTION OF NUCLEI
CH7
MOTION OF ELECTRONS
GTO
STO
One-electron basis functions
ADIABATIC APROXIMATION
CH6
VARIATIONAL METHOD
CH5
PERTURBATIONAL METHOD
Normal modes
Evolution
Time-dependent states
CH2
Fermi golden rule
Spin
Dirac eq.
Klein-Gordon eq.
CH3
Relativity
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, Vol. 1 corresponds to its bottom part). 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 is 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* (►),
- *the most important computational methods of quantum chemistry* (♢),
- *the future of quantum chemistry* (Ⅎ), and
- *"magical" aspects of quantum physics* (✠).

Ideas of Quantum Chemistry

Volume 1: From Quantum Physics to Chemistry

Third Edition

Things appear, ideas persist
Plato

Ideas of Quantum Chemistry

Volume 1: From Quantum Physics to Chemistry

Third Edition

Lucjan Piela

Department of Chemistry
University of Warsaw
Warsaw, Poland

ELSEVIER

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: 978-0-444-64246-2

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

ELSEVIER Book Aid International
Working together
to grow libraries in
developing countries
www.elsevier.com • www.bookaid.org

To all on the quest for the Truth

Contents

Chapter 6: A Key Concept of Molecular 3D Structure – Separation of Electronic and Nuclear Motions *305*

Chapter 8: Orbital Model of Electronic Motion in Atoms and Molecules *457*

RESULTS OF THE HARTREE–FOCK METHOD

Introduction

Quantum scimus gutta est, ignoramus mare
What we know is a drop, what we do not know is a sea
(Latin sentence)

This book (volume 1) is about how to understand the reason for existence of molecules, which the Earth, Nature, and ourselves are composed of.

Reality and its images

Have you ever seen the Milky Way on the night sky? An exceptional and breathtaking experience! You feel yourself looking at a great mystery, as if standing on the shores of the Universe, of something beyond our imagination of space and time. The warmth of the campfire flames and the cold of the night, the Moon, the Milky Way, and the stars – all that offered a wonderful, unique, and puzzling spectacle that, for millennia, challenged our ancestors' imagination and posed the Big Questions: *what do we see* and *who are we*?

The Greek philosopher Plato (427–347 BC) was already aware that looking at the sky is like looking at shadows of some unknown Reality seen on the wall of a cave (lit by a campfire at its entrance), Fig. 0.1. It is our senses that connect us and the "shadows" somehow to this mysterious Reality, which we call the Universe. We feel the Universe's presence, while at the same time we *are part of it* – a fascinating thing by itself. Can we understand what happens around us and in us? Many people, among them Plato, suspected that what we see exhibits a kind of order or regularity, and that maybe we *can* understand it. It is moving for me to feel the spiritual link between the Academia founded by Plato in the sacred piece of land of Academos in Athens and all of us in all universities of the world, who are seeking the truth, among them you, my friend, who are reading these words. Usually after painful work, if we are lucky, in a flash of enlightenment a secret of Nature may be disclosed before our eyes, a great feeling – our prize, the most precious one.

Sensory operations are the direct result of interactions, both between molecules and between light and matter. *All* of these phenomena deal with chemistry, physics, biology, and even psychology. In these complex events it is impossible to discern precisely where the disciplines of

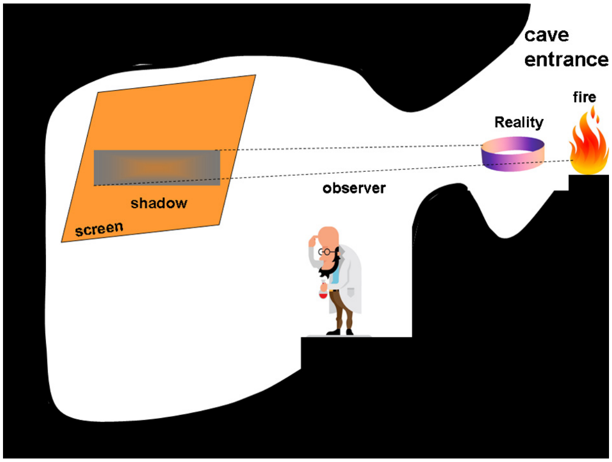


Fig. 0.1. Plato's cave. A caveman is able to see only some shadows on the cave wall. The shadows are produced by an unknown Reality and a fire outside the cave.

chemistry, physics, biology, and psychology begin and end. Any separation of these domains is artificial. The only reason for making such separations is to focus our attention on *some* aspects of one indivisible phenomenon. Sight, hearing, smell, taste, and touch – are these our only links and information channels to the Universe? How little we know about it! To feel that, just look up at the sky. A myriad of stars around us point to new worlds, which will remain unknown forever (because of distance). It is true that by ingenious spectrometry we have serious grounds to believe these stars are built from the same kind of matter we have around us. This bold conclusion was questioned quite recently, since we do not have the slightest idea what kind of particles represent 90% of the matter that does not shine (black matter). This pertains to the macroscale. On the other end, imagine how incredibly complicated the chemistry of (unconditional) maternal love must be, the most beautiful phenomenon in the Universe! Science is a certain response of humans to reduce the Unknown, but it cannot answer all legitimate questions a human being may ask when sitting at a campfire. Science is able to discover laws of Nature, but is unable to answer a question like: *why does our world conform to any laws at all*[1]? Such questions go beyond science.

We try to understand what might be really around us by constructing in our minds a kind of simplified picture, which represents to some extent Reality. It contains seemingly essential elements, being devoid of those elements that we think are irrelevant. These pictures we call models. Any model relies on the one hand on our perception of reality (on the appropriate scale of masses and time) emanating from our experience, and on the other hand on our ability to

---

[1] "*The most incomprehensible thing about the world is that it is at all comprehensible*" (Albert Einstein).

abstract by creating ideal beings. These ideal beings seem to be close to the concept of ideas ("forms") that Plato loved most. Many such models will be described in this book.

It is fascinating that man is able to magnify the realm of his senses by using sophisticated tools, e.g., to see quarks sitting in a proton[2] or to discover an amazingly simple equation of motion[3] that describes both cosmic catastrophes, with intensity beyond our imagination, and the flight of a butterfly. A water molecule has exactly the same properties in the Pacific Ocean as on Mars or in another galaxy. The conditions over there may be quite different from those in our laboratory, but we *assume* that if these conditions could be imposed in the lab, the molecule would behave in exactly the same way. We hold out hope that a set of universal physical laws apply for the entire Universe.

The model of these basic laws is not yet complete or unified. Thanks to the progress and important generalizations of physics, much is currently understood. For example, forces with seemingly disparate sources have been reduced to only three kinds:

- those attributed to *strong interactions* (acting in nuclear matter),
- those attributed to *electroweak interactions* (the domain of chemistry, biology, as well as β-decay),
- those attributed to *gravitational interactions* (showing up mainly in astrophysics).

Many scientists believe other reductions are possible, perhaps up to a single fundamental interaction, one that explains Everything (quoting Feynman: "the frogs as well as the composers"). This assertion is based on the conviction, which seems to be supported by developments in modern physics, that the laws of Nature are not only universal, but also simple.

Which of the three basic interactions is the most important? This is an ill-conceived question. The answer depends on the external conditions imposed (pressure, temperature) and the magnitude of the energy exchanged amongst the interacting objects. A measure of the energy exchanged (ΔE) may be taken to be the percentage of the accompanying mass deficiency (Δm) according to Einstein's relation $\Delta E = \Delta mc^2$. At a given magnitude of exchanged energies some particles are stable. Strong interactions produce the huge pressures that accompany the gravitational collapse of a star and lead to the formation of neutron stars, where the mass deficiency Δm approaches 40%. At smaller pressures, where individual nuclei may exist and

---

[2] A proton is 10^{15} times smaller than a human being and nevertheless in 1970 Jerome Friedman, Henry Kendall, and Richard Taylor were able to take a proton's photograph. They have shown us three quarks and unknown electrically neutral matter that binds the quarks together ("gluons")!

[3] Acceleration is directly *proportional* to force. Higher derivatives of the trajectory with respect to time do not enter this equation, neither does the nature or cause of the force. The equation is also invariant with respect to any possible starting point (position, velocity, and mass). What a remarkable simplicity and generality (within limits; see Chapter 3)!

undergo nuclear reactions (strong interactions[4]), the mass deficiency is of the order of 1%. At much lower pressures the electroweak forces dominate, nuclei are stable, and atomic and molecular structures emerge. Life (as we know it) becomes possible. The energies exchanged are much smaller and correspond to a mass deficiency of the order of only about $10^{-7}\%$. The weakest of the basic forces is gravitation. Paradoxically, this force is the most important on the macroscale (galaxies, stars, planets, etc.). There are two reasons for this. Gravitational interactions share with electric interactions the longest range known (both decay as $1/r$). However, unlike electric interactions[5] those due to gravitation are not shielded. For this reason the Earth and the Moon attract each other by a huge gravitational force[6] while their electric interaction is negligible. This is how David conquers Goliath, since at *any distance* electrons and protons attract each other by electrostatic forces, about 40 orders of magnitude stronger than their gravitational attraction.

Gravitation does not have any measurable influence on the collisions of molecules leading to chemical reactions, since reactions are due to much stronger electric interactions.[7]

Ten degrees only

Due to strong interactions, protons overcome mutual electrostatic repulsion and form (together with neutrons) stable nuclei, leading to the variety of chemical elements. Therefore, strong interactions are the prerequisite of any chemistry (except hydrogen chemistry). However, chemists deal with already prepared stable nuclei[8] and these strong interactions have a very small range (of about 10^{-13} cm) as compared to inter-atomic distances (of the order of 10^{-8} cm). *This is why a chemist may treat nuclei as stable point charges that create an electrostatic field.* Test tube conditions allow for the presence of electrons and photons, thus completing the set of particles that one might expect to see (some exceptions are covered in this book). This has to do with the order of magnitude of energies exchanged; under the conditions of our chemical reactions, the energies exchanged exclude practically all nuclear reactions.

---

[4] With a corresponding large energy output; the energy coming from the fusion $D + D \to He$ taking place on the Sun makes our existence possible.

[5] In electrostatic interactions charges of opposite sign attract each other while charges of the same sign repel each other (Coulomb's law). This results in the fact that large bodies (built of a huge number of charged particles) are nearly electrically *neutral* and interact electrically only very weakly. This dramatically reduces the range of their electrical interactions.

[6] Huge tides and deformations of the whole Earth are witness to that.

[7] It does not mean that gravitation has no influence on reactants' concentrations. Gravitation controls the convection flow in liquids and gases (and even solids) and therefore a chemical reaction or even crystallization may proceed in a different manner on the Earth's surface, in the stratosphere, in a centrifuge, or in space.

[8] At least on the time scale of chemical experiments. Instability of some nuclei is used by nuclear chemistry and radiation chemistry.

On the vast scale of attainable temperatures[9] chemical structures may exist in the narrow temperature range of 0 K to thousands of K. Above this range one has plasma, which represents a soup made of electrons and nuclei. Nature, in its vibrant living form, requires a temperature range of about 200–320 K, a margin of only 120 K. One does not require a chemist for chemical structures to exist. However, to develop a chemical science one has to have a chemist. This chemist can survive a temperature range of 273 ± 50 K, i.e., a range of only 100 K. The reader has to admit that a chemist may think of his job only in the narrow range of 290–300 K, only 10 K.

Grand unification and mission of chemistry

Suppose our dream comes true and the grand unification of the three remaining basic forces is accomplished one day. We would then know the first principles of constructing Everything. One of the consequences of such a feat is a catalog of all the elementary particles; maybe the catalog will be finite,[10] hopefully it will be simple. We might have a catalog of the conserved symmetries (which seem to be more elementary than the particles). Of course, knowing such first principles would have an enormous conceptual impact on all the physical sciences. It could create an impression that everything is clear, because science is complete. Even though such structures and processes are governed by first principles, it would still be very difficult to predict their existence by such principles alone. The resulting structures would depend not only on the principles, but also on the initial conditions, complexity, self-organization, etc.[11] *Therefore, if it does happen, the Grand Unification will not change the goals of chemistry.*

Organization of the book

TREE

Any book has a linear appearance, i.e., the text goes page after page 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.

---

[9] Millions of degrees.

[10] None of this is certain. Much of elementary particle research relies on large particle accelerators. This research resembles discerning the components of a car by dropping it from increasing heights from a large building. Dropping it from the first floor yields five tires and a jack. Dropping from the second floor reveals an engine and 11 screws of similar appearance. Eventually a problem emerges: after landing from a very high floor new components appear (having nothing to do with the car) and reveal that some of the collision energy has been converted to the new particles!

[11] The fact that Uncle John likes to drink coffee with cream at 5 p.m. possibly follows from the first principles, but it would be very difficult to trace that dependence.

Such a diagram allows for multiple branches emanating from a given chapter, particularly if the branches are placed logically on an equal footing. Such logical connections are illustrated in this book as a TREE diagram (beginning of the book). This TREE diagram plays a very important role in our book and is intended to be a study guide. It is used to lead the reader in a certain direction; from the TREE diagram, the reader can observe what this direction is, why he/she needs this direction, what will follow, and what benefits he/she will gain after such study. If studying were easy and did not require time, a TREE diagram might be of little importance. However, the opposite is usually true. In addition, knowledge represents much more than a registry of facts. Any understanding gained from seeing relationships among those facts and methods plays a key role.[12] The primary function of the TREE diagram is to make these relationships clear.

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

The use of hypertext in information science is superior to a traditional linear presentation. It relies on a tree structure. However, it has a serious drawback. Sitting on a branch, we have no idea what that branch represents in the whole diagram, whether it is an important branch or a remote tiny one, whether it leads further to important parts of the book or whether it is just a dead end, and so on. At the same time, a glimpse at the TREE diagram shows us that the thick trunk is the most important structure. What do we mean by important? At least two criteria may be used. Important for the majority of *readers*, or important because the material is fundamental for an understanding of the *laws of Nature*. I have chosen the first.[13] Thus, the trunk of the TREE diagram corresponds to the pragmatic way to study this book.

The trunk is the backbone of this book.

- It begins by presenting postulates, which play a vital role in formulating the foundation of quantum mechanics.
- Next, 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 or Born–Oppenheimer approximation, *the central idea of the present book and chemistry in general*).

---

[12] This advice comes from antiquity: "*knowledge is more precious than facts, understanding is more precious than knowledge, wisdom is more precious than understanding*."

[13] For example, relativity theory plays a pivotal role as a foundation of the physical sciences, but for the vast majority of chemists its practical importance and impact are much smaller. Should relativity be represented therefore as the base of the trunk, or as a minor branch? Contemporary inorganic chemistry and metallo-organic chemistry concentrate currently on heavy elements, where relativity effects are important. We have decided to make the second choice, to *not* create the impression that this topic is absolutely necessary for the student.

- It then develops the mean-field theory of electronic structure, and
- finally, it develops and describes methods that take into account electronic correlation.

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 diagram (appropriate for graduate students). In particular, it makes it possible to reach the crown of the TREE, where the reader may find tasty fruit. Examples include the theory of molecule–electric field interactions, as well as the theory of intermolecular interactions (including chemical reactions), which form the very essence of chemistry. We also see that our TREE diagram has an important branch concerned with nuclear motion, including molecular mechanics and several variants of molecular dynamics. At its base, the trunk has two thin branches: one pertains to relativity mechanics and the other to the time-dependent Schrödinger equation. The motivation for this presentation is different in each case. I do not highlight relativity theory for the reasons already explained. The time-dependent Schrödinger equation is not highlighted, because, for the time being, quantum chemistry accentuates stationary states. I am confident, however, that the 21st century will see significant developments in the methods designed for time-dependent phenomena.

The TREE helps tailoring your own book

The TREE not only serves 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.[14] This means tailoring the reader's own book.

Of course, all readers are welcome to find 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 the book 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 (▲). I imagine readers studying material science, biology, biochemistry, or a similar subject. They have heard that quantum chemistry explains chemistry, and want to get the flavor and grasp the most important information. They should read only 47 pages.

---

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

- *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* △. 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 through Volumes 1 and 2

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* (■),
- *chemical reactions* (℧),
- *spectroscopy* (Ⓢ),
- *exact calculations* on atoms or small molecules[15] (⧫),
- *relativistic and quantum electrodynamics effects* (▶), and
- *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.

- Just before an exam, read in each chapter the sections "*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, the sections labeled "*Ad futurum*" in each chapter, and the chapters designated by (⅃) are proposed.

---

[15] Suppose the reader is interested in an accurate theoretical description of small molecules. I imagine such a PhD student working in quantum chemistry. Following their itinerary, they have, in addition to the minimum program (300 pages), an additional 230 pages, which gives about 530 pages plus the appropriate appendices, in total about 700 pages (in Vol. 1 and Vol. 2).

- For people interested in the "magical" aspects of quantum physics (e.g., bilocation, reality of the world, teleportation, creation of matter, tunneling) we suggest sections with the label (✠).

The target audience

I hope that the TREE structure presented above will be useful for those with varying levels of knowledge in quantum chemistry, as well as for those whose goals and interests differ from traditional quantum chemistry.

This book is a direct result of my lectures at the Department of Chemistry, University of Warsaw, for students specializing in theoretical rather than experimental chemistry. Is this the target audience of this book? Yes, but not exclusively. In the beginning I assumed that the reader would have completed a basic quantum chemistry course[16] and, therefore, in the first version the basic material was omitted. However, that version became inconsistent, devoid of several fundamental problems. This is why I decided to explain, mainly very briefly,[17] these problems too. Therefore, a student who chooses the *minimum* path along the TREE diagram (mainly along the TREE trunk) will obtain an introductory course in quantum chemistry. However, the complete collection of chapters provides the student with a set of advanced topics in quantum chemistry, appropriate for graduate students. For example, a number of chapters, such as those on relativity mechanics, global molecular mechanics, solid state physics and chemistry, electron correlation, density function theory, intermolecular interactions, and theory of chemical reactions, present material that is usually accessible in monographs or review articles.

My goal – ideas of quantum chemistry

In writing this book I imagined students sitting in front of me. In discussions with students I often see their enthusiasm, their eyes showing me a glimpse of curiosity. First of all, this book is an acknowledgment of my young friends, my students, and an expression of the joy of being with them. Work with them formulated and influenced the way in which I decided to write this book. When reading textbooks one often has the impression that all the outstanding problems in a particular field have been solved, that everything is complete and clear, and that the student is just supposed to learn and absorb the material at hand. In science the opposite is true. All areas can benefit from careful probing and investigation. Your insight, your different perspective or point of view, even on a fundamental question, may open new doors for others.

---

[16] Say, at the level of P.W. Atkins, "*Physical Chemistry*," sixth edition, Oxford University Press, Oxford, 1998, Chapters 11–14.

[17] Except where I wanted to stress some particular topics.

Fostering this kind of new insight is one of my main goals. I have tried, whenever possible, to present the reasoning behind a particular method and to avoid rote citation of discoveries. I have tried to avoid details, because I know how difficult it is then for a new student to see the forest through the trees. I want to focus on the main ideas of quantum chemistry.

I have tried to stress this integral point of view, and this is why the book sometimes deviates from what is normally considered as quantum chemistry. I sacrificed, not only in full consciousness, but also voluntarily, "quantum cleanness" in favor of exposing the interrelationships of problems. In this respect, any division between physics and chemistry, organic chemistry and quantum chemistry, quantum chemistry for chemists and quantum chemistry for biologists, and intermolecular interactions for chemists, physicists, or biologists is completely artificial, and sometimes even absurd.[18] I have tried to cross these borders by supplying examples and comparisons from the various disciplines, even from everyday life, by incorporating into intermolecular interactions not only supramolecular chemistry, but also molecular computers, and particularly, by writing a "holistic" (last) chapter about the mission of chemistry.

My experience tells me that the new talented student who loves mathematics courts danger. They like complex derivations of formulae so much that it seems that the more complex the formalism, the happier the student. However, all these formulae represent no more than an approximation, and sometimes it would be better to have a simple formula. The simple formula, even if less accurate, may tell us more and bring more understanding than a very complicated one. Behind complex formulae usually hide some very simple concepts, e.g., that two molecules are unhappy when occupying the same space, or that in a tedious iteration process we approach the final ideal wave function in a way similar to a sculptor shaping his masterpiece. All the time, in everyday life, we unconsciously use these variational and perturbational methods – the most important tools in quantum chemistry. This book may be considered by some students as "too easy." However, I prize easy explanations very highly. In later years the student will not remember long derivations, but will know exactly why something *must* happen. Also, when deriving formulae, I try to avoid presenting the final result right away, but instead proceed with the derivation step by step.[19] The reason is psychological. Students have a much stronger motivation knowing they *control* everything, even by simply *accepting* every step of derivation. It gives them a kind of psychological integrity, very important in any study. Some formulae may be judged as right just by inspection. This is especially valuable for students and I always try to stress this.

In the course of study, students should master material that is both simple and complex. Much of this involves familiarity with the set of mathematical tools repeatedly used throughout this

---

[18] The above described itineraries cross these borders.

[19] Sometimes this is not possible. Some formulae require painstaking effort to be derived. This was the case, for example, in the coupled cluster method on p. V2-157.

book. The appendices provide ample reference to such a toolbox. These include matrix algebra, determinants, vector spaces, vector orthogonalization, secular equations, matrix diagonalization, point group theory, delta function, finding conditional extrema (Lagrange multipliers, penalty function methods), and Slater–Condon rules, as well as second quantization. I would suggest that the reader review (before reading this book) the elementary introduction to matrix algebra (Appendix A) and to vector spaces and operators (Appendix B). These appendices are often used throughout this book.

The book contains numerical examples in many places. Their goal is always a semiquantitative description of a *phenomenon*, not so much the description of this particular system. This is because I prefer to get a trend of changes, also an order of magnitude of the things to be illustrated rather than highly accurate numbers. My private conviction behind this approach is quite strange and unusual: Nature is so rich (think of all elements as possible substitutions, influence of neighboring atoms that could modify the properties, using pressure, etc.), that there is a good probability of finding a system exhibiting the phenomenon we got in our calculations... Well, at least we hope there is.

One more thing: Writing this book, I imagined students sitting in a lecture hall. The tone of this book should bring to mind a lecture, in interactive mode. To some, this is not the way books are supposed to be written. I apologize to any readers who may not feel comfortable with this approach.

I invite cordially all readers to share with me their comments (piela@chem.uw.edu.pl).

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. This web page offers the most powerful professional computer programs. The programs calculate millions of integrals needed, but the reader does not see them. Using this tool is straightforward and instructive. I propose the reader to check this immediately.

Web annex

The role of the annex is to expand the readers' knowledge after they read a given chapter. At the heart of the web annex are the links to other people's websites. The annex adds at least four new dimensions to the book: color, motion, an interactive mode of learning, and connection to the web (with a plethora of possibilities to go further and further). When on the web, the reader may choose to come back (automatically) to the annex at any time.

How to begin

It is suggested that the reader start with the following.

- Study the TREE diagram.
- Read the table of contents and compare it with the TREE.
- Ask yourself what *your* goal is, i.e., why would you like to read such a book?
- Choose a personal path on the TREE; the suggested itineraries may be of some help.[20]
- Become acquainted with the organization of any chapter.

Chapter organization

Once an itinerary is chosen the student will cover definite chapters. All the chapters have the same structure, and are divided into the following sections.

- **Where are we**
  In this section the reader is made aware of his/her current position on the TREE diagram. In this way, they know the relationship of the current chapter to other chapters, what chapters they are expected to have covered already, and the remaining chapters for which the current chapter provides a preparation. The position shows *whether he/she should invest time and effort in studying the current chapter*. In this section a mini-TREE is also shown indicating the current position.
- **An example**
  Here the reader is confronted with a practical problem that the current chapter addresses.
- **What is it all about**
  In this section the essence of the chapter is presented and a detailed exposition follows. The recommended paths are also provided.
- **Why is this important**
  Not all chapters are of equal importance for the reader. At this point, the reader has the opportunity to judge whether the arguments about the importance of a current chapter are convincing.

---

[20] This choice may still be tentative and may become clear in the course of reading this book. The index at the end may serve as a significant help. For example, a reader interested in drug design, which is based in particular on enzymatic receptors, should cover the chapters with ▲ (those considered most important) and then those with △ (at the very least, intermolecular interactions). He/she will gain the requisite familiarity with the energy which is minimized in computer programs. The reader should then proceed to those branches of the TREE diagram labeled with □ (large molecules). Initially the reader may be interested in force fields (where the abovementioned energy is approximated), and then in molecular mechanics and molecular dynamics (♠). It is true that our students might begin this course with only the ♠ labels. However, such a course would leave them without any link to quantum mechanics.

- **What is needed**
  This section lists the prerequisites necessary for the successful completion of the current chapter. Material required for understanding the text is provided in the appendices at the end of this book. The reader is asked not to take this section too literally, since a tool may be needed only for a minor part of the material covered, and would therefore be of secondary importance.
- **Classical works**
  Every field of science has a founding parent. They identified the seminal problems, introduced basic ideas, and selected the necessary tools. Wherever appropriate, these classical investigators and their most important contributions are mentioned.
- **The chapter's body**
  The main body of each chapter is presented in this section.
- **Summary**
  The main body of a chapter is still a big thing to digest and a student may be lost in the logical structure of each chapter.[21] A short summary assures the student of the motivation for presenting the material at hand and explains why one should expend the effort and pain, what the main benefits are, and why I have attached importance to this subject. This is a useful point for reflection and consideration. What have we learned, where are we heading, and where will this knowledge be used and applied?
- **Main concepts, new terms**
  New terms, definitions, concepts, and relationships that have been introduced in the chapter are listed here.
- **From the research front**
  It is often ill-advised to present state-of-the-art results to students. For example, of what value is it to present a wave function consisting of thousands of terms for the helium atom? The logistics of such a presentation are difficult to contemplate. There is significant didactic value in presenting a wave function with only one or a few terms where significant concepts are communicated. However, the student should be made aware of recent progress in generating new results and how well they agree with experimental observations.
- **Ad futurum**
  What is the prognosis for future developments in this area according to the author? These are often perplexing questions and the reader deserves an honest answer.
- **Additional literature**
  The present text offers only a general panorama of quantum chemistry. In most cases there exists an extensive body of literature, where the reader will find more detailed information.

---

[21] This is most dangerous. A student at *any* stage of study has to be able to answer easily what the purpose of each stage is.

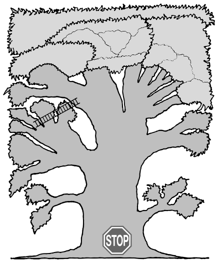
- **Questions**
  In this section the reader will find ten topics, each containing four yes-or-no questions related to the current chapter. Sometimes the choice will come down to the truth or an absurdity, and the answer will come easily.
- **Answers**
  Here the answers to the above questions are provided.

Acknowledgments

Only a unique combination of friendship, love of science, high competence, spiritual strength, extreme systematicity, and perseverance led my friend Professor Andrzej Sadlej (1941–2010) from the Nicolaus Copernicus University in Toruń to perform an unusual job: commenting on my book, from comment 1 to comment 553 on p. 207 of his notes, despite his terminal sickness, always smiling. I would like to thank also another friend of mine, Professor Leszek Z. Stolarczyk from the University of Warsaw, for daily stimulating discussions helping both of us to understand what we see around. Many thanks go to Professor Jacek Klinowski (University of Cambridge, UK) and his wife Dr. Margaret Klinowska for reading, commenting, and supporting.

I am very indebted to my friends – quantum chemists mainly from the University of Warsaw, but also from many other universities from all over the world for their invaluable expertise and helpful reviews and comments. Only with this help the present book may appear. I appreciate very much the spiritual support and freedom I always felt from the side of my university and my country.

Many thanks go to my wonderful wife Basia for her understanding and love.



CHAPTER 1

The Magic of Quantum Mechanics

> *Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.*
>
> **Albert Einstein**

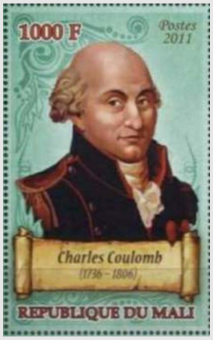
Where are we?

We are at the beginning of all the paths, at the base of the TREE.

An example

Since 1911 we have known that atoms and molecules are built of two kinds of particles: electrons and nuclei. Experiments show the particles may be treated as point-like objects of certain mass and electric charge. The electronic charge is equal to $-e$, while the nuclear charge amounts to Ze, where $e = 1.6 \cdot 10^{-19}$ C and Z is a natural number.[1]

Charles Augustin de Coulomb (1736–1806), French military engineer, one of the founders of quantitative physics. In 1777 he constructed a torsion balance for measuring very weak forces, with which he was able to demonstrate the inverse square law for electric and magnetic forces.
He also studied charge distribution on the surfaces of dielectrics.



---

[1] In quantum chemistry and, therefore, also in the present book these facts will be taken for granted. However, they are intriguing by themselves. Just a few questions: why is the electric charge quantized (in e units)? Why do all electrons have the same charge and mass? Why is the experimental $|e|$ the same for electrons and protons (within the accuracy of 10^{-20})? Up to now science is unable to answer these questions (there are some important attempts; see, e.g., A. Staruszkiewicz, *Acta Phys. Polon.*, 33(2002)2041).

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00009-4

Electrons and nuclei interact according to the Coulomb law, and classical mechanics and electrodynamics predict that *any atom or molecule is bound to collapse in a matter of a femtosecond*, emitting an infinite amount of energy. Hence, according to the classical laws, the complex matter we see around us should simply not exist at all.

However, atoms and molecules do exist, and their existence may be described in detail by quantum mechanics using what is known as the wave function. The postulates of quantum mechanics provide the rules for finding this function and for the calculation of all the observable properties of atoms and molecules. These calculations agree extremely well with experiments.

What is it all about?

Any branch of science has a list of postulates, on which the entire construction is built.[2] For quantum mechanics, six such postulates have been established. The postulates have evolved in the process of reconciling theory and experiment, and may sometimes be viewed as nonintuitive. They stand behind any tool of quantum mechanics used in practical applications. They also lead to some striking conclusions

---

[2] And which are not expected to be proved.

concerning the reality of our world, for example, the possibilities of bilocation and teleportation. These unexpected conclusions have recently been experimentally confirmed.

Why is this important?

The postulates given in this chapter represent the *foundation* of quantum mechanics, and justify all that follows in this book. In addition, our ideas of what the world is really like acquire a new and unexpected dimension.

What is needed?

- complex numbers (necessary),
- operator algebra and vector spaces, p. 595 (necessary),
- angular momentum, p. 665 (necessary),
- some background in experimental physics: black body radiation, photoelectric effect (recommended).

Classical works

The beginning of quantum theory was the discovery, by Max Planck, of the electromagnetic energy quanta *emitted* by a black body. The work was published under the title "*Über das Gesetz der Energieverteilung im Normalspektrum*,"[3] in *Annalen der Physik,* 4(1901)553. ★ Four years later Albert Einstein published the paper "*Über die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*" in *Annalen der Physik,* 17(1905)132, in which he explained the photoelectric effect by assuming that the energy is *absorbed* by a metal as quanta of energy. ★ In 1911 Ernest Rutherford discovered that atoms are composed of a massive nucleus and electrons in "*The Scattering of the* α *and* β *Rays and the Structure of the Atom*," in *Proceedings of the Manchester Literary and Philosophical Society*, IV, 55(1911)18. ★ Two years later Niels Bohr introduced a planetary model of the hydrogen atom in "*On the Constitution of Atoms and Molecules*," in *Philosophical Magazine*, Series 6, vol. 26(1913). ★ Louis de Broglie generalized the corpuscular and wave character of any particle in his PhD thesis "*Recherches sur la théorie des quanta*," Sorbonne, 1924. ★ The first mathematical formulation of quantum mechanics was developed by Werner Heisenberg in "*Über quantentheoretischen*

---

[3] Or "*On the Energy Distribution Law in the Normal Spectrum*" with a note saying that the material had already been presented (in another form) at the meetings of the German Physical Society on October 19 and December 14, 1900.

On p. 556 one can find the following historical sentence on the total energy denoted as U_N: "*Hierzu ist es notwendig,* U_N *nicht als eine stetige, unbeschränkt teilbare, sondern als eine diskrete, aus einer ganzen Zahl von endlichen gleichen Teilen zusammengesetzte Grösse aufzufassen,*" which translates as: "*Therefore, it is necessary to assume that* U_N *does not represent any continuous quantity that can be divided without any restriction. Instead, one has to understand that it is a discrete quantity composed of a finite number of equal parts.*"

Umdeutung kinematischer und mechanischer Beziehungen," in *Zeitschrift für Physik,* 33(1925)879. ★ Max Born and Pascual Jordan recognized matrix algebra in the formulation in "*Zur Quantenmechanik*," in *Zeitschrift für Physik*, 34(1925)858, and then all three expounded a coherent mathematical basis for quantum mechanics in the famous "*Dreimännerarbeit*" entitled "*Zur Quantenmechanik. II*," published in *Zeitschrift für Physik*, 35(1925)557. ★ Wolfgang Pauli introduced his "two-valuedness" for the nonclassical electron coordinate in "*Über den Einfluss der Geschwindigkeitsabhängigkeit der Elektronenmasse auf den Zeemaneffekt*," published in *Zeitschrift für Physik*, 31(1925)373. The next year George Uhlenbeck and Samuel Goudsmit described their concept of particle spin in "*Spinning Electrons and the Structure of Spectra*," *Nature*, 117(1926)264. ★ Wolfgang Pauli published his famous exclusion principle in "*Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*," which appeared in *Zeitschrift für Physik B*, 31(1925)765. ★ The series of papers by Erwin Schrödinger "*Quantisierung als Eigenwertproblem*", in *Annalen der Physik*, 79(1926)361 (other references in Chapter 2) was a major advance. He proposed a different mathematical formulation (from Heisenberg's) and introduced the notion of the wave function. ★ In the same year Max Born, in "*Quantenmechanik der Stossvorgänge*," which appeared in *Zeitschrift für Physik*, 37(1926)863, gave an interpretation of the wave function. ★ The uncertainty principle for position and momentum of a particle was discovered by Werner Heisenberg and described in "*Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*," *Zeitschrift für Physik*, 43(1927)172.★ A similar uncertainty between energy and time has been proved in 1945 by the Russian physicists Leonid I. Mandelshtam and Igor E. Tamm in "*The uncertainty relation between energy and time in nonrelativistic quantum mechanics*," *J. Phys. (USSR)* 9(1945)249. ★ Paul Adrien Maurice Dirac reported an attempt to reconcile quantum and relativity theories in a series of papers from 1926–1928 (references in Chapter 3). ★ Albert Einstein, Boris Podolsky, and Natan Rosen proposed a (then a *Gedankenexperiment* or thought experiment, now a real one) test of quantum mechanics in "*Can quantum-mechanical description of physical reality be considered complete?*," published in *Physical Review*, 47(1935)777. ★ Richard Feynman, Julian Schwinger, and Shinichiro Tomonaga (independently) developed quantum electrodynamics in 1948 (description of the electromagnetic forces), which in 1973 has been supplemented by David Gross, Frank Wilczek, and David Politzer by creating the theory of the strong interactions, what is known today as chromodynamics ★ John Bell, in "*On the Einstein–Podolsky–Rosen Paradox*," *Physics*, 1(1964)195 reported inequalities which were able to verify the very foundations of quantum mechanics. ★ Alain Aspect, Jean Dalibard, and Gérard Roger in "*Experimental Test of Bell's Inequalities Using Time-Varying Analyzers*," *Physical Review Letters*, 49(1982)1804 reported measurements which violated the Bell inequality and proved the nonlocality and/or (in a sense) nonreality of our world. ★ The first two-slit interference experiments proving the wave nature of electrons have been performed in 1961 by Claus Jönsson from Tübingen Universität in Germany and published in "*Elektroneninterferenzen an mehreren künstlich hergestellter Feinspalten*," in *Zeitschrift für Physik*, 161(1961)454], while the experimental proof for interference of a single electron has been presented by Pier Giorgio Merli, Gianfranco Missiroli, and Gulio Pozzi from University of Milan (Italy) in the article "*On the Statistical Aspect of electron interference phenomena*," *American Journal of Physics*, 44(1976)306. ★ Charles H. Bennett, Gilles Brassart, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters, in "*Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*," in *Physical Review Letters*, 70(1993)1895, designed a teleportation experiment, which has subsequently been successfully accomplished by Dik Bouwmeester, Jan-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, and Anton Zeilinger, in “*Experimental Quantum Teleportation*,” in *Nature*, 390(1997)575.

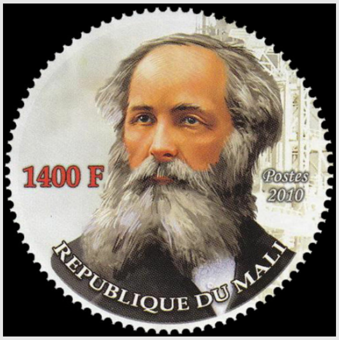
* * *

1.1 History of a revolution

The end of the 19th century saw itself as a proud period for physics, which seemed to finally achieve a state of coherence and clarity. Physics at that time believed the world consisted of two kingdoms: a kingdom of particles and a kingdom of electromagnetic waves. Motion of particles had been described by Isaac Newton’s equation, with its striking simplicity, universality, and beauty. Similarly, electromagnetic waves had been accurately described by James Clerk Maxwell’s simple and beautiful equations.

Young Planck was advised to abandon the idea of studying physics, because everything had already been discovered. As Leon Lederman wrote[4]: “*The physics was elegantly packed in a box and tied with a bow*.” This beautiful idyll was only slightly incomplete, because of a few annoying details: the strange black body radiation, the photoelectric effect, and the mysterious atomic spectra. Just some rather peripheral problems to be fixed in the near future by the omnipotent Newton–Maxwell machine...

James Clerk Maxwell (1831–1879), British physicist, professor at the University of Aberdeen, Kings College, London, and Cavendish Professor in Cambridge. His main contributions are famous equations for electromagnetism (1864) and the earlier discovery of velocity distribution in gases (1860).



As it turned out, they opened a New World. The history of quantum theory, one of most revolutionary and successful theories ever designed by man, will briefly be given below. Many of these facts have their continuation in the present textbook.

---

[4] L. Lederman and D. Teresi, “*The God Particle*,” Dell Publishing, 1993.

1900 – Max Planck

Max Karl Ernst Ludwig Planck (1858–1947), German physicist, professor at the universities in Munich, Kiel, and Berlin, first director of the Institute of Theoretical Physics in Berlin. Planck was born in Kiel, where his father was a university professor of law. Max Planck was a universally talented school pupil, then an outstanding physics student at the University of Berlin, where he was supervised by Gustaw Kirchhoff and Hermann Helmholtz. Music was his passion throughout his life, and he used to play piano duets with Einstein (who played the violin). This hard-working, middle-aged, old-fashioned professor of thermodynamics made a major breakthrough as if in an act of scientific desperation. In 1918 Planck received the Nobel Prize "*for services rendered to the advancement of Physics by his discovery of energy quanta*." Einstein recalls jokingly Planck's reported lack of full confidence in general relativity theory: "*Planck was one of the most outstanding people I have ever known (...) In reality, however, he did not understand physics. During the solar eclipse in* 1919 *he stayed awake all night, to see whether light bending in the gravitational field will be confirmed. If he understood the very essence of the general relativity theory, he would quietly go to bed, as I did*" (cited by Ernst Straus in "*Einstein: A Centenary Volume*," p. 31).

Black body radiation

John William Strutt, Lord Rayleigh (1842–1919), British physicist, Cavendish Professor at Cambridge, contributed greatly to physics (wave propagation, light scattering theory – Rayleigh scattering). In 1904 Rayleigh received the Nobel Prize "*for his investigations of the densities of the most important gases and for his discovery of argon in connection with these studies.*" ©The Nobel Foundation.

Max Planck wanted to understand black body radiation. The black body may be modeled by a box with a small hole (Fig. 1.1). We heat the box up, wait for the system to reach a stationary state (at a fixed temperature) and see what kind of electromagnetic radiation (intensity as a function of frequency) comes out of the hole. In 1900 Rayleigh and Jeans[5] tried to apply classical mechanics to this problem, and calculated correctly that the black body would emit electromagnetic radiation having a distribution of frequencies. However, the larger the frequency the larger its intensity, leading to what is known as ultraviolet catastrophe, an absurd conclusion. Experiment contradicted theory (Fig. 1.1).

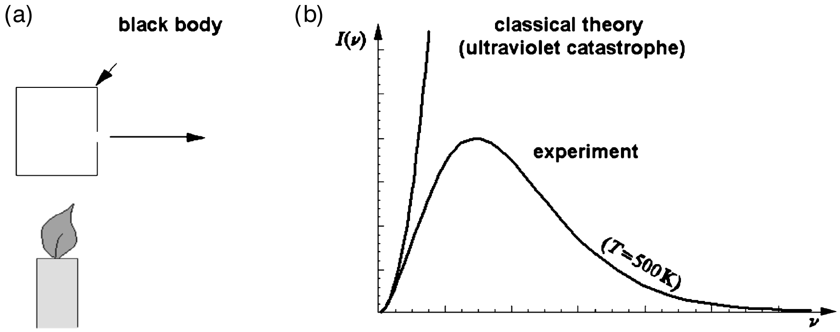


Fig. 1.1. Black body radiation. (a) As one heats a box to temperature T, the hole emits electromagnetic radiation with a wide range of frequencies. (b) The distribution of intensity as a function of frequency ν. There is a serious discrepancy between the results of classical theory and the experiment, especially for large frequencies. Only after assuming the existence of energy quanta can theory and experiment be reconciled.

At a given temperature T the intensity distribution (at a given frequency ν, Fig. 1.1b) has a single maximum. As the temperature increases, the maximum should shift towards higher frequencies (a piece of iron appears red at 500°C, but bluish at 1000°C). Just like Rayleigh and Jeans, Max Planck was unable to derive this simple qualitative picture from classical theory – something had to be done. On December 14, 1900, the generally accepted date for the birth of quantum theory, Planck presented his theoretical results for the black body treated as an ensemble of harmonic oscillators. With considerable reluctance, he postulated[6] that matter cannot emit radiation otherwise than by equal portions ("quanta") of energy $h\nu$, proportional to the frequency ν of vibrations of a single oscillator of the black body. The famous Planck constant h followed soon after ($h = 6.62607 \cdot 10^{-34}$ Js; but in this book, we will use

---

[5] James Hopwood Jeans (1877–1946), British physicist, professor at the University of Cambridge and at the Institute for Advanced Study in Princeton. Jeans also made important discoveries in astrophysics (e.g., the theory of double stars).

[6] He felt uncomfortable with this idea for many years.

a more convenient constant[7] $\hbar = \frac{h}{2\pi}$). It is exactly this hypothesis about energy quanta that led to the agreement of theory with experiment and the elimination of the ultraviolet catastrophe.

1905 – Albert Einstein

Photoelectric effect

The second worrying problem, apart from the black body, was the photoelectric effect.[8] Light knocks electrons[9] out of metals, but only when its frequency exceeds a certain threshold. Classical physics was helpless. In classical theory, light energy should be stored in the metal in a continuous way and *independent of the frequency used, after a sufficient period of time, the electrons should be ejected from the metal*. Nothing like that was observed. Einstein introduced the idea of electromagnetic radiation quanta as *particles*, later baptized *photons* by Gilbert Lewis. Note that Planck's idea of a quantum concerned energy transfer *from the black body to the electromagnetic field*, while Einstein introduced it *for the opposite direction* with the energy corresponding to Planck's quantum. Planck considered the quantum as a portion of energy, while for Einstein, the quantum meant a particle.[10] Everything be-

---

[7] Known as "h bar."

[8] Experimental work on the effect had been done by Philipp Eduard Anton Lenard (1862–1947), German physicist, professor at Breslau (now Wrocław), Köln, and Heidelberg. Lenard discovered that the number of photoelectrons is proportional to the intensity of light, and that their kinetic energy *does not depend at all on the intensity*, depending instead on the *frequency* of light. Lenard received the Nobel Prize in 1905 "*for his work on cathode rays*." A faithful follower of Adolf Hitler, and devoted to the barbarous Nazi ideas, Lenard terrorized German science. He demonstrates that scientific achievement and decency are two separate human characteristics.

[9] The electron was already known, having been predicted as early as 1868 by the Irish physicist George Johnstone Stoney (1826–1911), and finally discovered in 1897 by the British physicist Joseph John Thomson (1856–1940). Thomson also discovered a strange pattern: the number of electrons in light elements was equal to about one half of their atomic mass. Free electrons were obtained much later (1906). The very existence of atoms was still a hypothesis. The atomic nucleus was to be discovered only in 1911. Physicists were also anxious about the spectra of even the simplest substances such as hydrogen. Johann Jacob Balmer, a teacher from Basel, was able to design an astonishingly simple formula which fitted perfectly some of the observed lines in the hydrogen spectrum ("Balmer series"). All that seemed mysterious and intriguing.

[10] It is true that Einstein wrote about "point-like quanta" four years later, in a careful approach identifying the quantum as a particle. Modern equipment enables us to count photons, the individual particles of light, but the human eye is also capable of detecting 6–8 photons striking a neuron.

came clear: energy goes to electrons by *quanta* and this is why only quanta exceeding some threshold (the binding energy of an electron in the metal) are able to eject electrons from a metal.

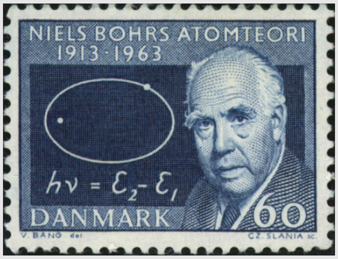
Gilbert Newton Lewis (1875–1946), the greatest American chemist, who advanced American chemistry internationally through his research and teaching. In a 1926 article in *Nature* Lewis introduced the name of the "photon." He also developed an early theory of chemical bonding ("Lewis structures") based on counting the valence electrons and forming "octets" from them. The idea that atoms in molecules tend to form octets in order to complete their electron shells turned out to be surprisingly useful in predicting bond patterns in molecules. A drawback for this concept is that it was not closely connected to the ideas of theoretical physics. It is an example of an extremely clever concept rather than of a coherent theory. Lewis also introduced a new definition of acids and bases, which is still in use.

1911 – Ernest Rutherford

Rutherford proved experimentally that atoms have a massive nucleus that is very small when compared to the size of the atom. The positive charge is concentrated in the nucleus, which is about 10^{-13} cm in size. The density of the nuclear matter boggles the imagination: 1 cm^3 has a mass of about 300 million tones. This is how researchers found out that an atom is composed of a massive nucleus and electrons.

In 1905, the accuracy of experimental data was too poor to confirm Einstein's theory as the only one which could account for the experimental results. Besides, the wave nature of light was supported by thousands of crystal clear experiments. Einstein's argument was so breathtaking (...particles???) that Robert Millikan decided to falsify experimentally Einstein's hypothesis. However, after 10 years of investigations, Millikan acknowledged that he was forced to undoubtedly support Einstein's explanation "however absurd it may look" (*Rev. Modern Phys.*, 21(1949)343). This conversion of a sceptic inclined the Nobel Committee to grant Einstein the Nobel Prize in 1923 "*for his work on the elementary charge of electricity and on the photo-electric effect*."

1913 – Niels Bohr



Niels Hendrik Bohr (1885–1962), Danish physicist, professor at Copenhagen University, played a key role in the creation and interpretation of quantum mechanics (see the end of this chapter). Bohr was born in Copenhagen, the son of a professor of physiology. He graduated from Copenhagen university and in 1911 obtained his doctorate there. Then he went to Cambridge to work under the supervision of J.J. Thomson, the discoverer of the electron. The collaboration did not work out, and in 1912 Bohr began to cooperate with Ernest Rutherford at the University of Manchester. In Manchester Niels Bohr made a breakthrough by introducing a planetary model of the hydrogen atom. He *postulated* that the angular orbital momentum must be quantized. Using this, Bohr reproduced the experimental spectrum of hydrogen atom with high accuracy. In 1922 Bohr received the Nobel Prize "*for his investigation of the structure of atoms*." In the same year he became the father of Aage Niels Bohr – a future winner of the Nobel Prize (1975, for studies of the structure of nuclei). In October 1943, Bohr and his family fled from Denmark to Sweden, and then to Great Britain and the USA, where he worked on the Manhattan Project. After the war the Bohr family returned to Denmark.

The model of the hydrogen atom

Atomic spectra were the third great mystery of early 20th-century physics. Even interpretation of the spectrum of the hydrogen atom represented a challenge. At the age of 28 (in 1913) Bohr proposed a simple planetary model of this atom, in which the electron, *contrary to classical mechanics*, did not fall onto the nucleus. Instead, it changed its orbit with accompanying absorption or emission of energy quanta. Bohr assumed that angular orbital momentum is quantized and that the centrifugal force is compensated by the Coulomb attraction between the electron and the nucleus. He was able to reproduce part of the spectrum of the hydrogen atom very accurately. Bohr then began work on the helium atom, which turned out to be a disaster, but he was successful again with the helium cation[11] He^+.

Niels Bohr played an inspiring role in the development and popularization of quantum mechanics. His Copenhagen Institute for Theoretical Physics, founded in 1921, where many young theoreticians from all over the world worked on quantum mechanical problems, was the leading

---

[11] Bohr did not want to publish without good results for all other atoms, something he would never achieve. Rutherford argued: "*Bohr, you explained hydrogen, you explained helium, people will believe you for other atoms*."

world center in the 1920s and 1930s.[12] Bohr, with Werner Heisenberg, Max Born, and John von Neumann, contributed greatly to the elaboration of the philosophical foundations of quantum mechanics. According to this, quantum mechanics represents a coherent and complete model of reality ("the world"), and the discrepancies with the classical mechanics have a profound and fundamental character,[13] and both theories coincide in the limit $h \to 0$ (where h is the Planck constant), and thus the predictions of quantum mechanics reduce to those of classical mechanics (known as Bohr's correspondence principle).

1916 – Arnold Sommerfeld

"Old quantum theory"

In 1916 Arnold Sommerfeld generalized the Bohr quantization rule beyond the problem of the one-electron atom. Known as "old quantum theory," it did not represent any coherent theory of general applicability. As a matter of fact, this quantization was achieved by assuming that for every periodic variable (like an angle), an integral is equal to an integer times the Planck constant.[14] Sommerfeld also tried to apply the Bohr model to atoms with a single valence electron (he had to modify the Bohr formula by introducing the quantum defect, i.e., a small change in the principal quantum number; see p. 232).

Arnold Sommerfeld (1868–1951), German physicist, professor at the Mining Academy in Clausthal, then at the Technical University of Aachen, in the key period 1906–1938, was professor at Munich University. Sommerfeld considered not only circular (Bohr-like) orbits, but also elliptical ones, and introduced the angular quantum number. He also investigated X-rays and the theory of metals. The scientific father of many Nobel Prize winners, he did not get this distinction himself.

---

[12] John Archibald Wheeler recalls that, when he first came to the Institute, he met a man working in the garden and asked him where he could find Professor Bohr. The gardener answered: "*That's me*."

[13] The center of the controversy was that quantum mechanics is indeterministic, while classical mechanics is deterministic, although this indeterminism is not all it seems. As will be shown later in this chapter, quantum mechanics is a *fully deterministic theory in the Hilbert space* (the space of all possible wave functions of the system), its indeterminism pertains to the physical space in which we live.

[14] Similar periodic integrals were used earlier by Bohr.

1923 – Louis de Broglie

Louis-Victor Pierre Raymond de Broglie (1892–1987) was studying history at the Sorbonne, carefully preparing himself for a diplomatic career quite natural in a princely family. His older brother Maurice, a radiographer, aroused his interest in physics. The First World War (Louis did his military service in a radio communications unit) and the study of history delayed his start in physics.

He was 32 when he presented his doctoral dissertation, which embarrassed his supervisor, Paul Langevin. The thesis, on the wave nature of all particles, was so revolutionary, that only a positive opinion from Einstein, who was asked by Langevin to take a look of the dissertation, convinced the doctoral committee. Only five years later (in 1929), Louis de Broglie received the Nobel Prize "*for his discovery of the wave nature of electrons*."

Waves of matter

In his doctoral dissertation, stuffed with mathematics, Louis de Broglie introduced the concept of "waves of matter." He postulated that not only photons, but *also any other particle*, has, besides its corpuscular characteristics, some wave properties (those corresponding to light had been known for a long, long time). According to de Broglie, the wave length corresponds to momentum p,

$$p = \frac{h}{\lambda}$$

where h is again the Planck constant! What kind of momentum can this be, in view of the fact that momentum depends on the laboratory coordinate system chosen? Well, it is the momentum measured in the same laboratory coordinate system as that used to measure the corresponding wave length.

1923 – Arthur Compton[15]

Electron–photon scattering

It turned out that an electron–photon collision obeys the same laws of dynamics as those describing collision of two particles: the energy conservation law and the momentum con-

---

[15] Arthur Holly Compton (1892–1962), American physicist, professor at the universities of Saint Louis and Chicago. He obtained the Nobel Prize in 1927 "*for the discovery of the effect named after him*," i.e., for investigations of electron–photon scattering.

servation law. This result confirmed the wave-corpuscular picture emerging from experiments.

1925 – George E. Uhlenbeck and Samuel A. Goudsmit

Discovery of spin

Two Dutch students explained an experiment (Stern–Gerlach) in which a beam of silver atoms passing through a magnetic field split into two beams. In a short paper, they suggested that the silver atoms have (besides their orbital angular momentum) an additional internal angular momentum (spin), *similar* to a macroscopic body, which besides its center-of-mass motion, also has a rotational (spinning) motion.[16] Moreover, the students demonstrated that the atomic spin follows from the spin of the electrons: among the 47 electrons of the silver atom, 46 have their spin compensated (23 "down" and 23 "up"), while the last "unpaired" electron gives the net spin of the atom.

1925 – Wolfgang Pauli[17]

Pauli exclusion principle

Pauli postulated that *in any system two electrons cannot be in the same state* (including their spins). This "Pauli exclusion principle" was deduced from spectroscopic data (some states were not allowed). It turned out the corner stone of understanding chemistry.

1925 – Werner Heisenberg

Matrix quantum mechanics

A paper by the 24-year-old Werner Heisenberg turned out to be a breakthrough in quantum theory.[18] He wrote in a letter: "*My whole effort is to destroy without a trace the idea of orbits*." Max Born recognized matrix algebra in Heisenberg's formulation (who, himself, had not

---

[16] Caution: *identifying* the spin with the rotation of a rigid body leads to physical inconsistencies.

[17] Pauli also introduced the idea of spin when interpreting spectra of atoms with a single valence electron. He was inspired by Sommerfeld, who interpreted the spectra by introducing the quantum number $j = l \pm \frac{1}{2}$, where the quantum number l quantized the orbital angular momentum of the electron. Pauli described spin as *a bivalent nonclassical characteristic of the electron* in W. Pauli, *Zeit. Phys. B*, 3(1925)765.

[18] On June 7, 1925, Heisenberg was so tired after a bad attack of hay fever that he decided to go and relax on the North Sea island of Helgoland. Here, he divided his time between climbing the mountains, learning Goethe's poems by heart, and (despite his intention to rest) hard work on the spectrum of the hydrogen atom with which he was obsessed. It was at night on 7 or 8 June that he saw something – the beginning of the new mechanics. In

yet realized it) and in the same year a more solid formulation of the new mechanics ("matrix mechanics") was proposed by Werner Heisenberg, Max Born, and Pascual Jordan.[19]

1926 – Erwin Schrödinger

Schrödinger equation

In November 1925, Erwin Schrödinger delivered a lecture at the Technical University (ETH) in Zurich, in which he presented the results of de Broglie. Professor Peter Debye stood up and asked the speaker:

Peter Joseph Wilhelm Debye, or more exactly, Peter Josephus Wilhelmus Debye (1884–1966), Dutch physicist and chemist, professor at the Technical University (ETH) of Zurich (1911, 1920–1937) as well as at Göttingen, Leipzig, and Berlin, won the Nobel Prize in chemistry in 1936 "*for his contribution to our knowledge of molecular structure through his investigations on dipole moments and on the diffraction of X-rays and electrons in gases*." Debye emigrated to the USA in 1940, where he obtained a professorship at Cornell University in Ithaca, NY (and remained in this beautiful town to the end of his life). His memory is still alive there. Professor Scheraga remembers him as an able chair in seminar discussions, in the tradition of the Zurich seminar of 1925.

"*You are telling us about waves, but where is the wave equation in your talk?*" Indeed, there wasn't any! Schrödinger began to work on this and the next year formulated what is now called wave mechanics based on the wave equation. Both formulations, Heisenberg's and Schrödinger's,[20] turned out to be equivalent and are now known as (nonrelativistic) quantum mechanics.

---

later years he wrote in his book "*Der Teil und das Ganze*": "*It was about three o'clock in the morning when the final result of the calculation lay before me. At first I was deeply shaken. I was so excited that I could not think of sleep. So I left the house and awaited the sunrise on the top of a rock*." The first man with whom Heisenberg shared his excitement a few days later was his schoolmate Wolfgang Pauli, and, after another few days, also with Max Born.

[19] Jordan, despite his talents and achievements, felt himself underestimated and even humiliated in his native Germany. For example, he had to accept a position at Rostock University, which the German scientific elite used to call the "Outer Mongolia of Germany." The best positions seemed to be reserved. When Hitler came to power, Jordan became a fervent follower.

[20] And the formulation proposed by Paul A.M. Dirac.

1926 – Max Born

Statistical interpretation of wave function

Max Born (1882–1970), German physicist, professor at the universities of Göttingen, Berlin, Cambridge, and Edinburgh, born in Breslau (now Wrocław) to the family of a professor of anatomy. Born studied first in Wrocław, then at Heidelberg and Zurich. He received his PhD in physics and astronomy in 1907 in Göttingen, where he began his swift academic career. Born obtained a chair at the University of Berlin in 1914, and returned to Göttingen in 1921, where he founded an outstanding school of theoretical physics, which competed with the famous institute of Niels Bohr in Copenhagen. Born supervised Werner Heisenberg, Pascual Jordan, and Wolfgang Pauli. It was Born who recognized, in 1925, that Heisenberg's quantum mechanics could be formulated in terms of matrix algebra. Together with Heisenberg and Jordan, he created the first consistent quantum theory (the famous "*Dreimännerarbeit*"). After Schrödinger's formulation of quantum mechanics, Born proposed the probabilistic interpretation of the wave function. Despite such seminal achievements, the Nobel Prizes in the 1930s were received by his colleagues. Finally, when in 1954 Born obtained the Nobel Prize "*for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wave-function*," there was a great relief among his famous friends. ©The Nobel Foundation.

Max Born proposed interpreting the square of the complex modulus of Schrödinger's wave function as the probability density for finding the particle.

1927 – Werner Heisenberg

Uncertainty principle

When analyzing a hypothetical microscope observation of an electron Heisenberg concluded that it is not possible to measure simultaneously the position (x) and momentum (p_x) of a particle with any desired accuracy. The more exactly we measure the position (small Δx), the larger the error we make in measuring the momentum (large Δp_x) and *vice versa.*

1927 – Clinton Davisson, Lester H. Germer, George Thomson[21]

Electron diffraction

Davisson and Germer, and Thomson, demonstrated in ingenious experiments that indeed electrons do exhibit wave properties (using crystals as diffraction gratings).

1927 – Walter Heitler, Fritz Wolfgang London

The birth of quantum chemistry

Walter Heitler and Fritz Wolfgang London convincingly explained why two neutral atoms (like hydrogen) attract each other with a force so strong as to be comparable with the Coulomb forces between ions. Applying the Pauli exclusion principle when solving the Schrödinger equation is of key importance. Their paper was received on June 30, 1927, by *Zeitschrift für Physik*, and this may be considered as *the birthday of quantum chemistry*.[22]

1928 – Paul Dirac

Dirac equation for the electron and positron

Paul Dirac's main achievements are the foundations of quantum electrodynamics and construction of the relativistic wave equation (1926–1928) which now bears his name. The equation not only describes the electron, but also its antimatter counterpart – the positron (predicting antimatter). Spin was also inherently present in the equation.

1929 – Werner Heisenberg and Wolfgang Pauli

Quantum field theory

Two classmates developed a theory of matter, and the main features still survive. In this theory, the elementary particles (the electron, photon, and so on) were viewed as excited states of the corresponding fields (the electron field, electromagnetic field, and so on).

---

[21] Clinton Joseph Davisson (1881–1958), American physicist at Bell Telephone Laboratories. He discovered the diffraction of electrons with L.H. Germer, and they received the Nobel Prize in 1937 "*for their experimental discovery of the diffraction of electrons by crystals*." The prize was shared with George Paget Thomson (1892–1975), son of the discoverer of the electron, Joseph John Thomson, and professor at Aberdeen, London, and Cambridge, who used a different diffraction method.

[22] The term "quantum chemistry" was first used by Arthur Haas in his lectures to the Physicochemical Society of Vienna in 1929 (A. Haas, "*Die Grundlagen der Quantenchemie. Eine Einleitung in vier Vortragen*," Akademische Verlagsgesellschaft, Leipzig, 1929).

1932 – Carl Anderson[23]

Discovery of antimatter (the positron)

One of Dirac's important results was the observation that his relativistic wave equation is satisfied, not only by the electron but also by a mysterious unknown particle, the positive electron (positron). This antimatter hypothesis was confirmed by Carl Anderson, who found the positron experimentally, a victorious day for quantum theory.

1948 – Richard Feynman, Julian Schwinger, Shinichiro Tomonaga[24]

Quantum electrodynamics

The Dirac equation did not take all the physical effects into account. For example, the strong electric field of the nucleus polarizes the vacuum so much, that electron–positron pairs emerge from the vacuum and screen the electron–nucleus interaction. The quantum electrodynamics developed by Feynman, Schwinger, and Tomonaga accounts for this and similar effects, and brings theory and experiment to an agreement of unprecedented accuracy.

1964 – John Bell

Bell inequalities

The mathematician John Bell proved that, if particles have certain properties *before measurement* (so that they were small but *classical objects*), then the measurement results would have to satisfy some inequalities which contradict the predictions of quantum mechanics (further details at the end of this chapter).

1982 – Alain Aspect

Is the world nonlocal?

Experiments with photons showed that the Bell inequalities are *not* satisfied. This means that either there is *instantaneous communication* even between extremely distant particles ("entangled states"), or that the particles *do not have some definite properties* before the measurement is performed (more details at the end of this chapter).

---

[23] More details in Chapter 3.

[24] All received the Nobel Prize in 1965 "*for their fundamental work in quantum electrodynamics, with fundamental implications for the physics of elementary particles*."

1997 – Anton Zeilinger

Teleportation of the photon state

A research group at the University of Innsbruck used entangled quantum states (see p. 48) to perform teleportation of a photon state,[25] that is, to prepare at a distance any state of a photon with simultaneous disappearance of this state from the teleportation site (details at the end of this chapter).

1.2 Postulates of quantum mechanics

All science is based on a number of postulates. Quantum mechanics has also elaborated a system of postulates that have been formulated to be as simple as possible and yet to remain in accordance with experimental results. Postulates are not supposed to be proved; their justification is efficiency. Quantum mechanics, the foundations of which date from 1925–1926, still represents the basic theory of phenomena within atoms and molecules. This is the domain of chemistry, biochemistry, and atomic and nuclear physics. Further progress (quantum electrodynamics, quantum field theory, elementary particle theory) permitted deeper insights into the structure of the atomic nucleus, but did not produce any fundamental revision of our understanding of atoms and molecules. Matter as described at a nonrelativistic[26] quantum mechanics level represents a system of electrons and nuclei, treated as point-like particles with a definite mass and electric charge, moving in three-dimensional space and interacting by *electrostatic* forces.[27] This model of matter is at the core of quantum chemistry; see Fig. 1.2.

The assumptions on which quantum mechanics is based may be given in the form of the following Postulates I–VI. For simplicity, we will restrict ourselves to a single particle moving along a single coordinate axis x (the mathematical foundations of quantum mechanics are given in Appendix B on p. 595).

---

[25] M. Eibl, H. Weinfurter, A. Zeilinger, *Nature*, 390(1997)575.

[26] Assuming that the speed of light is infinite.

[27] Yes, we take only electrostatics, that is, Coulomb interactions, into account. It is true that a moving charged particle creates a magnetic field, which influences its own and other particles' motion. This however (the Lorentz force) is taken into account in the *relativistic* approach to quantum mechanics.

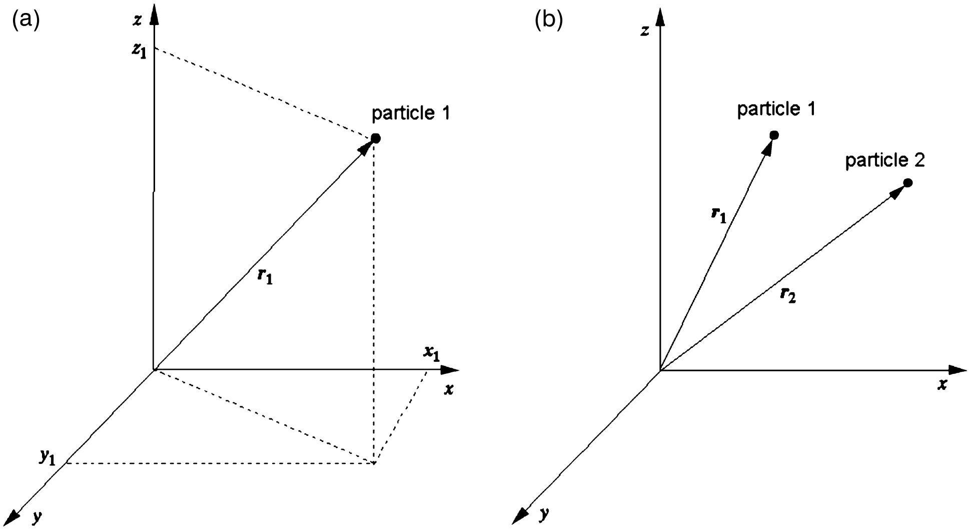


Fig. 1.2. An atom (molecule) in nonrelativistic quantum mechanics. (a) A Cartesian ("laboratory") coordinate system is introduced into three-dimensional space. (b) We assume that all the particles (electrons and nuclei) are point-like (the figure shows their instantaneous positions) and interact only by electrostatic (Coulomb) forces.

Postulate I (on the quantum mechanical state)

> The state of the system is described by the wave function $\Psi = \Psi(x,t)$, which depends on the coordinate of particle x at time t. Wave functions in general are complex functions of real variables. The symbol $\Psi^*(x,t)$ denotes the complex conjugate of $\Psi(x,t)$. The quantity
>
> $$p(x,t) = \Psi^*(x,t)\Psi(x,t)\,dx \tag{1.1}$$
>
> gives the *probability that at time t the x coordinate of the particle lies in the small interval* $[x, x+dx]$ (Fig. 1.3a). The probability of the particle being in the interval (a,b) on the x axis is given by Fig. 1.3b.

The probabilistic interpretation of the wave function was proposed by Max Born.[28] By analogy with the formula mass = density × volume, the quantity $\Psi^*(x,t)\Psi(x,t)$ is called the *probability density* that a particle at time t has position x.

---

[28] M. Born, *Zeitschrift für Physik*, 37(1926)863.

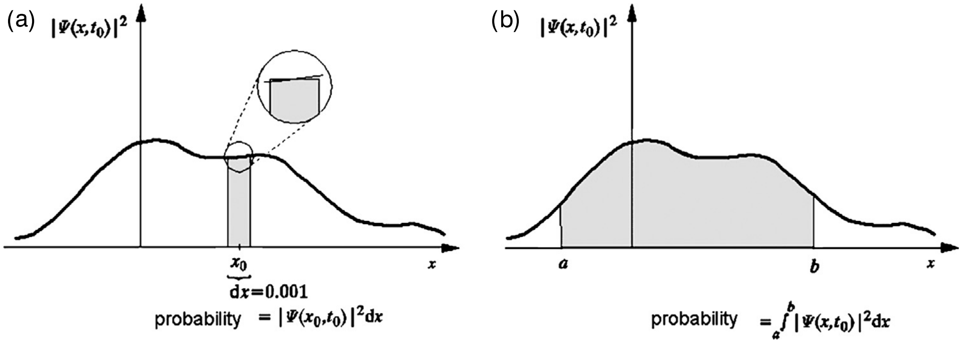


Fig. 1.3. A particle moves along the x axis and is in the state described by the wave function $\Psi(x,t)$. (a) The figure shows how the probability of finding a particle in an *infinitesimally small* section of the length dx at x_0 (at time $t = t_0$) is calculated. It is not important where exactly in section $[x, x+dx]$ the number x_0 really is, because of the infinitesimal length of the section. In the figure the number is positioned in the middle of the section. (b) The figure shows how to calculate the probability of finding the particle at $t = t_0$ in a *section* (a,b).

In order to treat the quantity $p(x,t)$ as a probability, at any instant t the wave function must satisfy the *normalization condition*

$$\int_{-\infty}^{\infty} \Psi^*(x,t)\Psi(x,t)\,dx = 1\,. \tag{1.2}$$

All this may be generalized for more complex situations. For example, in three-dimensional space, the wave function of a single particle depends on position $\mathbf{r} = (x, y, z)$ and time, $\Psi(\mathbf{r}, t)$, and the *normalization condition* takes the form

$$\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz \Psi^*(x,y,z,t)\Psi(x,y,z,t) \equiv \int \Psi^*(\mathbf{r},t)\Psi(\mathbf{r},t)\,dV \equiv \int \Psi^*(\mathbf{r},t)\Psi(\mathbf{r},t)d\mathbf{r} = 1\,. \tag{1.3}$$

For simplicity, the last two integrals are given without the integration limits, but they are there *implicitly, and this convention will be used throughout the book unless stated otherwise*.

For n particles (Fig. 1.4), shown by vectors $\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_n$ in three-dimensional space, the interpretation of the wave function is as follows. The probability P that, at a given time $t = t_0$, particle 1 is in the domain V_1, particle 2 is in the domain V_2, etc., is computed as

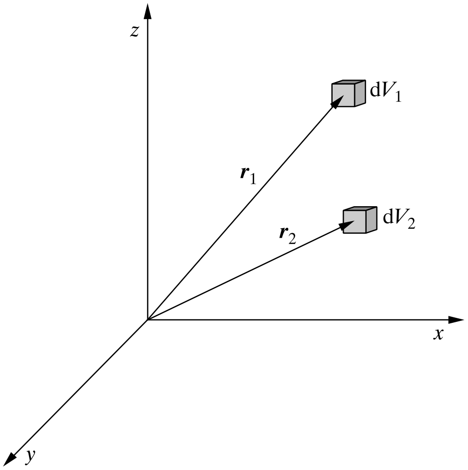


Fig. 1.4. Interpretation of a many-particle wave function, an example for two particles. The number $|\psi(\mathbf{r}_1, \mathbf{r}_2, t_0)|^2 dV_1 dV_2$ represents the probability that at $t = t_0$ particle 1 is in its box of volume dV_1 shown by vector $\mathbf{r}_1$ and particle 2 in its box of volume dV_2 indicated by vector $\mathbf{r}_2$.

$$\begin{aligned} P &= \int_{V_1} dV_1 \int_{V_2} dV_2 ... \int_{V_n} dV_n \Psi^*(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_n, t_0)\Psi(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_n, t_0) \\ &\equiv \int_{V_1} d\mathbf{r}_1 \int_{V_2} d\mathbf{r}_2 ... \int_{V_n} d\mathbf{r}_n \Psi^*(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_n, t_0)\Psi(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_n, t_0). \end{aligned}$$

Often in this book we will perform what is called normalization of a function, which is required if a probability is to be calculated. Suppose we have a nonnormalized function[29] ψ, i.e.,

$$\int_{-\infty}^{\infty} \psi(x,t)^* \psi(x,t)\, dx = A, \tag{1.4}$$

with $0 < A \neq 1$. To compute the probability ψ must be normalized, i.e., multiplied by a *normalization constant* N, such that the new function $\Psi = N\psi$ satisfies the normalization condition, $1 = \int_{-\infty}^{\infty} \Psi^*(x,t)\Psi(x,t)\, dx = N^*N \int_{-\infty}^{\infty} \psi^*(x,t)\psi(x,t)\, dx = A|N|^2$. Hence, $|N| = \frac{1}{\sqrt{A}}$. How is N computed? One person may choose it equal to $N = \frac{1}{\sqrt{A}}$, another as $N = -\frac{1}{\sqrt{A}}$, a third as $N = e^{1989i}\frac{1}{\sqrt{A}}$, and so on. There are, therefore, an infinite number of legitimate choices of the *phase* ϕ of the wave function $\Psi(x,t) = e^{i\phi}\frac{1}{\sqrt{A}}\psi$. Yet, when $\Psi^*(x,t)\Psi(x,t)$ is calculated, everyone will obtain *the same* result, $\frac{1}{A}\psi^*\psi$, because the phase disappears. In most applications, this is what will happen and therefore the computed physical properties will not depend on the choice of phase. There are cases, however, where the phase will be of importance.

---

[29] Eq. (1.3) is not satisfied.

Postulate II (on operator representation of mechanical quantities)

The mechanical quantities that describe the particle (energy, the components of vectors of position, momentum, angular momentum, etc.) are represented by linear operators acting in the Hilbert space (see Appendix B). There are two important examples of the operators: the operator of the particle's position $\hat{x} = x$ (i.e., multiplication by x, or $\hat{x} = x\cdot$; Fig. 1.5) and the operator of the (x component) momentum $\hat{p}_x = -i\hbar\frac{d}{dx}$, where i stands for the imaginary unit.

Mechanical quantity	**Classical formula**	**Operator acting on f**
coordinate	x	$\hat{x} f \overset{\text{def}}{=} x f$
momentum component	p_x	$\hat{p}_x f \overset{\text{def}}{=} -i\hbar\frac{\partial f}{\partial x}$
kinetic energy	$T = \frac{mv^2}{2} = \frac{p^2}{2m}$	$\hat{T} f \overset{\text{def}}{=} -\frac{\hbar^2}{2m}\Delta f$

Fig. 1.5. Mechanical quantities and the corresponding operators.

Note that the mathematical form of the operators is always defined with respect to a *Cartesian* coordinate system.[30] From the given operators (Fig. 1.5) the operators of some other quantities may be constructed. The potential energy operator $\hat{V} = V(x)$, where $V(x)$ (the multiplication operator by the function $\hat{V} f = V(x) f$) represents a function of x called a potential. The kinetic energy operator of a single particle (in one dimension) $\hat{T} = \frac{\hat{p}_x{}^2}{2m} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}$, and in three dimensions

$$\hat{T} = \frac{\hat{p}^2}{2m} = \frac{\hat{p}_x{}^2 + \hat{p}_y{}^2 + \hat{p}_z{}^2}{2m} = -\frac{\hbar^2}{2m}\Delta, \tag{1.5}$$

where the (ubiquitous in this book) Laplacian operator Δ is defined as

$$\Delta \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{1.6}$$

and m denotes the particle's mass. The total energy operator, or *Hamiltonian*, is the most frequently used:

---

[30] Although they may *then* be transformed to other coordinates systems.

$$\hat{H} = \hat{T} + \hat{V}. \tag{1.7}$$

An important feature of operators is that they may not *commute*, i.e., for two particular operators $\hat{A}$ and $\hat{B}$ it may happen that $\hat{A}\hat{B} - \hat{B}\hat{A} \neq 0$. This property has important physical consequences (see below, Postulate IV and the Heisenberg uncertainty principle). Because of the possible noncommutation of the operators, transformation of the classical formula (in which the commutation or noncommutation did not matter) may be nonunique. In such a case, from all the possibilities one has to choose an operator which is Hermitian. The operator $\hat{A}$ is Hermitian if for any functions ψ and ϕ from its domain one has

$$\int_{-\infty}^{\infty} \psi^*(x)\hat{A}\phi(x)\,dx = \int_{-\infty}^{\infty} [\hat{A}\psi(x)]^*\phi(x)\,dx\,. \tag{1.8}$$

Using what is known as *Dirac notation*, Fig. 1.6, the above equality may be written in a concise form:

$$\langle\psi|\hat{A}\phi\rangle = \langle\hat{A}\psi|\phi\rangle\,. \tag{1.9}$$

$\int \psi^*\phi\,\mathrm{d}\tau \equiv \langle\psi\|\phi\rangle$	Scalar product of two functions
$\int \psi^*\hat{A}\phi\,\mathrm{d}\tau \equiv \left\langle\psi\|\hat{A}\phi\right\rangle$ or $\left\langle\psi\|\hat{A}\|\phi\right\rangle$	A matrix element of the operator $\hat{A}$
$\hat{Q} = \|\psi\rangle\langle\psi\|$	Projection operator on the direction of the vector ψ
$1 = \sum_k \|\psi_k\rangle\langle\psi_k\|$	Spectral resolution of identity. Its sense is best seen when acting on function χ: $\chi = \sum_k \|\psi_k\rangle\langle\psi_k\|\chi\rangle = \sum_k \|\psi_k\rangle\langle\psi_k\|\chi\rangle =$ $\sum_k \|\psi_k\rangle c_k$

Fig. 1.6. Dirac notation.

In Dirac notation[31] (Fig. 1.6) the key role is played by vectors *bra*: $\langle\,|$ and *ket*: $|\,\rangle$, denoting respectively $\psi^* \equiv \langle\psi|$ and $\phi \equiv |\phi\rangle$. Writing bra and ket side by side as $\langle\psi||\phi\rangle$ denotes $\langle\psi|\phi\rangle$,

---

[31] Its deeper meaning is discussed in many textbooks of quantum mechanics, e.g., A. Messiah, "*Quantum Mechanics*," vol. I, Amsterdam (1961), p. 245. Here we treat it as a convenient tool.

or the scalar product of ψ and ϕ in a unitary space (Appendix B), while writing it as $|\psi\rangle\langle\phi|$ means the operator $\hat{Q} = |\psi\rangle\langle\phi|$, because of its action on function $\xi = |\xi\rangle$ shown as $\hat{Q}\xi = |\psi\rangle\langle\phi|\xi = |\psi\rangle\langle\phi|\xi\rangle = c\psi$, where $c = \langle\phi|\xi\rangle$:

- $\langle\psi|\phi\rangle$ denotes a scalar product of two functions (i.e., vectors of the Hilbert space) ψ and ϕ, also known as the overlap integral of ψ and ϕ,
- $\left\langle\psi|\hat{A}\phi\right\rangle$, or $\left\langle\psi|\hat{A}|\phi\right\rangle$, stands for the scalar product of two functions: ψ and $\hat{A}\phi$, or the matrix element of operator $\hat{A}$,
- $\hat{Q} = |\psi\rangle\langle\psi|$ means the projection operator on the vector ψ (in the Hilbert space),
- the last formula (with $\{\psi_k\}$ representing the complete set of functions) represents what is known as "spectral resolution of identity," best demonstrated when acting on an arbitrary function χ:

$$\chi = \sum_k |\psi_k\rangle\langle\psi_k|\chi\rangle = \sum_k |\psi_k\rangle c_k .$$

We have obtained the decomposition of the function (i.e., a vector of the Hilbert space) χ on its components $|\psi_k\rangle c_k$ along the basis vectors $|\psi_k\rangle$ of the Hilbert space. The coefficient $c_k = \langle\psi_k|\chi\rangle$ is the corresponding scalar product, the basis vectors ψ_k are normalized. This formula says something trivial: any vector can be retrieved when adding *all* its components together.

Postulate III (on time evolution of the state)

TIME-DEPENDENT SCHRÖDINGER EQUATION

The time evolution of the wave function Ψ is given by the equation

$$i\hbar\frac{\partial\Psi(x,t)}{\partial t} = \hat{H}\Psi(x,t), \tag{1.10}$$

where $\hat{H}$ is the system Hamiltonian (see Eq. (1.7)); $\hat{H}$ may be time-dependent (energy changes in time, interacting system) or time-independent (energy conserved, isolated system). Eq. (1.10) is called the time-dependent Schrödinger equation (Fig. 1.7).

When $\hat{H}$ *is time-independent*, the general solution to (1.10) can be written as

$$\Psi(x,t) = \sum_{n=1}^{\infty} c_n \Psi_n(x,t) , \tag{1.11}$$

where $\Psi_n(x,t)$ represent special solutions to (1.10) that have the form

$$\Psi_n(x,t) = \psi_n(x)\, e^{-i\frac{E_n}{\hbar}t} , \tag{1.12}$$

$$\boxed{\begin{array}{c}\Psi(x,t_0)\\ \downarrow\\ \hat{H}\Psi(x,t_0)\\ \downarrow\\ i\hbar\left(\frac{\partial\Psi}{\partial t}\right)_{t=t_0}\\ \downarrow\\ \Psi(x,t_0+\mathrm{d}t)=\Psi(x,t_0)-\frac{i}{\hbar}\hat{H}\Psi\mathrm{d}t\end{array}}$$

Fig. 1.7. Time evolution of a wave function. Knowing $\Psi(x,t)$ at a certain $t=t_0$ makes it possible to compute $\hat{H}\Psi(x,t_0)$, and from this (using Eq. (1.10)) one can calculate the time derivative $\frac{\partial\Psi(x,t_0)}{\partial t}=-\frac{i\hat{H}\Psi(x,t_0)}{\hbar}$. Knowledge of the wave function at time $t=t_0$, and of its time derivative, is sufficient to calculate the function a little later ($t=t_0+\mathrm{d}t$).

and c_n stand for some constants. Substituting the special solution to (1.10) leads to[32] what is known as the *time-independent Schrödinger equation*:

SCHRÖDINGER EQUATION FOR STATIONARY STATES

$$\hat{H}\psi_n=E_n\psi_n\,,\quad n=1,2,\ldots,\infty\,. \tag{1.13}$$

---

[32] $i\hbar\frac{\partial\Psi_n(x,t)}{\partial t}=i\hbar\frac{\partial\psi_n(x)\,e^{-i\frac{E_n}{\hbar}t}}{\partial t}=i\hbar\psi_n(x)\,\frac{\partial e^{-i\frac{E_n}{\hbar}t}}{\partial t}=i\hbar\psi_n(x)\,(-i\frac{E_n}{\hbar})e^{-i\frac{E_n}{\hbar}t}=E_n\psi_n e^{-i\frac{E_n}{\hbar}t}$. However, $\hat{H}\Psi_n(x,t)=\hat{H}\psi_n(x)\,e^{-i\frac{E_n}{\hbar}t}=e^{-i\frac{E_n}{\hbar}t}\,\hat{H}\psi_n(x)$, because the Hamiltonian does not depend on t. Hence, after dividing both sides of the equation by $e^{-i\frac{E_n}{\hbar}t}$ one obtains the time-independent Schrödinger equation. Therefore, the stationary state $\psi_n(x)\exp(-i\frac{E_n}{\hbar}t)$ is time-dependent, but this time dependence comes only from the factor $\exp(-i\frac{E_n}{\hbar}t)$. When, in the future, we would calculate the probability density $\Psi_n^*\Psi_n$, we would not need the factor $\exp(-i\frac{E_n}{\hbar}t)$, but its modulus only, whereas $\left|\exp(-i\frac{E_n}{\hbar}t)\right|=1$ for any value of t. The time dependence through $\exp(-i\frac{E_n}{\hbar}t)$ means that, as the time goes on, function ψ_n is multiplied by an oscillating complex number, which however never attains zero. The number oscillates on a circle of radius 1 within the complex plane. For example, limiting ourselves to the angles $m\cdot 90°, m=0,1,2,3,...$, in the time evolution of a stationary state, function $\psi_n(x)$ is multiplied by $1,i,-1,-i,...$, respectively. It is a bit frustrating though that the frequency of this rotation $\omega=\frac{E_n}{\hbar}$ depends on number E_n, because this number depends on yourself (you may add an arbitrary constant to the potential energy and the world will be functioning exactly as before). This is true, but whatever you compare with experiment you calculate $\Psi_m^*\,\Psi_n$, which annihilates the arbitrary constant and we get as time dependence $\exp(-i\omega t)$ with $\omega=\frac{E_n-E_m}{\hbar}$. Already at this point one can see that such oscillations might be damped by something, preferably of frequency just equal to ω. We will show this in detail in Chapter 2, and indeed it will turn out that such a damping by oscillating external electric field represents the condition to change the state Ψ_n to Ψ_m (and *vice versa*).

The equation represents an example of an *eigenvalue equation of the operator*; the functions ψ_n are called the eigenfunctions, and E_n are the eigenvalues of the operator $\hat{H}$ (we have assumed here that their number is equal to ∞). It can be shown that E_n are real (see Appendix B, p. 595). The eigenvalues are the permitted energies of the system, and the corresponding eigenfunctions Ψ_n are defined in Eqs. (1.12) and (1.13). These states have a special character, because the probability given by (1.1) does not change in time[33] (Fig. 1.8):

$$p_n(x,t) = \Psi_n^*(x,t)\Psi_n(x,t)\,dx = \psi_n^*(x)\psi_n(x)\,dx = p_n(x)\,. \tag{1.14}$$

Therefore, in determining these states, known as stationary states, one can apply the time-independent formalism based on the Schrödinger equation (1.13).

Postulate IV (on interpretation of experimental measurements)

This postulate pertains to ideal measurements, i.e., such that no error is introduced through imperfections in the measurement apparatus. We assume the measurement of the physical quantity A, represented by its time-independent operator $\hat{A}$ and, for the sake of simplicity, that the system is composed of a single particle (with one variable only).

- The result of a single measurement of a mechanical quantity A can be *only* an eigenvalue a_k of the operator $\hat{A}$.

  The eigenvalue equation for operator $\hat{A}$ reads

$$\hat{A}\phi_k = a_k\phi_k\,, \quad k = 1, 2, \ldots, M\,. \tag{1.15}$$

  The eigenfunctions ϕ_k are orthogonal[34] (cf. Appendix on p. 595). When the eigenvalues do not form a continuum, they are *quantized*, and then the corresponding eigenfunctions ϕ_k,

---

[33] There is a problem though. *Experiments show that this is true only for the ground state, not for the excited states, which turn out to be quasistationary only*. These experiments prove that in the excited states the system emits photons until it reaches the ground state. This excited states instability (which goes beyond the nonrelativistic approximation our book is focused on) comes from coupling with the electromagnetic field of the vacuum, the phenomenon ignored in presenting the postulates of quantum mechanics. The coupling is a real thing, because there are convincing experiments showing that the vacuum is not just nothing. It is true that the vacuum's mean electric field is zero, but the electromagnetic field fluctuates even in the absence of photons (the mean square of the electric field does not equal zero).

[34] If two eigenfunctions correspond to the same eigenvalue, they are not necessarily orthogonal, but they *can* still be orthogonalized (if they are linearly independent; see Appendix K, p. 697). Such orthogonal functions still remain the eigenfunctions of $\hat{A}$. Therefore, *one can always construct the orthonormal set of the eigenfunctions of a Hermitian operator*.

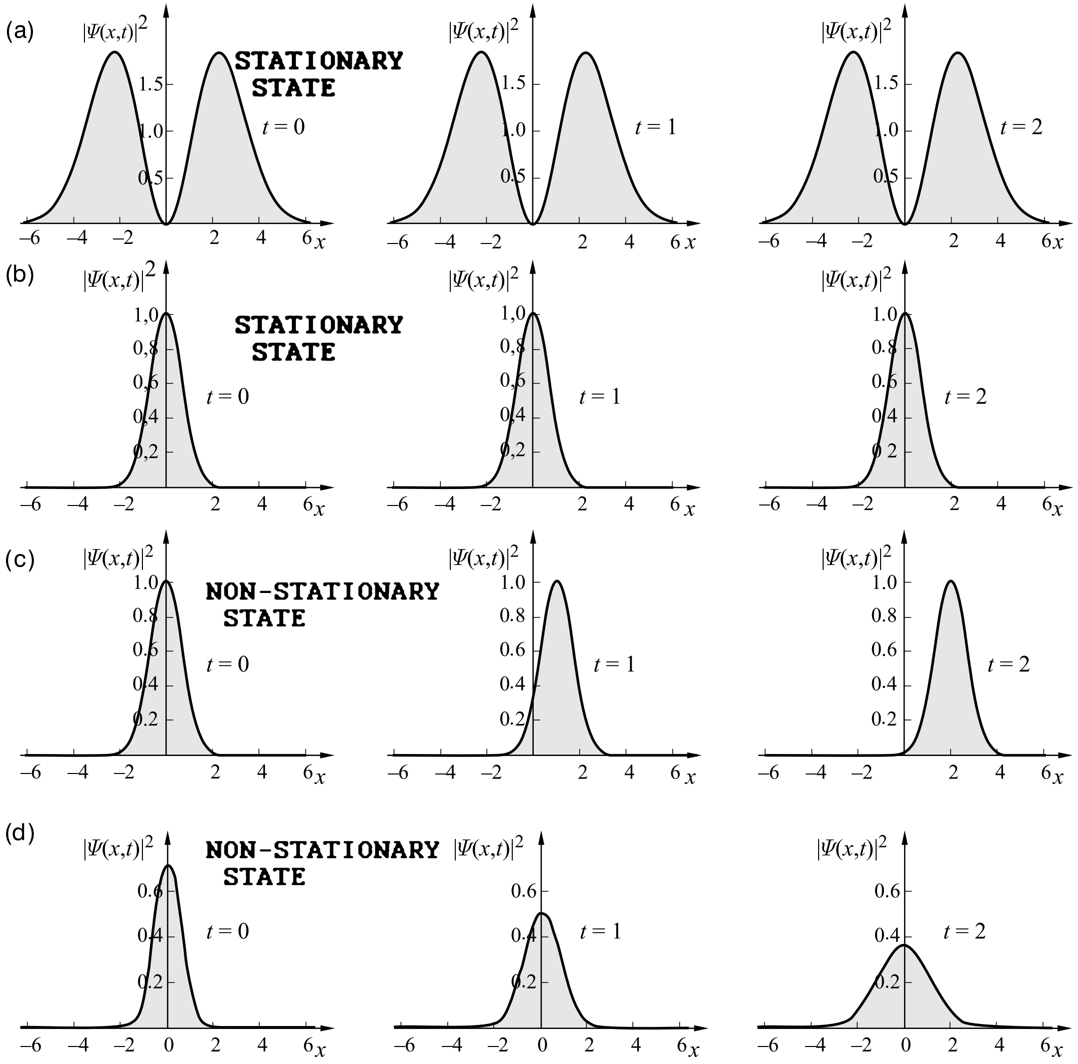


Fig. 1.8. Evolution of several starting wave functions Ψ (rows a–d) for a system shown as $|\Psi(x,t)|^2$ for three snapshots $t = 0, 1, 2$ (columns). In all cases the area under $|\Psi(x,t)|^2$ equals 1 (normalization of $\Psi(x,t)$). Cases (a) and (b) show $|\Psi(x,t)|^2$ as time-independent – these are stationary states, satisfying the time-independent Schrödinger equation. Contrary to this, in cases (c) and (d) function $|\Psi(x,t)|^2$ changes very much when the time goes on: in a translational motion with shape preserving in case (c) $|\Psi(x,t)|^2$ and irregularly in case (d). These are nonstationary states. The nonstationary states always represent linear combinations of stationary ones.

$k = 1, 2, \ldots, \infty$, satisfy the orthonormality relations[35]:

$$\int_{-\infty}^{\infty} \phi_k^*(x)\phi_l(x)\,dx \equiv \langle\phi_k|\phi_l\rangle \equiv \langle k|l\rangle = \delta_{kl} \equiv \begin{cases} 1, & \text{when} \quad k = l, \\ 0, & \text{when} \quad k \neq l, \end{cases} \tag{1.16}$$

---

[35] If ϕ_k belong to continuum they cannot be normalized, but still can be made mutually orthogonal.

where we have given several equivalent notations of the scalar product, which will be used in the present book, and δ_{kl} is known as the Kronecker delta.

- Since eigenfunctions $\{\phi_k\}$ form the complete set, the wave function of the system may be expanded as (M is quite often equal to ∞)

$$\psi = \sum_{k=1}^{M} c_k \phi_k \,, \tag{1.17}$$

where c_k are in general complex coefficients. For ψ and ϕ_k normalized and all ϕ_k chosen to be mutually orthogonal, one gets, after multiplying the equation by[36] ϕ_m^* and integrating,

$$c_k = \langle \phi_k | \psi \rangle \,. \tag{1.18}$$

From the normalization condition for ψ we have[37]

$$\sum_{k=1}^{M} c_k^* c_k = 1. \tag{1.19}$$

> According to the postulate, the probability that the result of the measurement is a_k is equal to $c_k^* c_k$.

If the wave function that describes the state of the system has the form given by (1.17) and does not reduce to a single term $\psi = \phi_k$, then the result of the measurement of the quantity A *cannot* be foreseen. We will measure *some* eigenvalue of the operator $\hat{A}$, but cannot predict which one. After the measurement is completed the wave function of the system represents the eigenstate that corresponds to the measured eigenvalue (known as the *collapse of the wave function*). According to the postulate, the only thing one may say about the measurements is that the *mean value* $\overline{a}$ of the quantity A (from many measurements) is to be compared with the following theoretical result[38] (Fig. 1.9):

---

[36] We have singled out some particular m and therefore a sequence of equalities holds, i.e., $\langle \phi_m | \psi \rangle = \sum_{k=1}^{M} c_k \langle \phi_m | \phi_k \rangle = \sum_{k=1}^{M} c_k \delta_{km} = c_m$ for $m = 1, 2, \ldots M$.

[37] $\langle \psi | \psi \rangle = 1 = \sum_{k=1}^{M} \sum_{l=1}^{M} c_k^* c_l \langle \phi_k | \phi_l \rangle = \sum_{k,l=1} c_k^* c_l \delta_{kl} = \sum_{k=1}^{M} c_k^* c_k$.

[38] $\left\langle \psi | \hat{A} \psi \right\rangle = \left\langle \sum_{l=1}^{M} c_l \phi_l | \hat{A} \sum_{k=1}^{M} c_k \phi_k \right\rangle = \sum_{k=1}^{M} \sum_{l=1}^{M} c_k^* c_l \left\langle \phi_l | \hat{A} \phi_k \right\rangle = \sum_{k=1}^{M} \sum_{l=1}^{M} c_k^* c_l a_k \langle \phi_l | \phi_k \rangle = \sum_{k=1}^{M} \sum_{l=1}^{M} c_k^* c_l a_k \delta_{kl} = \sum_{k=1}^{M} c_k^* c_k a_k$. In case of degeneracy ($a_k = a_l = \ldots$) the probability is $c_k^* c_k + c_l^* c_l + \ldots$. This is how one computes the mean value of anything. Just take all possible distinct results of measurements, multiply each by its probability, and sum up all resulting numbers.

$$\bar{a} = \sum_{k=1}^{M} c_k^* c_k a_k = \frac{\left\langle \psi | \hat{A}\psi \right\rangle}{\langle \psi | \psi \rangle}. \tag{1.20}$$

If we have a special case, $\psi = \phi_k$ (all coefficients $c_l = 0$, except $c_k = 1$), the measured quantity is exactly equal to a_k. From this it follows that, if the wave function is an eigenfunction of operators of several quantities (this happens when the operators commute; see Appendix B), then all these quantities when measured produce with certainty the eigenvalues corresponding to the eigenfunction.

- Eqs. (1.18) and (1.20) show that *the more the eigenfunction* ϕ_k *corresponding to* a_k *resembles the wave function* ψ (a large $|\langle \phi_k | \psi \rangle|$ and therefore also $|\langle \phi_k | \psi \rangle|^2$), the more frequently the eigenvalue a_k will be measured.

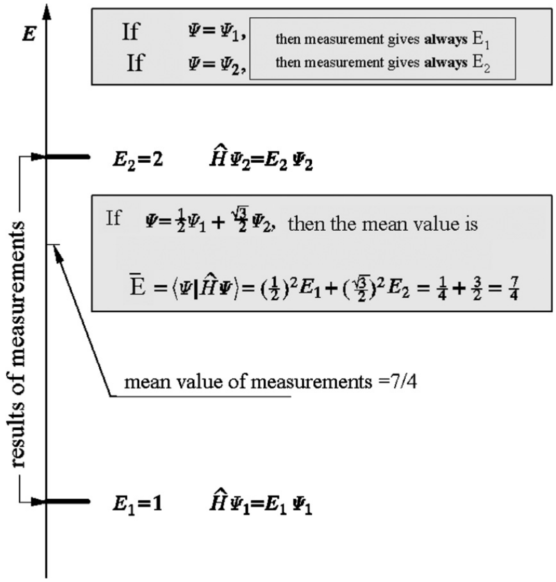


Fig. 1.9. The results of measurements of a quantity A *are the eigenvalues* of the operator $\hat{A}$.

Postulate V (spin angular momentum)

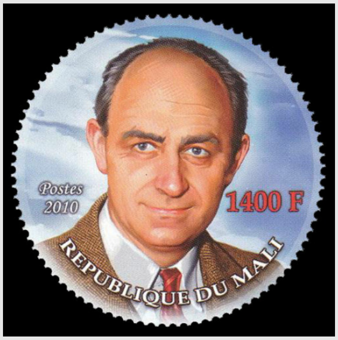
Spin of elementary particles. As will be shown in Chapter 3 (about relativistic effects) spin angular momentum will appear in a natural way. However, in nonrelativistic theory the existence of spin is *postulated*.[39]

---

[39] This has been forced by experimental facts, e.g., energy level splitting in a magnetic field suggested two possible electron states connected to internal angular momentum.

An elementary particle has, besides its orbital angular momentum $\mathbf{r} \times \mathbf{p}$, an internal angular momentum (analogous to that associated with the rotation of a body about its own axis) called spin $\mathbf{S} = (S_x, S_y, S_z)$. Two quantities are measurable: the square of the spin length, $|\mathbf{S}|^2 = S_x^2 + S_y^2 + S_z^2$, and one of its components, by convention, S_z. These quantities only take some particular values $|\mathbf{S}|^2 = s(s+1)\hbar^2$, $S_z = m_s \hbar$, where the spin magnetic quantum number $m_s = -s, -s+1, ..., s$. The spin quantum number s, characteristic of the type of particle (often called simply its spin), can be written as $s = n/2$, where n may be zero or a natural number ("an integer or half-integer" number).

The particles with a half-integer[40] s (e.g., $s = \frac{1}{2}$ for electrons, protons, neutrons, neutrinos) are called *fermions*, the particles with an integer s (e.g., $s = 1$ for deuterons, photons[41]; $s = 0$ for mesons π and mesons K) are called *bosons*.



Enrico Fermi (1901–1954), Italian physicist, professor at universities in Florence, Rome, New York, and in 1941–1946 at the University of Chicago. Fermi introduced the notion of statistics for the particles with a half-integer spin number (called fermions) during the Florence period. Dirac made the same discovery independently, hence this property is called the Fermi–Dirac statistics. Young Fermi was notorious for being able to derive a formula from any domain of physics faster than someone sent to find it in textbooks. His main topic was nuclear physics. He played an important role in the atomic bomb construction at Los Alamos, and in 1942 he built the world's first nuclear reactor on a tennis court at the University of Chicago. Fermi was awarded the Nobel Prize in 1938 "*for his demonstration of the existence of new radioactive elements and for results obtained with them, especially with regard to artificial radioactive elements.*"

---

[40] Note the length of the spin vector for an elementary particle is given by Nature once and for all. Thus, if there is any relation between the spin and the rotation of the particle about its own axis, it has to be a special relation. One cannot change the angular momentum of such a rotation.

[41] The photon represents a particle of zero mass. As a consequence one can show that, instead of three possible m_s one has only two: $m_s = 1, -1$. We call these two possibilities "polarizations" ("parallel" and "perpendicular").

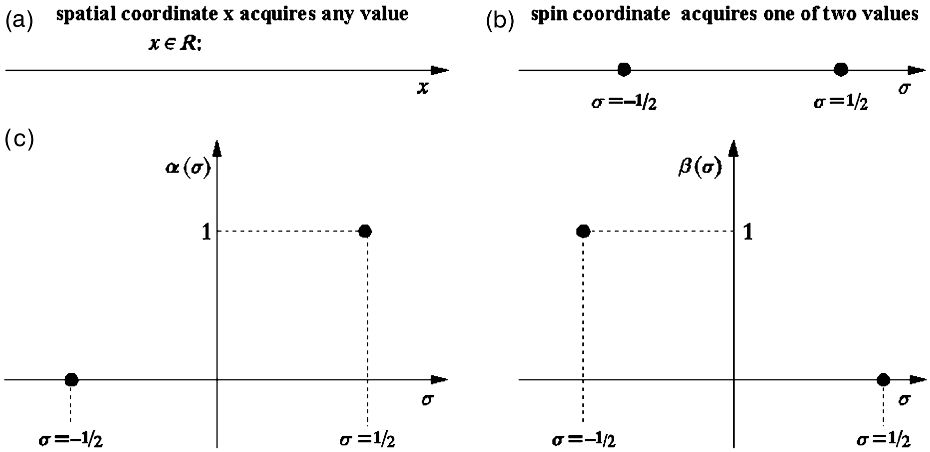


Fig. 1.10. Main differences between the spatial coordinate (x) and spin coordinate (σ) of an electron. (a) The spatial coordinate is *continuous:* it may take any value being a real number. (b) The spin coordinate σ has a granular character (*discrete* values): for $s = \frac{1}{2}$ it can take only one of two values. One of the values is represented by $\sigma = -\frac{1}{2}$, the other by $\sigma = \frac{1}{2}$. In (c), two widely used basis functions, $\alpha(\sigma)$ and $\beta(\sigma)$, are shown in the spin space, respectively.

The magnetic[42] spin quantum number m_s quantizes the z component of the spin angular momentum.

Thus,

> a particle with spin quantum number s has an additional (spin) degree of freedom, or an additional coordinate – *spin coordinate* σ. The spin coordinate differs widely from a spatial coordinate, because it takes only $2s+1$ *discrete values* (Figs. 1.10 and 1.11) associated to $-s, -s+1, \ldots, 0, \ldots, +s$.

Most often one will have to deal with electrons. For electrons, the spin coordinate σ takes two values, often called "up" and "down." We will (arbitrarily) choose $\sigma = -\frac{1}{2}$ and $\sigma = +\frac{1}{2}$.

---

[42] The name is related to energy level splitting in a magnetic field, from which the number is deduced. A nonzero s value is associated to the magnetic dipole, which in magnetic field acquires $2s+1$ energetically nonequivalent positions.

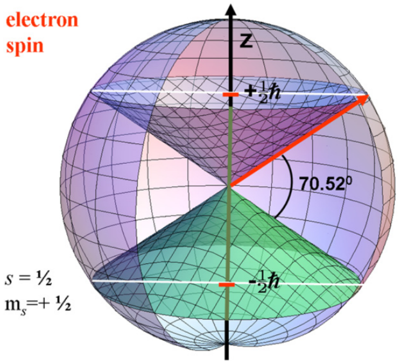


Fig. 1.11. Diagram of the spin angular momentum vector for a particle with spin quantum number $s = \frac{1}{2}$. The only measurable quantities are the spin length $\sqrt{s(s+1)}\hbar = \sqrt{\frac{3}{4}}\hbar$ and the projection of the spin on the quantization axis (chosen as coincident with the perpendicular axis z), which takes only the values $-s, -s+1, \ldots, +s$ in units $\hbar$, i.e., $S_z = -\frac{1}{2}\hbar, \frac{1}{2}\hbar$, which means $m_s = -\frac{1}{2}$ or $m_s = +\frac{1}{2}$. The m_s may be chosen as the spin coordinate σ of the electron, therefore $\sigma \in \left[-\frac{1}{2}, +\frac{1}{2}\right]$. Since the x and y components of the spin remain indefinite, one may visualize the same by locating the spin vector (of constant length $\sqrt{s(s+1)}\hbar$) *anywhere* on a cone surface that ensures a given z component. In the case of $s = \frac{1}{2}$ one has $2s + 1 = 2$ such cones, which are shown in the figure. They form an angle of $70.52° = 70°31'$, so the opening cone angle of each of them is equal to $54.74° = 54°44'$.

Satyendra Nath Bose (1894–1974), Indian physicist, professor at Dakka and Calcutta, first recognized that particles with integer spin number have different statistical properties. Einstein contributed to a more detailed description of this statistics.

According to the postulate (p. 29), the square of the spin length is always the same and equal to $s(s+1)\hbar^2 = \frac{3}{4}\hbar^2$. The maximum projection of a vector on a chosen axis is equal to $\frac{1}{2}\hbar$, while the length of the vector is larger, equal to $\sqrt{s(s+1)}\hbar = \frac{\sqrt{3}}{2}\hbar$. We conclude that the vector of the spin angular momentum makes an angle θ with the quantization axis, with $\cos\theta = \frac{1}{2}/\frac{\sqrt{3}}{2} = \frac{1}{\sqrt{3}}$. From this one obtains[43] $\theta = \arccos\frac{1}{\sqrt{3}} \approx 54°44'$. Fig. 1.11 shows that the spin angular momentum has indefinite x and y components, while always preserving its length and projection on the z axis.

---

[43] In the general case, the spin of a particle may take the following angles with the quantization axis: $\arccos\frac{m_s}{\sqrt{s(s+1)}}$ for $m_s = -s, -s+1, \ldots, +s$.

Spin basis functions for $s = \frac{1}{2}$. One may define (see Fig. 1.10c) the complete set of orthonormal basis functions of the spin space of an electron:

$$\alpha(\sigma) = \begin{cases} 1 & \text{for} \quad \sigma = \frac{1}{2} \\ 0 & \text{for} \quad \sigma = -\frac{1}{2} \end{cases} \qquad \text{and} \quad \beta(\sigma) = \begin{cases} 0 & \text{for} \quad \sigma = \frac{1}{2} \\ 1 & \text{for} \quad \sigma = -\frac{1}{2} \end{cases}$$

or, in a slightly different notation, as orthogonal unit vectors[44]:

$$|\alpha\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |\beta\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Orthogonality follows from $\langle\alpha|\beta\rangle \equiv \sum_\sigma \alpha(\sigma)^*\beta(\sigma) = 0 \cdot 1 + 1 \cdot 0 = 0$. Similarly, normalization means that $\langle\alpha|\alpha\rangle \equiv \sum_\sigma \alpha(\sigma)^*\alpha(\sigma) = \alpha\left(-\frac{1}{2}\right)^* \alpha\left(-\frac{1}{2}\right) + \alpha\left(\frac{1}{2}\right)^* \alpha\left(\frac{1}{2}\right) = 0 \cdot 0 + 1 \cdot 1 = 1$, etc.

We shall now construct operators of the spin angular momentum.

The following definition of spin operators is consistent with the postulate about spin.

$\hat{S}_x = \frac{1}{2}\hbar\sigma_x$,

$\hat{S}_y = \frac{1}{2}\hbar\sigma_y$,

$\hat{S}_z = \frac{1}{2}\hbar\sigma_z$,

where the Pauli matrices of rank 2 are defined as

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Indeed, after applying $\hat{S}_z$ to the spin basis functions one obtains

$$\hat{S}_z|\alpha\rangle \equiv \hat{S}_z \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\hbar|\alpha\rangle,$$

$$\hat{S}_z|\beta\rangle \equiv \hat{S}_z \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}\hbar \begin{pmatrix} 0 \\ -1 \end{pmatrix} = -\frac{1}{2}\hbar|\alpha\rangle.$$

Therefore, functions α and β represent the eigenfunctions of the $\hat{S}_z$ operator with corresponding eigenvalues $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$. How to construct the operator $\hat{\mathbf{S}}^2$? From Pythagoras' theorem, after applying Pauli matrices one obtains

---

[44] In the same spirit as wave functions represent vectors: vector components are values of the function for various values of the variable.

Wolfgang Pauli (1900–1958), German physicist, professor in Hamburg, at the Technical University of Zurich, and at the Institute for Advanced Studies in Princeton (USA), son of a physical chemistry professor at Vienna and a classmate of Werner Heisenberg. At the age of 20 he wrote a famous 200-page article on relativity theory for *Mathematical Encyclopedia*, afterwards edited as a book. A year later Pauli defended his doctoral dissertation under the supervision of Sommerfeld in Munich. The renowned Pauli exclusion principle was proposed in 1924. Wolfgang Pauli received the Nobel Prize in 1945 "*for the discovery of the Exclusion Principle, also called the Pauli Principle.*"

$$
\hat{S}^2 |\alpha\rangle = \hat{S}^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left(\hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2\right) \begin{pmatrix} 1 \\ 0 \end{pmatrix} =
$$

$$
\frac{1}{4}\hbar^2 \left\{ \begin{matrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + \\ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{matrix} \right\} \begin{pmatrix} 1 \\ 0 \end{pmatrix} =
$$

$$
\frac{1}{4}\hbar^2 \begin{pmatrix} 1+1+1 & 0+0+0 \\ 0+0+0 & 1+1+1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{3}{4}\hbar^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}
$$

$$
= \frac{3}{4}\hbar^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \left[\frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2\right] |\alpha\rangle .
$$

The function $|\beta\rangle$ gives an identical result. These are therefore the pure states. One should, however, remember that a particle can be prepared also in a mixed spin state, which is a common procedure in the modern nuclear magnetic resonance technique.

Therefore, both basis functions α and β represent the eigenfunctions of $\hat{S}^2$ and correspond to the same eigenvalue. Thus, the definition of spin operators through Pauli matrices gives results identical to those postulated for S^2 and S_z, and the two formulations are equivalent. From Pauli matrices, it follows that the functions α and β are not eigenfunctions of $\hat{S}_x$ and $\hat{S}_y$ and that the following relations are satisfied[45]:

---

[45] The last three formulae are easy to memorize, since the sequence of the indices is always "rotational," i.e., $x, y, z, x, y, z, \ldots$.

$$
\begin{gathered}
[\hat{S}^2, \hat{S}_z] = 0, \\
[\hat{S}_x, \hat{S}_y] = i\hbar \hat{S}_z, \\
[\hat{S}_y, \hat{S}_z] = i\hbar \hat{S}_x, \\
[\hat{S}_z, \hat{S}_x] = i\hbar \hat{S}_y,
\end{gathered}
$$

which is in agreement with the general properties of angular momenta[46] (see the appendix on p. 665).

The $|\alpha\rangle$ and $|\beta\rangle$ functions form a complete basis set and represent the eigenfunctions of the $\hat{S}^2$ and $\hat{S}_z$ operators. It does not mean however that they are the only spin states possible for the particle. As usually in quantum mechanics the particle (electron, proton, etc.) may exist in a *mixed* spin state

$$\psi = a\,|\alpha\rangle + b\,|\beta\rangle\,, \tag{1.21}$$

with the coefficients a and b satisfying the normalization relation $|a|^2 + |b|^2 = 1$. This state still represents an eigenfunction of $\hat{S}^2$, because $\hat{S}^2\psi = \hat{S}^2(a\,|\alpha\rangle + b\,|\beta\rangle) = a\hat{S}^2\,|\alpha\rangle + b\hat{S}^2\,|\beta\rangle = \left[\frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2\right](a\,|\alpha\rangle + b\,|\beta\rangle) = \left[\frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2\right]\psi$, but is not an eigenfunction of $\hat{S}_z : \hat{S}_z\psi = \hat{S}_z(a\,|\alpha\rangle + b\,|\beta\rangle) = a\hat{S}_z\,|\alpha\rangle + b\hat{S}_z\,|\beta\rangle = a\frac{1}{2}\hbar\,|\alpha\rangle - b\frac{1}{2}\hbar\,|\beta\rangle \neq const\;\psi$. When performing a measurement of S_z for the particle in state ψ each time we will obtain however only *one* of two values: either $\frac{1}{2}\hbar$ (with the probability $|a|^2$) or $-\frac{1}{2}\hbar$ (with the probability $|b|^2$). Thus, after repeating many times the measurement we would get the mean value of S_z. According to Eq. (1.20) this result should agree with the number calculated in quantum mechanics, i.e., $\left\langle\psi|\hat{S}_z\psi\right\rangle = \left\langle a\alpha + b\beta|\hat{S}_z(a\alpha + b\beta)\right\rangle = |a|^2\,\frac{1}{2}\hbar + |a|^2\,(-\frac{1}{2}\hbar) = \frac{1}{2}\hbar(|a|^2 - |b|^2)$.

Spin of nonelementary particles. The postulate on spin pertains to an elementary particle. What about a system composed of such particles? Do they have spin? Spin represents angular mo-

---

[46] Also, note that the mean values of S_x and S_y are both equal to zero in the α and β states, e.g., for the α state one has

$$\left\langle\alpha|\hat{S}_x\alpha\right\rangle = \left\langle\begin{pmatrix}1\\0\end{pmatrix}|\hat{S}_x\begin{pmatrix}1\\0\end{pmatrix}\right\rangle = \frac{1}{2}\hbar\left\langle\begin{pmatrix}1\\0\end{pmatrix}|\begin{pmatrix}0&1\\1&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\right\rangle = \frac{1}{2}\hbar\left\langle\begin{pmatrix}1\\0\end{pmatrix}|\begin{pmatrix}0\\1\end{pmatrix}\right\rangle = 0.$$

This means that in an external vector field (of direction z), when the space is no longer isotropic, only the projection of the total angular momentum on the field direction is conserved. A way to satisfy this is to recall the behavior of a top in a gravitational field. The top rotates about its own axis, but the axis precesses about the field axis. This means that the total electron spin momentum moves on the cone surface making an angle of $54°44'$ with the external field axis in the α state and an angle $180° - 54°44'$ in the β state. Whatever the motion, it must satisfy $\left\langle\alpha|\hat{S}_x\alpha\right\rangle = \left\langle\alpha|\hat{S}_y\alpha\right\rangle = 0$ and $\left\langle\beta|\hat{S}_x\beta\right\rangle = \left\langle\beta|\hat{S}_y\beta\right\rangle = 0$. No more information is available, but one may imagine the motion as a precession just like that of the top.

mentum (a vector) and therefore the angular momentum vectors of the elementary particles have to be added. A system composed of a number of elementary particles (each with its spin $\mathbf{s}_i$) has as an observable the square

$$|\mathbf{S}|^2 = S(S+1)\hbar^2$$

of the total spin vector

$$\mathbf{S} = \mathbf{s}_1 + \mathbf{s}_2 + ...\mathbf{s}_N$$

and one of the components of $\mathbf{S}$ (denoted by S_z):

$$S_z = M_S\hbar, \quad \text{for } M_S = -S, -S+1, ..., S,$$

where the number S stands, just as for a single particle, for an integer or half-integer nonnegative number. Particular values of S (often called simply the spin) and of the spin magnetic number M_S depend on the directions of vectors $\mathbf{s}_i$. It follows that no excitation of a nonelementary boson (that causes another summing of the individual spin vectors) can change the particle to a fermion and *vice versa*. Systems with an even number of fermions are always bosons, while those with an odd number of fermions are always fermions.

Nuclei. The ground states of the important *nuclei* ^{12}C and ^{16}O correspond to $S = 0$, while those of ^{13}C, ^{15}N, and ^{19}F have $S = \frac{1}{2}$.

Atoms and molecules. Does an atom as a whole represent a fermion or a boson? This depends on which atom and which molecule one considers. Consider the hydrogen atom, composed of two fermions (proton and electron, both with spin number $\frac{1}{2}$). This is sufficient to know that one deals with a boson. For similar reasons, the sodium atom with 23 nucleons (each of spin $\frac{1}{2}$) in the nucleus and 11 electrons moving around it also represents a boson.

Let us consider two electrons and add together two electron spin vectors, $\mathbf{s}_1 + \mathbf{s}_2$. Then the maximum z component of the spin angular momentum will be (in $\hbar$ units) $|M_S| = m_{s1} + m_{s2} = \frac{1}{2} + \frac{1}{2} = 1$. This corresponds to the vectors $\mathbf{s}_1$, $\mathbf{s}_2$, called "parallel" to each other, while the minimum $|M_S| = m_{s1} + m_{s2} = \frac{1}{2} - \frac{1}{2} = 0$ means an "antiparallel" configuration of $\mathbf{s}_1$ and $\mathbf{s}_2$ (Fig. 1.12).

The first situation indicates that for the state with parallel spins $S = 1$, and for this S the possible $M_S = 1, 0, -1$. This means there are three states: $(S, M_S) = (1, 1), (1, 0), (1, -1)$. If no

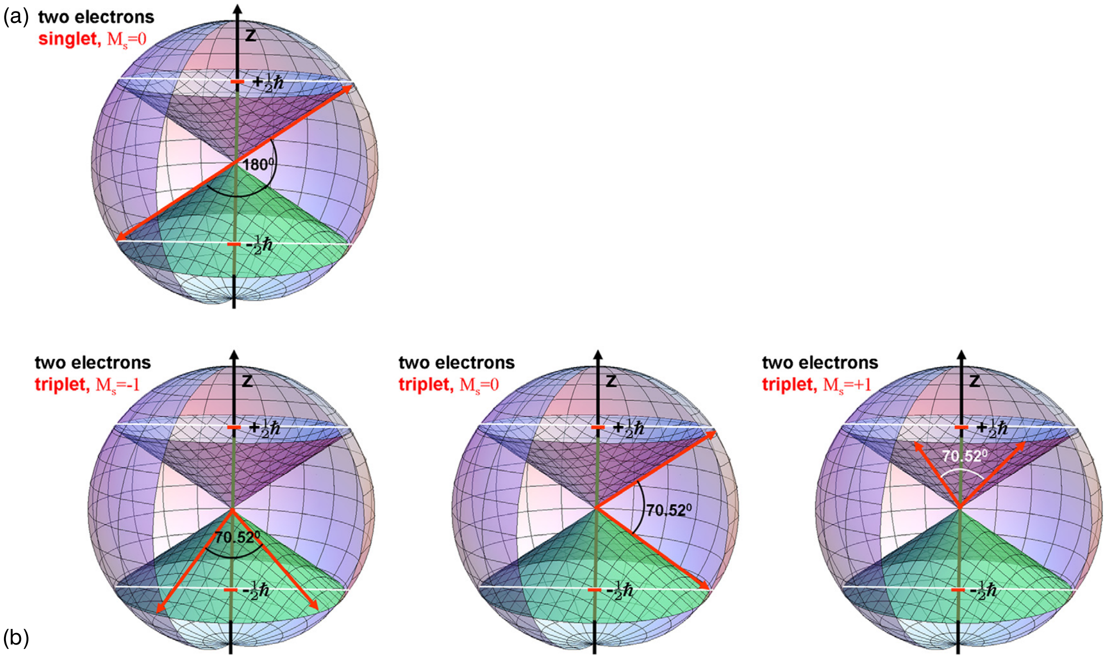


Fig. 1.12. Spin angular momentum for a system with two electrons (in general, particles with $s = \frac{1}{2}$). The quantization axis is arbitrarily chosen as the vertical axis z. Then, the spin vectors of the individual electrons (see Fig. 1.11) may be thought to reside somewhere on the upper cone that corresponds to $m_s = \frac{1}{2}$, or on the lower cone corresponding to $m_s = -\frac{1}{2}$. For two electrons there are two spin eigenstates of $\hat{\mathbf{S}}^2$. One has total spin quantum number $S = 0$ (singlet state); the other is triply degenerate (triplet state), and the three components of the state have $S = 1$ and $S_z = 1, 0, -1$ in $\hbar$ units. (a) In the singlet state the vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ remain on the cones of different orientation, and have opposite ("*antiparallel*") orientations, so that $\mathbf{s}_1 + \mathbf{s}_2 = \mathbf{0}$. Although their exact positions on the cones are undetermined (and moreover the cones themselves follow from the arbitrary choice of the quantization axis in space), they are always pointing in opposite directions. (b) The three triplet components, corresponding to $S = 1$, differ by the direction of the total spin angular momentum (of constant length $\sqrt{S(S+1)}\hbar = \sqrt{2}\hbar$). The three directions correspond to three projections $M_S\hbar$ of spin momentum, i.e., $\hbar, -\hbar, 0$. In each of the three cases the angle between the two spins equals $\omega = 70.52^\circ = 70^\circ 31'$ (although in textbooks – including this one – they are said to be "*parallel*," in fact they are not; see the text). In case of $M_S = 0$ the z axis and the two spins are in the same plane.

direction in space is privileged, then all the three states correspond to the same energy (triple degeneracy). This is why such a set of three states is called a *triplet state*. The second situation witnesses the existence of a state with $S = 0$, which obviously corresponds to $M_S = 0$. This state is called a *singlet state*.

Let us calculate the angle ω between the individual electronic spins:

$$|\mathbf{S}|^2 = (\mathbf{s}_1 + \mathbf{s}_2)^2 = s_1^2 + s_2^2 + 2\mathbf{s}_1 \cdot \mathbf{s}_2 = s_1^2 + s_2^2 + 2s_1 \cdot s_2 \cos\omega =$$

$$\frac{1}{2}\left(\frac{1}{2}+1\right)\hbar^2 \cdot 2 + 2\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)}\sqrt{\frac{1}{2}\left(\frac{1}{2}+1\right)}\hbar^2 \cos\omega =$$

$$\left(\frac{3}{2}+\frac{3}{2}\cos\omega\right)\hbar^2 = \frac{3}{2}\left(1+\cos\omega\right)\hbar^2.$$

SINGLET AND TRIPLET STATES

For the singlet state $|\mathbf{S}|^2 = S(S+1)\hbar^2 = 0$, hence $1 + \cos\omega = 0$ and $\omega = 180°$. *This means the two electronic spins in the singlet state are antiparallel.*

For the triplet state $|\mathbf{S}|^2 = S(S+1)\hbar^2 = 2\hbar^2$, and hence $\frac{3}{2}(1+\cos\omega)\hbar^2 = 2\hbar^2$, i.e., $\cos\omega = \frac{1}{3}$, or $\omega = 70.52°$; see Fig. 1.12. *Despite forming the angle $\omega = 70.52°$ the two spins in the triplet state are often said to be "parallel."*

The two electrons which we have considered may, for example, be part of a hydrogen molecule. Therefore, when considering electronic states, we may have to deal with singlets or triplets. However, in the same hydrogen molecule we have two protons, whose spins may also be "parallel" (orthohydrogen) or antiparallel (parahydrogen). In parahydrogen the nuclear spin is $S = 0$, while in orthohydrogen $S = 1$. In consequence, there is only one state for parahydrogen ($M_S = 0$), and three states for orthohydrogen ($M_S = 1, 0, -1$).[47]

Postulate VI (on the permutational symmetry)

Unlike classical mechanics, quantum mechanics is radical: it requires that two particles of the same kind (two electrons, two protons, etc.) should play the same role in the system, and therefore in its description enshrined in the wave function.[48] Quantum mechanics *guarantees* that the roles played in the Hamiltonian by two identical particles are identical. Within this philosophy, exchange of the labels of two identical particles (i.e., the exchange of their coordinates $x_1, y_1, z_1, \sigma_1 \leftrightarrow x_2, y_2, z_2, \sigma_2$, or in short, $1 \leftrightarrow 2$) leads, at most, to a change of the phase ϕ of the wave function: $\psi(2,1) \rightarrow e^{i\phi}\psi(1,2)$, because in such a case $|\psi(2,1)| = |\psi(1,2)|$ (this guarantees equal probabilities of both situations). However, when we exchange the two labels

---

[47] Since all the states have very similar energy (and therefore at high temperatures the Boltzmann factors are practically the same), there are three times as many molecules of orthohydrogen as of parahydrogen. Both states (ortho and para) differ slightly in their physicochemical characteristics.

[48] Everyday experience tells us the opposite, e.g., a car accident involving a Mercedes does not cause all copies of that particular model to have identical crash traces.

once more, we have to return to the initial situation: $\psi(1,2) = e^{i\phi}\psi(2,1) = e^{i\phi}e^{i\phi}\psi(1,2) = \left(e^{i\phi}\right)^2\psi(1,2)$. Hence, $\left(e^{i\phi}\right)^2 = 1$, i.e., $e^{i\phi} = \pm 1$. *Postulate VI says that $e^{i\phi} = +1$ refers to bosons, while $e^{i\phi} = -1$ refers to fermions.*

> The wave function ψ which describes identical *bosons* (i.e., spin integer particles) $1, 2, 3, ..., N$ has to be *symmetric* with respect to the exchange of coordinates x_i, y_i, z_i, σ_i and x_j, y_j, z_j, σ_j, i.e., if $x_i \leftrightarrow x_j$, $y_i \leftrightarrow y_j$, $z_i \leftrightarrow z_j$, $\sigma_i \leftrightarrow \sigma_j$, then $\psi(1,2,...,i,...,j,...,N) = \psi(1,2,...,j,...,i,...,N)$. If particles i and j denote identical *fermions*, the wave function must be *antisymmetric*, i.e., $\psi(1,2,...,i,...,j,...,N) = -\psi(1,2,...,j,...,i,...,N)$.

Let us see the probability density that two fermions (particles 1 and 2) occupy the same position in space and, additionally, that they have the same spin coordinate $(x_1, y_1, z_1, \sigma_1) = (x_2, y_2, z_2, \sigma_2)$. We have $\psi(1,1,3,4,...,N) = -\psi(1,1,3,4,...,N)$, hence $\psi(1,1,3,4,...,N) = 0$ and, of course, $|\psi(1,1,3,4,...,N)|^2 = 0$. We conclude two electrons of the same spin coordinate (we will sometimes say "of the same spin") avoid each other. This is called the exchange or Fermi hole around each electron.[49] The reason for the hole is the

---

[49] Electrons represent fermions ($s = \frac{1}{2}, \sigma = +\frac{1}{2}, -\frac{1}{2}$) and therefore there are two forms of repulsion among them: a Coulomb one because of their electric charge, and the one between same spin electrons only, which follows from the Pauli exclusion principle. As will be shown in Chapter V2-3, the Coulomb repulsion is much less important than the effect of the Pauli exclusion principle.

Positions of the nuclei determine which parts of the space are preferred by electrons (just because of the electron–nucleus Coulomb attraction): the vicinity of the nuclei. However, considering one such part (a particular nucleus), the probability of finding there electrons of the same spin coordinate is negligible (from the Pauli exclusion principle together with the continuity condition for the wave function). However, the Pauli exclusion principle does not pertain to the electrons of opposite spin coordinates; therefore *two such electrons can occupy the same small volume* even despite the energy increase because of their Coulomb repulsion. Hence, we have a compromise of two opposite effects: the electronic pair is *attracted* by a nucleus (proportionally to the nuclear charge), but the interelectronic distance should not be too small, because of the electron–electron Coulomb *repulsion*. A third electron seeking its low energy in the vicinity of this nucleus has no chance: because its spin coordinate is necessarily the same as that of one of the electrons. The *third electron therefore has to leave this area, even at the expense of a large increase in energy*, and find another low (Coulomb) energy region, maybe together with his colleague of the opposite spin (thus forming another electron pair). This picture does not mean one is able to discern the electrons – they play the same role in the system; this is guaranteed by the antisymmetry of the wave function. Therefore, we will not know *which* electrons form a particular electronic pair, but only that there are *two* of them in it and they have the *opposite spin* coordinates.

Let us turn to molecules. For electrons, which are very weakly bound (valence electrons) a space with Coulomb attraction of *two* nuclei might be a good choice (we will see this in Chapter 8). This space may also be shared by two electrons of the opposite spins, since this is still better than to send the partner elsewhere in space.

Thus, already at this stage we foresee a fundamental role of electron pairs (the opposite spin coordinates within the pair) leading to the electronic shell structure in atoms and molecules and to chemical bonds in molecules.

antisymmetry of the electronic wave function, or in other words, the Pauli exclusion principle.[50]

> Thus, the probability density of finding two identical fermions *in the same position* and *with the same spin coordinate* is equal to zero. There is no such restriction for two identical bosons or two identical fermions with *different* spin coordinates. They can be at the same point in space.

This conclusion sounds incredible, especially for the atomic or molecular bosons, and is related to what is known as Bose–Einstein condensation, which represents however an experimental fact.[51] Does it mean that we may pack, say, 50 000 bosonic atoms in a single point? It seems like saying that one may place 50 000 couples in a single tent without any problem whatsoever. In principle you can, but you are certainly asking for troubles.

In the Bose–Einstein condensate the bosons (alkali metal atoms) are in the same place, but in a peculiar sense. The total wave function for the bosons is, to a first approximation, a product of *identical* nodeless wave functions for the particular bosons (this ensures proper symmetry). Each of the wave functions extends considerably in space (the Bose–Einstein condensate is as large as a fraction of a millimeter), but all have been *centered in the same point in space.* An exact wave function is not product-like and the individual bosons will avoid each other within the Bose–Einstein condensate at least because they are composed of identical fermions, and the latter ones have to obey the Pauli exclusion principle.

* * *

Among the above postulates, the strongest controversy has always been associated with Postulate IV, which says that, except some special cases, one cannot predict the result of a particular single measurement, but only its probability. More advanced considerations devoted to Postulate IV lead to the conclusion that there is no way (neither experimental protocol nor theoretical reasoning) to predict when and in which direction an excited atom will emit a photon. This means that quantum mechanics is not a deterministic theory.

---

[50] The Pauli exclusion principle is sometimes formulated in another way: two electrons cannot be in the same state (including spin). The connection of this strange phrasing (what does electron state mean?) with the above will become clear in Chapter 8.

[51] Shown to happen by Eric A. Cornell, Carl E. Wieman, and Wolfgang Ketterle (Nobel Prize 2001 "*for discovering a new state of matter*").

The indeterminism appears however only in the physical space, while in the space of all states (Hilbert space) everything is perfectly deterministic. The wave function evolves in a deterministic way according to the time-dependent Schrödinger equation (1.10).

The puzzling way in which indeterminism operates will be shown below.

1.3 The Heisenberg uncertainty principle

Consider two mechanical quantities A and B, for which the corresponding Hermitian operators (constructed according to Postulate II), $\hat{A}$ and $\hat{B}$, give the commutator $\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A} = i\hat{C}$, where $\hat{C}$ is a Hermitian operator.[52] This is what happens for example for $A = x$ and $B = p_x$. Indeed, for any differentiable function ϕ one has $[\hat{x}, \hat{p_x}]\phi = -xi\hbar\phi' + i\hbar(x\phi)' = i\hbar\phi$, and therefore the operator $\hat{C}$ in this case means simply multiplication by $\hbar$.

From axioms of quantum mechanics one can prove that a product of errors (in the sense of standard deviation) of measurements of two mechanical quantities is greater than or equal to $\frac{1}{2}\left\langle [\hat{A}, \hat{B}] \right\rangle$, where $\left\langle [\hat{A}, \hat{B}] \right\rangle$ is the mean value of the commutator $[\hat{A}, \hat{B}]$.

This is known as the Heisenberg uncertainty principle.

Werner Heisenberg did not provide any formal proof; instead he carried out a *Gedankenexperiment* (an ideal thought experiment) with an electron interacting with an electromagnetic wave ("Heisenberg's microscope").

The formal proof goes as follows.

---

[52] This is guaranteed. Indeed, $\hat{C} = -i\left[\hat{A}, \hat{B}\right]$ and then the Hermitian character of $\hat{C}$ is shown by the following chain of transformations: $\left\langle f|\hat{C}g \right\rangle = -i\left\langle f|\left[\hat{A}, \hat{B}\right]g \right\rangle = -i\left\langle f|(\hat{A}\hat{B} - \hat{B}\hat{A})g \right\rangle = -i\left\langle (\hat{B}\hat{A} - \hat{A}\hat{B})f|g \right\rangle = \left\langle -i(\hat{A}\hat{B} - \hat{B}\hat{A})f|g \right\rangle = \left\langle \hat{C}f|g \right\rangle$.

Werner Karl Heisenberg (1901–1976) was born in Würzburg (Germany), attended high school in Munich, and then (with his friend Wolfgang Pauli) studied physics at Munich University under Sommerfeld's supervision. In 1923 he defended his doctoral thesis on turbulence in liquids. Reportedly, during the doctoral examination he had problems writing down the chemical reaction in lead batteries. He joined the laboratory of Max Born at Göttingen (following his friend Wolfgang) and in 1924 the Institute of Theoretical Physics in Copenhagen, working under the supervision of Niels Bohr. A lecture delivered by Niels Bohr decided the future direction of his work. Heisenberg later wrote: "*I was taught optimism by Sommerfeld, mathematics in Göttingen, physics by Bohr*." In 1925 (only a year after being convinced by Bohr) Heisenberg developed a formalism, which became the first successful quantum theory. Then, in 1926 Heisenberg, Born, and Jordan elaborated the formalism, which resulted in a coherent theory ("matrix mechanics"). In 1927 Heisenberg obtained a chair at Leipzig University, which he held until 1941 (when he became director of the Kaiser Wilhelm Physics Institute in Berlin). Heisenberg received the Nobel Prize in 1932 "*for the creation of quantum mechanics, the application of which has, inter alia, led to the discovery of the allotropic forms of hydrogen*." In 1937 Werner Heisenberg was at the height of his powers. He was nominated professor and got married. However, just after returning from his honeymoon, the rector of the university called him, saying that there was a problem. In the SS weekly, an article by Professor Johannes Stark (a Nobel Prize winner and faithful Nazi) was about to appear claiming that Professor Heisenberg is not such a good patriot as he pretends, because he socialized in the past with Jewish physicists. Soon Professor Heisenberg was invited to the SS headquarters at Prinz Albert Strasse in Berlin. The interrogation took place in the basement. On the raw concrete wall the scoffing slogan "*Breathe deeply and quietly*" was written. One of the questioners was a PhD student from Leipzig, who had once been examined by Heisenberg. The terrified Heisenberg told his mother about the problem. She recalled that in her youth she had made the acquaintance of Heinrich Himmler's mother. Frau Heisenberg paid a visit to Frau Himmler and asked her to pass a letter from her son to Himmler. At the beginning Himmler's mother tried to separate her maternal feelings for her beloved son from politics. She was finally convinced after Frau Heisenberg said "*we mothers should care about our boys*." After a certain time, Heisenberg received a letter from Himmler saying that his letter "coming through unusual channels" has been examined especially carefully. He promised to stop the attack. In the *post scriptum* there was a precisely tailored phrase: "*I think it best for your future, if for the benefit of your students, you would carefully separate scientific achievements from the personal and political beliefs of those who carried them out. Your faithfully, Heinrich Himmler*" (after D. Bodanis, "$E = mc^2$," Fakty, Warsaw, 2001, p. 130) ©The Nobel Foundation.

Recall the definition of the variance, or the square of the standard deviation $(\Delta A)^2$, of measurements of the quantity A:

$$(\Delta A)^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2, \tag{1.22}$$

where $\langle X \rangle$ means the mean value of many measurements of the quantity X. The *standard deviation* ΔA *represents the width of the distribution of* A, i.e., *the error made.* Eq. (1.22) is equivalent to[53]

$$(\Delta A)^2 = \langle (\hat{A} - \langle \hat{A} \rangle)^2 \rangle. \tag{1.23}$$

Consider the product of the standard deviations for the operators $\hat{A}$ and $\hat{B}$, taking into account that $\langle \hat{u} \rangle$ denotes (Postulate IV) the integral $\langle \Psi | \hat{u} | \Psi \rangle$ according to (1.20). One obtains (denoting $\hat{\mathcal{A}} = \hat{A} - \langle \hat{A} \rangle$ and $\hat{\mathcal{B}} = \hat{B} - \langle \hat{B} \rangle$; of course, $[\hat{\mathcal{A}}, \hat{\mathcal{B}}] = [\hat{A}, \hat{B}]$)

$$(\Delta A)^2 \cdot (\Delta B)^2 = \langle \Psi | \hat{\mathcal{A}}^2 \Psi \rangle \langle \Psi | \hat{\mathcal{B}}^2 \Psi \rangle = \langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{A}} \Psi \rangle \langle \hat{\mathcal{B}} \Psi | \hat{\mathcal{B}} \Psi \rangle,$$

where the Hermitian character of the operators $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$ is used. Now, let us use the Schwarz inequality (Appendix B) $\langle f_1 | f_1 \rangle \langle f_2 | f_2 \rangle \geq |\langle f_1 | f_2 \rangle|^2$:

$$(\Delta A)^2 \cdot (\Delta B)^2 = \langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{A}} \Psi \rangle \langle \hat{\mathcal{B}} \Psi | \hat{\mathcal{B}} \Psi \rangle \geq |\langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{B}} \Psi \rangle|^2.$$

Next,

$$\begin{aligned} \langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{B}} \Psi \rangle &= \langle \Psi | \hat{\mathcal{A}} \hat{\mathcal{B}} \Psi \rangle = \langle \Psi | \{ [\hat{\mathcal{A}}, \hat{\mathcal{B}}] + \hat{\mathcal{B}} \hat{\mathcal{A}} \} \Psi \rangle = i \langle \Psi | \hat{C} \Psi \rangle + \langle \Psi | \hat{\mathcal{B}} \hat{\mathcal{A}} \Psi \rangle \\ &= i \langle \Psi | \hat{C} \Psi \rangle + \langle \hat{\mathcal{B}} \Psi | \hat{\mathcal{A}} \Psi \rangle = i \langle \Psi | \hat{C} \Psi \rangle + \langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{B}} \Psi \rangle^*. \end{aligned}$$

Hence,

$$i \langle \Psi | \hat{C} \Psi \rangle = 2i \ \mathrm{Im}\{ \langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{B}} \Psi \rangle \}.$$

This means that $\mathrm{Im}\{ \langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{B}} \Psi \rangle \} = \frac{\langle \Psi | \hat{C} \Psi \rangle}{2}$, which gives $|\langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{B}} \Psi \rangle| \geq \frac{|\langle \Psi | \hat{C} \Psi \rangle|}{2}$. Hence,

$$(\Delta A)^2 \cdot (\Delta B)^2 \geq |\langle \hat{\mathcal{A}} \Psi | \hat{\mathcal{B}} \Psi \rangle|^2 \geq \frac{|\langle \Psi | \hat{C} \Psi \rangle|^2}{4} \tag{1.24}$$

or, taking into account that $|\langle \Psi | \hat{C} \Psi \rangle| = |\langle \Psi | [\hat{A}, \hat{B}] \Psi \rangle|$, we have[54]

$$\Delta A \cdot \Delta B \geq \frac{1}{2} |\langle \Psi | [\hat{A}, \hat{B}] \Psi \rangle|. \tag{1.25}$$

---

[53] Because $\langle (\hat{A} - \langle \hat{A} \rangle)^2 \rangle = \langle \hat{A}^2 - 2\hat{A}\langle \hat{A} \rangle + \langle \hat{A} \rangle^2 \rangle = \langle \hat{A}^2 \rangle - 2\langle \hat{A} \rangle^2 + \langle \hat{A} \rangle^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2$.

[54] This particular result was used later by Mandelshtam and Tamm to derive the energy-time uncertainty – an analogy to the Heisenberg uncertainty principle; see p. 100.

There are two important special cases.

(a) $\hat{C} = 0$, i.e., the operators $\hat{A}$ and $\hat{B}$ commute. We have $\Delta A \cdot \Delta B \geq 0$, i.e., *the errors can be arbitrarily small*. Both quantities therefore *can be* measured simultaneously without error.

(b) $\hat{C} = \hbar$, as in the case of $\hat{x}$ and $\hat{p}_x$. Then, $(\Delta A) \cdot (\Delta B) \geq \frac{\hbar}{2}$.

In particular, for $\hat{A} = \hat{x}$ and $\hat{B} = \hat{p}_x$, if quantum mechanics is valid, *one cannot determine the exact position and the exact momentum of a particle*. The sense of this statement is the following: even if it *may happen* that we get as a result of a measurement the exact position and the exact momentum of the particle, by repeating many times the measurement (in order to be sure about the error made) we are unable to get zero standard deviation (error) for position and energy measurements. When the precision with which x is measured increases, the particle's momentum has so wide a distribution that the error of determining p_x is huge.[55] Fig. 1.13 shows why the Heisenberg principle does not represent any mystery, but is a natural consequence of the wave nature of particles. Also, the principle does not rely on any putative perturbation of position or momentum of the particle during measurement.

The power of the Heisenberg uncertainty principle is seen when it is used for estimation, just from scratch, of the size of some systems.

Example 1 (Size of the hydrogen atom). How on earth this can be estimated from virtually no information? Let us see.

We assume that the electron moves, while the nucleus does not.[56] Whatever the electron does in the hydrogen atom, it has to conform to the uncertainty principle $\Delta x \cdot \Delta p_x \geq \frac{\hbar}{2}$, i.e., in the most compact state we may expect the *equality* $\Delta x \cdot \Delta p_x = \frac{\hbar}{2}$. We may estimate Δx as the radius r of the atom, while $\Delta p_x = \sqrt{\langle p_x^2 \rangle - \langle p_x \rangle^2} = \sqrt{\langle p_x^2 \rangle - 0} = \sqrt{\langle p_x^2 \rangle}$. Therefore, we have an estimation $r \cdot \sqrt{\langle p_x^2 \rangle} = \frac{\hbar}{2}$, or $\sqrt{\langle p_x^2 \rangle} = \frac{\hbar}{2r}$. The total energy may be estimated as the sum of the kinetic and potential energies: $E = \frac{\langle p^2 \rangle}{2m} - \frac{e^2}{r} = \frac{\langle p_x^2 + p_y^2 + p_z^2 \rangle}{2m} - \frac{e^2}{r} = \frac{3\hbar^2}{8mr^2} - \frac{e^2}{r}$. Now, let us find the minimum of $E(r)$ as its probable value: $\frac{\mathrm{d}E}{\mathrm{d}r} = 0 = -2\frac{3\hbar^2}{8mr^3} + \frac{e^2}{r^2}$, or $e^2 = \frac{3}{4}\frac{\hbar^2}{mr}$.

---

[55] There is an apocryphal story about a *Polizei* patrol stopping Professor Heisenberg for speeding. The very serious man asks: "*Do you know how fast you were going when I stopped you?*" Professor Heisenberg answered: "*I have absolutely no idea, Herr Oberleutnant, but I can tell you precisely where you stopped me*."

[56] Note that by the momentum conservation law the nucleus moves 1840 times slower than the electron. This practically means the electron moves in the electric field of the immobilized nucleus.

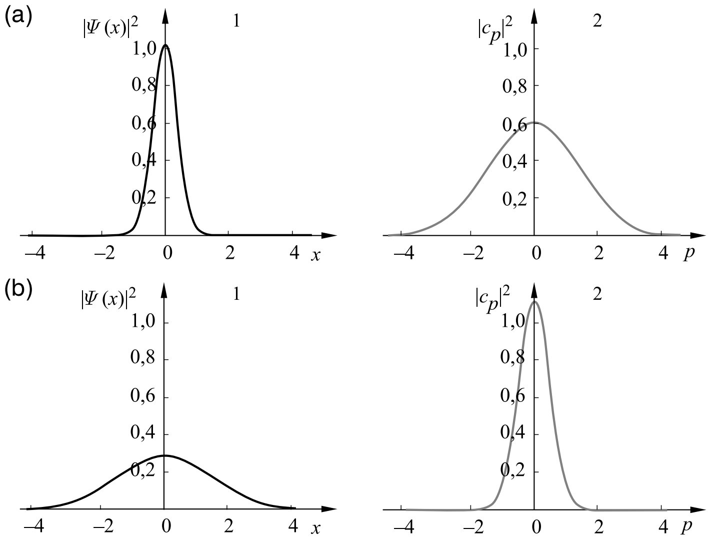


Fig. 1.13. Illustration of the Heisenberg uncertainty principle in case of a single particle moving on the x axis. (a1) $|\Psi(x)|^2$ (i.e., the probability density distribution of finding the particle) as function of coordinate x. The width of this distribution is related to the expected error of determining x. The curve is narrow, which means the error in determining the particle's position is expected to be small. The wave function $\Psi(x)$ can be expanded in the infinite series $\Psi(x) = \sum_p c_p \exp(ipx)$, where p denotes the momentum. Note that each individual function $\exp(ipx)$ is an eigenfunction of momentum, and therefore, if $\Psi(x) = \exp(ipx)$, a measurement of momentum would give exactly p. If, however, $\Psi(x) = \sum_p c_p \exp(ipx)$, then such a measurement yields a given p with the probability $|c_p|^2$. (a2) $|c_p|^2$ as function of p. As one can see a broad range of p (large uncertainty of momentum) ensures a sharp $|\Psi(x)|^2$ distribution (small uncertainty of position in (a1)). Simply the waves $\exp(ipx)$ to obtain a sharp peak of $\Psi(x)$ should exhibit a perfect constructive interference in a small region and a perfect destructive interference elsewhere. This requires a lot of different p's, i.e., a broad momentum distribution. This means that a small error in determining the particle's position is necessarily associated with a large error in determining the particle's momentum (Heisenberg uncertainty principle). (b1, b2) The same is shown, but this time a narrow p distribution gives a broad x distribution. The Heisenberg uncertainty principle has nothing mysterious in it and does not represent a unique feature of quantum mechanics. If one wanted to construct from the ocean waves (of various directions and wave lengths) a tall water pole in the middle of the still ocean, one would be forced to use many wave lengths to get the constructive interference in one spot and destructive interference in all the others.

Hence, we have an estimation, $r = \frac{3}{4}\frac{\hbar^2}{me^2} = 0.75 \cdot a_0$, where $a_0 = 0.529$ Å (as will be shown in Chapter 4, Eq. (4.41), p. 232) is known as the "Bohr first orbit radius of the hydrogen atom." Thus, just from the Heisenberg uncertainty principle we got a value, which has the correct order of magnitude!

We have an important result.

The Heisenberg uncertainty principle prevents atoms from collapsing (the electron falling down to the nucleus, $x = 0$, $\Delta x = 0$, its energy would be equal to $-\infty$), which would mean a catastrophe in the Universe (emission of an infinite amount of energy). The principle also gives a correct order of magnitude for the density of matter.

Example 2 (Size of nucleus). How large is an atomic nucleus? Well, again we may estimate its size from the Heisenberg uncertainty principle knowing only one thing: that the binding energy per nucleon is of the order of 8 MeV. We are now in the realm of nuclear forces acting among nucleons. It seems therefore our task is extremely difficult. We just ignore all these extremely complex forces, considering instead *a single nucleon moving in a mean potential, bound with the energy* $E = 8$ MeV, *and therefore its kinetic energy has to be of the same order of magnitude*. We focus on this kinetic energy[57] of the nucleon assuming simply that $E = \frac{\langle p^2 \rangle}{2m_N} = \frac{\langle (\Delta p)^2 \rangle}{2m_N}$. From this we calculate $\sqrt{\langle (\Delta p)^2 \rangle} = \Delta p = \sqrt{2m_N E}$ and, using the Heisenberg uncertainty principle, we have the uncertainty of the nucleon's position as $a = \frac{\hbar}{2\Delta p} = \frac{\hbar}{2\sqrt{2m_N E}}$. Calculating this in atomic units ($\hbar = 1$, $e = 1$, $m = 1$, where m is the electron mass) we get $a = \frac{\hbar}{2\sqrt{2m_N E}} = \frac{1}{2\sqrt{2 \cdot 1840 \cdot 8 \cdot 10^6 \cdot 3.67516 \cdot 10^{-2}}} = 0.0000152$ a.u. $\approx 10^{-13}$ cm, 100,000 times smaller than the hydrogen atom. Experiments confirm our number: this is indeed the order of magnitude of the sizes of atomic nuclei!

1.4 The Copenhagen interpretation of the world[58]

The picture of the world that emerged from quantum mechanics was "diffuse" with respect to classical mechanics. In classical mechanics one could measure a particle's position and momentum with a desired accuracy,[59] whereas the Heisenberg uncertainty principle states that this is simply *impossible*.

Bohr presented a philosophical interpretation of the world, which at its foundation had in a sense a *nonreality of the world*.

---

[57] We are interested in the order of its magnitude only.

[58] Schrödinger did not like the Copenhagen interpretation. Once, Bohr and Heisenberg invited him for a Baltic Sea cruise and tried to indoctrinate him so strongly that Schrödinger became ill and stopped participating in their discussions.

[59] This is an exaggeration. Classical mechanics also has its own problems with uncertainty. For example, obtaining the same results for a game of dice would require a perfect reproduction of the initial conditions, which is never feasible.

According to Bohr, before a measurement on a particle is made *nothing* can be said about the value of a given mechanical quantity, unless the wave function represents an eigenfunction of the operator of this mechanical quantity. Moreover, except this case, the particle *does not have* any fixed value of mechanical quantity at all.

A measurement gives a value of the mechanical property (A). Then, according to Bohr, after the measurement is completed, the state of the system changes (the so-called *wave function collapse* or *decoherence*) to the state described by *an eigenfunction of the corresponding operator* $\hat{A}$, and as the *measured value one obtains the eigenvalue corresponding to the wave function*. According to Bohr, there is no way to foresee which eigenvalue one will get as the result of the measurement. However, one can calculate the *probability* of getting a particular eigenvalue. This probability may be computed as the square of the overlap integral (cf. p. 29) of the initial wave function and the eigenfunction of $\hat{A}$.

1.5 How to disprove the Heisenberg principle? The Einstein–Podolsky–Rosen's recipe

The Heisenberg uncertainty principle came as a shock. Many scientists felt a strong imperative to prove that the principle is false. One of them was Albert Einstein, who used to play with ideas by performing some (as he used to say) ideal thought experiments (in German *Gedankenexperiment*) in order to demonstrate internal contradictions in theories. Einstein believed in the reality of our world. With his colleagues Podolsky and Rosen ("EPR team") he designed a special *Gedankenexperiment*.[60] It represented an attempt to disprove the Heisenberg uncertainty principle and to show that one *can* measure the position and momentum of a particle *without any error*. To achieve this, the gentlemen invoked the help of a second particle.

The key statement of the whole reasoning, given in the EPR paper, was the following: "*If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity*." In other words, the quantity to be measured *has* a particular, albeit maybe unknown, value. EPR considered a coordinate system fixed in space and two particles: particle 1 with coordinate x_1 and momentum p_{x1} and particle 2 with coordinate x_2 and momentum p_{x2}, the total system being in a well-defined total momentum, $P = p_{x1} + p_{x2}$, and a well-defined relative position, $x = x_1 - x_2$. The meaning of the words "well-defined" is that, according to quantum mechanics, there is a possibility of the *exact measurement* of the

---

[60] A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.*, 47(1935)777.

two quantities (x and P), because the two operators $\hat{x}$ and $\hat{P}$ do commute.[61] At this point, Einstein and his colleagues and the great interpreters of quantum theory agreed.

We now come to the crux of the real controversy.

The particles interact, then separate and fly far away (at any time we are able to measure exactly both x and P). When they are extremely far from each other (e.g., one close to us, the other one millions of light years away), we begin to suspect that each of the particles may be treated as free. Then, we decide to measure p_{x1}. However, after we do it, we know *with absolute certainty the momentum of the second particle* $p_{x2} = P - p_{x1}$, and this knowledge has been acquired *without any perturbation of particle* 2. According to the above cited statement, one has to admit that p_{x2} represents an element of physical reality. So far so good. However, we *might have decided* with respect to particle 1 to measure its *coordinate* x_1. If this happened, then we would know *with absolute certainty the position of the second particle*, $x_2 = x - x_1$, without perturbing particle 2 at all. Therefore, x_2, as p_{x2}, is an element of physical reality. The Heisenberg uncertainty principle says that it is *impossible* for x_2 and p_{x2} to be exactly measurable quantities. Conclusion: the Heisenberg uncertainty principle is wrong, and quantum mechanics is at least incomplete.

A way to defend the Heisenberg principle was to treat the two particles as an indivisible total system and reject the supposition that the particles are independent, even if they are millions of light years apart. This is how Niels Bohr desperately defended himself against Einstein (and his two colleagues). He said that the state of the total system in fact never fell apart into particles 1 and 2, and still is in what is known as an *entangled quantum state*[62] of the system of particles 1 and 2, and

> any measurement influences the state of the system as a whole, *independently of the distance of particles* A and B.

This reduces to the statement that measurement manipulations on particle 1 influence the results of measurements on particle 2. This correlation between measurements on particles 1 and 2 has to take place *immediately, regardless of the space that separates them.*[63] This is a shocking and

---

[61] Indeed, $\hat{x}\hat{P} - \hat{P}\hat{x} = (\hat{x}_1 - \hat{x}_2)(\hat{p}_{x1} + \hat{p}_{x2}) - (\hat{p}_{x1} + \hat{p}_{x2})(\hat{x}_1 - \hat{x}_2) = [\hat{x}_1, \hat{p}_{x1}] - [\hat{x}_2, \hat{p}_{x2}] + [\hat{x}_1, \hat{p}_{x2}] - [\hat{x}_2, \hat{p}_{x1}] = -i\hbar + i\hbar + 0 - 0 = 0$.

[62] To honor Einstein, Podolsky, and Rosen the entanglement of states is sometimes called the EPR effect.

[63] Nevertheless, the correlation is not quite clear. One may pose some questions. The statement about the instantaneous correlation between particles 1 and 2 in the EPR effect cannot be correct, because the measurements are separated in the space–time manifold and the simultaneity is problematic (see Chapter 3). What is the laboratory fixed coordinate system? How is information about particle 2 transferred to where we carry out the measurement on particle 1? This takes time. After that time particle 1 is elsewhere. Is there anything to say about the separation time? In which coordinate system is the separation time measured?

nonintuitive feature of quantum mechanics. This is why it is often said that quantum mechanics cannot be understood. One can apply it successfully and obtain an excellent agreement with experiment, but there is something strange in its foundations. This represents a challenge: an excellent theory, but based on some unclear foundations.

In the following, some precise experiments will be described, in which it is shown that quantum mechanics is right, however absurd it looks. Before that let us consider a cat.

1.6 The life and death of Schrödinger's cat

The abovementioned paper by Einstein, Podolsky, and Rosen represented a severe critique of quantum mechanics in the form it has been presented by its fathers. After the appearance of the article, Erwin Schrödinger published a series of works[64] showing some other problematic issues in quantum mechanics. In particular, he described a *Gedankenexperiment*, later known as the Schrödinger's cat paradox. According to Schrödinger it shows some absurd consequences of quantum mechanics.

Here is the paradox. There is a cat closed in an isolated steel box (filled with air). Together with the cat there is a Geiger counter in the box (protected from the cat). We put in the counter a bit of a radioactive substance, carefully prepared in such a way that every hour two events happen with the same probability: either a radioactive nucleus decays or no radioactive nucleus decays. If a nucleus really decays, it causes ionization and electric discharge in the counter tube, which in turn results in a hammer hitting a glass capsule with hydrocyanic acid (HCN) gas, which kills the cat. If one puts the cat in for an hour, then *before* opening the box, one cannot say the cat is dead or alive. According to the Copenhagen interpretation, this is reflected by a proper wave function Ψ (for the box with everything in it), which is a superposition of two states with equal probability amplitudes: one state corresponds to the cat being alive, the other to the dead cat. The function Ψ describes therefore a cat in an intermediate state, neither alive nor dead, just in the middle between life and death, which, according to Schrödinger, would represent a totally absurd description. Einstein joined Schrödinger enthusiastically in this mockery from the Copenhagen interpretation, adding in his style "an even better" idea about how to kill the cat (gun powder instead of cyanide). All in all, in this way an atomic-scale phenomenon (decay of a nucleus) can result in some drastic events in the macroscopic world. If, after an hour, an observer opens the box ("measurement" of the state of the content of the box), then according to the Copenhagen interpretation, this will result in the collapse of the corresponding wave function and the observer will find the cat *either* alive *or* dead. This is the crux of the paradox.

Since the early days of quantum mechanics many scholars were trying to rationalize the paradox, always relying on some particular interpretation of quantum mechanics. In one of the

---

[64] E. Schrödinger, *Naturwissenschaften*, 23(1935)807, 823, 844.

interpretations, quantum mechanics does not describe a single system, but rather an infinite set of systems. We have therefore plenty of cats and the same number of boxes, each of them with the same macabre gear inside. Then, the paradox disappears, because, after the boxes are open, in 50% of cases the cats will be alive and in 50% they will be dead. In another interpretation it is criticized that Schrödinger treats the box as a quantum system, while the observer is treated classically. In this interpretation not only Schrödinger plays the role of the observer, but also the cat, and even the box itself (since it may contain a camera). What happened may be described differently by each of the observers, depending on what information they have about the whole system. For example, in the cat (alive or dead) there is information what has happened even *before* the box is open. The human observer does not have this information. Therefore, the collapse of the wave function happened earlier for the cat than for the observer! Only after the box is open it will turn out for the cat and for the observer that the collapse happened to the same state.

Hugh Evereth stepped out with another, truly courageous, interpretation. According to Evereth we do not have our single Universe, as most of us might think naively, but a plethora of universes coexist, and in each of them some other things happen! When the box is open, the state of the box is entangled with the state of the observer, but then a collapse of the wave function happens, and in one of the universes the cat is alive, but in the other one it is dead (bifurcation). These two universes evolve independently ("parallel universes"), and they do not know about each other. Similar bifurcations happen massively in other events; hence, according to Evereth, the number of parallel Universes is astronomic.

The quantity of possible interpretations, large differences among them, and desperate character of some of them indicate that the problem of understanding quantum mechanics is still unsolved...

1.7 Bilocation

Assume that the world (stars, Earth, Moon, me, table, proton, electron, etc.) exists objectively. One may suspect this from everyday observations. For example, the Moon is seen by many people, who describe it in a similar way.[65] Instead of the Moon, let us begin with something

---

[65] This *may* indicate that the Moon exists independently of our observations and overcome importunate suspicions that the Moon ceases to exist when we do not look at it. Besides, there are people who claim to have seen the Moon from very close and even touched it (admittedly through a glove) and this slightly strengthens our belief in the Moon's existence. First of all, one has to be cautious. For example, some chemical substances, hypnosis, or an ingenious set of mirrors may cause some people to be convinced about the reality of some phenomena, while others do not see them. Yet, would it help if even everybody saw? We should not verify serious things by voting. The example of the Moon also intrigued others; cf. D. Mermin, "*Is the Moon there, when nobody looks?*", *Phys. Today*, 38(1985)38.

simpler: how is it with electrons, protons, or other elementary particles? This is an important question because the world as we know it – including the Moon – is mainly composed of protons.[66] Here one encounters a mysterious problem. I will try to describe it by reporting results of several experiments.

Following Richard Feynman,[67] imagine two slits in a wall. Every second (the time interval has to be large enough to be sure that we deal with properties of a *single* particle) we send an electron towards the slits. There is a screen behind the two slits, and when an electron hits the screen, there is a flash (fluorescence) at the point of collision. Nothing special happens. Some electrons will not reach the screen at all, but traces of others form a pattern, which seems quite chaotic. The experiment looks monotonous and boring. Just a flash here, and another there. One cannot predict where a particular electron will hit the screen. But suddenly we begin to suspect that there is some regularity in the traces (Fig. 1.14).

A strange pattern appears on the screen: regions with high concentrations of traces are separated by regions of low concentration. This resembles the interference of waves, e.g., a stone thrown into water causes interference behind two slits: an alternation of high and low amplitudes of water level. Well, but what has an electron in common with a wave on the water surface? The interference on water was possible, because there were two sources of waves (the Huygens principle) – two slits.

Common sense tells us that nothing like this could happen with the electron, because, firstly, the electron could not pass through *both* slits, and, secondly, unlike the waves, the electron has hit a tiny spot on the screen (transferring there its energy). Let us repeat the experiment with a *single* slit. The electrons go through the slit and make flashes on the screen here and there, but there is only a single major concentration region (just facing the slit) fading away from the center (with some minor minima).

This result should make you feel faint. Why? You would like the Moon, a proton, or an electron to be solid objects, wouldn't you? All investigations made so far indicate that the electron is a point-like elementary particle. If, in the experiments we have considered, the electrons were to be divided into two classes, i.e., those that passed through slit 1 and those that passed through slit 2, then the electron patterns would be different. The pattern with the two slits *had to be* the

---

[66] In the darkest communist times a colleague of mine came to my office. Conspiratorially, very excited, and occasionally looking over his shoulder, he whispered: "*The proton decays!!!*" He just read in a government newspaper that the lifetime of a proton turned out to be finite. When asked about the lifetime, he gave an astronomical number, something like 10^{30} years or so. I said: "*Why do you look so excited then and why all this conspiracy?*" He answered: "*The Soviet Union is built of protons, and therefore is bound to decay as well!*"

[67] After Richard Feynman, "The Character of Physical Law," MIT Press, 1967.

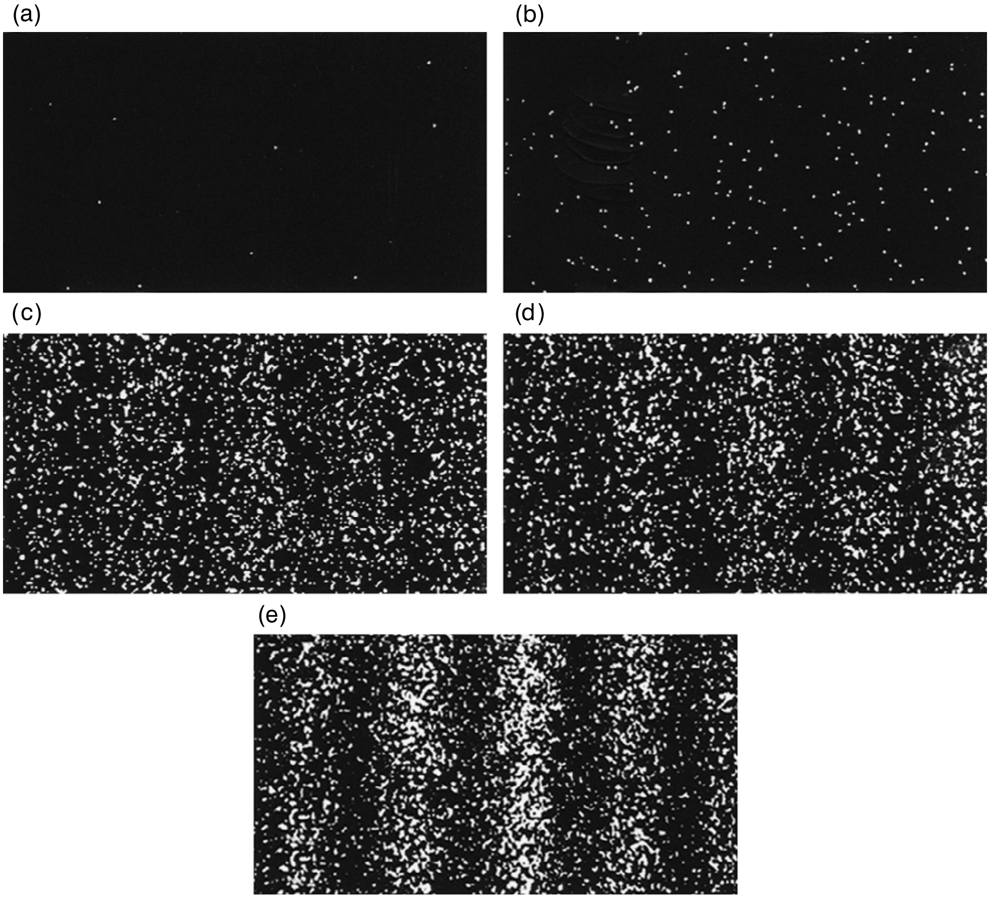


Fig. 1.14. Two-slit electron interference pattern registered by Akira Tonomura. (a) 10 electrons. (b) 100 electrons. (c) 3000 electrons – one begins to suspect something. (d) 20 000 electrons – no doubt, we will have a surprise. (e) 70 000 electrons – here it is! Conclusion: there is only one possibility – each electron went through the two slits (according to J. Gribbin, "*Q is for Quantum: An Encyclopedia of Particle Physics*," Weidenfeld and Nicolson, 1998).

sum of the patterns corresponding to only one open slit (facing slit 1 and slit 2). *We do not have that picture*.

The *only* explanation for this interference of the electron with itself is that with the two slits open it went through *both*.

Clearly, the two parts of the electron united somehow and caused the flash at a single point on the screen. The quantum world is really puzzling. Despite the fact that the wave function is delocalized, the measurement gives its single point position (decoherence). How could an electron pass simultaneously through two slits? We do not understand this, but this is what happens.

Maybe it is possible to pinpoint the electron passing through two slits? Indeed, one may think of the Compton effect: a photon collides with an electron and changes its direction, and this can be detected ("a flash on the electron"). When one prepares two such ambushes at the two open slits, it turns out that the flash is *always on a single slit, not on both*. This cannot be true! If it were true, then the pattern would be of a *noninterference* character (and had to be the sum of the two one-slit patterns), but we have the interference. No. There is no interference. *Now*, the pattern does not show the interference. The interference was when the electrons were not observed. When we observe them, there is no interference... Somehow we perturb the electron's momentum (the Heisenberg principle) and the interference disappears.

We have to accept that the electron passes through two slits. Maybe it only pertains to the electron, maybe the Moon is something completely different? A weak hope. The same thing happens to protons. Sodium atoms were also found to interfere.[68] A sodium atom, of diameter of a few Å, looks like an ocean liner when compared to a child's toy boat of a tiny electron (almost 42 000 times lighter). And this ocean liner passed through two slits separated by thousands of Å. At the end of 1999 similar interference was observed for fullerene,[69] a giant C_{60} molecule, about one million times heavier than the electron. It is worth noting that after such adventure the fullerene molecule remained intact: somehow all its atoms, with the details of their chemical bonds, preserved their nature. There is something intriguing in this.

1.8 The magic of erasing the past

John Archibald Wheeler was not completely satisfied by the description of the hypothetical two-slit experiment, if it were performed at an astronomically large scale. Indeed, there was something puzzling in it. Suppose we have the two slits far away from us. A photon goes through the slit region, then flies towards us for a long time, and finally arrives to our screen. At large numbers of such photons, we obtain *an interference picture on the screen that witnesses each photon went through two slits.* However, when a photon passed the slit region, leaving the slits behind and heading towards us, we had plenty of time (say, of the order of a billion of years) to think. In particular, we might have an idea to replace the screen by two telescopes, each directed on one slit. In a similar situation we never observe half of the photon in one telescope and half in the other one: there is always a photon seen in *one* telescope only. *The*

---

[68] To observe such phenomena the slit distance has to be of the order of the de Broglie wave length, $\lambda = h/p$, where h is the Planck constant and p is the momentum. Cohen-Tannoudji lowered the temperature to such an extent that the momentum was close to 0, and λ could be of the order of thousands of Å.

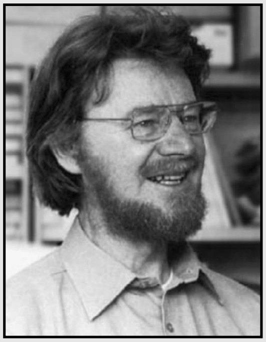
[69] M. Arndt, O. Nairz, J. Voss-Andreae, C. Keller, G. van der Zouw, A. Zeilinger, *Nature*, 401(1999)680.

telescopes allow us to identify unambiguously which slit the particular photon passed through. Therefore, we may divide all such photons into two classes (depending on the slit they were through), and their distribution cannot be interference-like; it has to be bullet-like (it is a sum of distributions of both classes). And now we have a paradox.

> *Quantum eraser*
> Our decision (taken *after* the photons passed the slits) to replace the screen by the telescopes visibly changed the way the photons have been passing the slits. The problem is that the photon's decision must have been anticipating our action by some billion years or so.

This strange behavior (for a much shorter time scale) has been confirmed experimentally[70] in 2007. Well, this is a kind of drama. How may such a thing happen, however absurd it looks? Nobody knows. Let us put that clearly: it seems there is something fundamentally wrong in our current understanding of the world. Therefore, it is perhaps more safe to limit ourselves to strict physical *measurements* rather than telling pure science-fiction-like stories about traveling particles and their decisions anticipating situations that will happen in some billion years. Note also that we have "changed the past" only in a very special sense, since we are speaking here about two separate experiments: one with the telescopes and afterwards the other one with the screen. This is not like undoing an airplane catastrophe that has already taken place, but rather like changing something before the catastrophe takes place in order not to have it come.

1.9 A test for common sense: the Bell inequality



John Stuart Bell (1928–1990), Irish mathematician at Centre Européen de la Recherche Nucleaire (CERN) in Geneva. In the 1960s Bell attacked an old controversy of locality versus nonlocality, hidden variables, etc., apparently exhausted after exchange of ideas between Einstein and Bohr.

John Bell proved a theorem in 1964 that pertains to the results of measurements carried out on particles and some of the inequalities they have to fulfill. The theorem pertains to the basic logic of the measurements and is valid independently of the kind of particles and of the nature of their interaction. The theorem soon became

---

[70] V. Jacques, E. Wu, F. Grosshans, F. Treussart, Ph. Grangier, A. Aspect, J.-F. Roch, *Science*, 315(2007)966.

very famous, because it turned out to be a useful tool allowing us to verify some fundamental features of our knowledge about the world.[71]

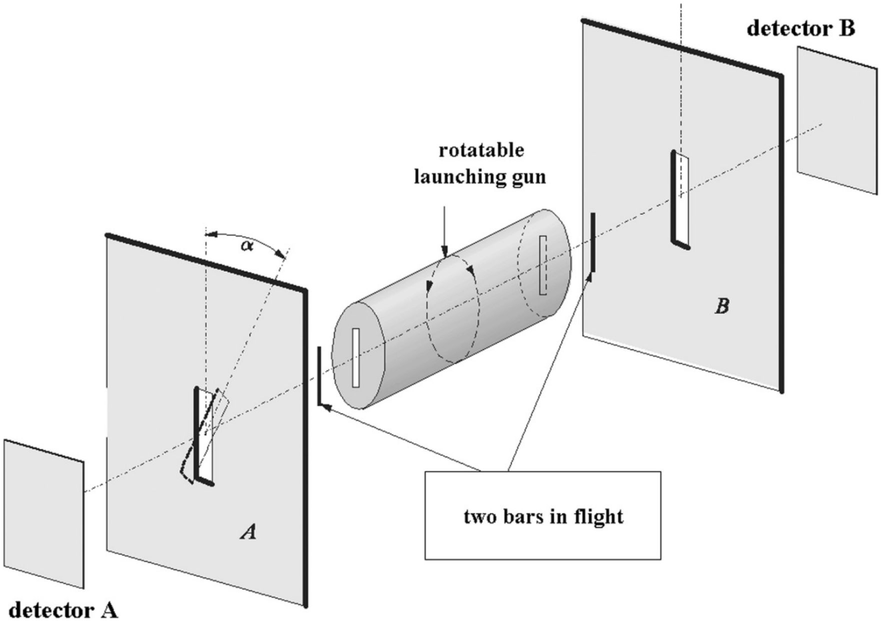


Fig. 1.15. Bell inequalities. A bar launching gun adopts stochastic positions (of equal probability) when rotating about the axis. Each time the full magazine of bars is loaded. The slits also may be rotated about the axis. The bars arrive at slits A and B. Some will go through and be detected behind the slits.

Imagine a launching gun (Fig. 1.15) which ejects a series of pairs of identical rectangular bars flying along a straight line (no gravitation) in opposite directions (opposite velocities). The axes of the bars are always parallel to each other and always perpendicular to the straight line. The launching machine is constructed in such a way that the probabilities of all orientations of the bars are equal, and that any two launching series are identical. At a certain distance from the launching machine there are two rectangular slits A and B (the same on both sides). If the bar's longer axis coincides with the longer dimension of the slit, then the bar will go through for sure and will be registered as "1," i.e., "it has arrived" by the detector. If the bar's longer axis coincides with the shorter axis of the slit, then the bar will not go through for sure, and will be detected as "0". For other angles between the bar and slit axes the bar will sometimes go through (when it fits the slit), sometimes not (when it does not fit the slit).[72]

---

[71] The description given below follows the line shown by W. Kołos, Proceedings of the IV Castel Gandolfo Symposium of John Paul II, 1986.

[72] Simple reasoning shows that for a bar of length L, two possibilities, "to go through" and "not to go through," are equally probable (for a bar of zero width) if the slit width is equal to $\frac{L}{\sqrt{2}}$.

Having prepared the launching gun (our magazine contains 16 pairs of bars) we begin our experiments. Four experiments will be performed. Each experiment will need the full magazine of bars. In the first experiment the two slits will be parallel. This means that the fate of both bars in any pair will be exactly the same: if they go through, they will both do so, and if they are stopped by the slits, they will both be stopped. Our detectors have registered (we group the 16 pairs in clusters of 4 to make the sequence more transparent):

Experiment I (angle 0)
Detector A: 1001 0111 0010 1001
Detector B: 1001 0111 0010 1001

Now, we repeat Experiment I, but this time slit A will be rotated by a small angle α (Experiment II). At slit B nothing has changed, and therefore we must obtain there exactly the same sequence of zeros and ones as in Experiment I. At slit A, however, the results may be different. Since the rotation angle is small, the difference list will be short. We might get the following result:

Experiment II (angle α)
Detector A: 10**1**1 0111 0010 **0**001
Detector B: 1001 0111 0010 1001

There are two differences (highlighted in bold) between the lists for the two detectors.

Now for Experiment III. This time slit A comes back to its initial position, but slit B is rotated by $-\alpha$. Because of our perfect gun, we must obtain at detector A the same result as in Experiment I. However, at B we find some difference with respect to Experiments I and II:

Experiment III (angle $-\alpha$)
Detector A: 1001 0111 0010 1001
Detector B: 1001 0**0**11 0**1**10 1001

There are two differences (bold) between the two detectors.

We now carry out Experiment IV. We rotate slit A by angle α and slit B by angle $-\alpha$. Therefore, at detector A we obtain the same results as in Experiment II, and at detector B the same as in Experiment III. Therefore, we detect:

Experiment IV (angle 2α)
Detector A: 10**1**1 0111 0010 **0**001
Detector B: 1001 0**0**11 0**1**10 1001

Now there are four differences between detector A and detector B. In Experiment IV *the number of differences could not be larger (Bell inequality)*. In our case it could be four or fewer. When would it be fewer? When accidentally the bold figures (i.e., the differences of Experiments II and III with respect to those of Experiment I) coincide. In this case this would be counted as a difference in Experiments II and III, while in Experiment IV it would not be counted as a difference.

Thus, we have demonstrated the following.

Bell inequality:

$$N(2\alpha) \leq 2N(\alpha), \tag{1.26}$$

where N stands for the number of measurement differences. The Bell inequality was derived under assumption that whatever happens at slit A does not influence what happens at slit B (this is how we constructed the counting tables) and that the two flying bars have, maybe unknown for the observer, only a real (definite) direction in space (the same for both bars).

It would be interesting to perform a real experiment similar to Bell's to confirm the Bell inequality; see the following section. This opens the way for deciding in a physical experiment whether

- elementary particles are classical (though extremely small) objects that have some well-defined attributes irrespective of whether we observe them (Einstein's view), or
- elementary particles do not have such attributes and only measurements themselves make them have measured values (Bohr's view).

1.10 Photons violate the Bell inequality

French scientists from the Institute of Theoretical and Applied Optics in Orsay published the results of their experiments with photons.[73] The excited calcium atom emitted pairs of photons (analogs of our bars) which moved in opposite directions and had the same polarization. After flying about 6 m they both met the polarizers – analogs of slits A and B in the Bell procedure. A polarizer allows a photon with polarization state $|0\rangle$, or "parallel" (to the polarizer axis), always to pass through, and always rejects any photon in the polarization state $|1\rangle$, or "perpendicular" (indeed perpendicular to the above "parallel" setting). When the polarizer is rotated about the optical axis by an angle, it will allow to pass through a percentage of the photons in state $|0\rangle$ and

---

[73] A. Aspect, J. Dalibard, G. Roger, *Phys. Rev. Lett.*, 49(1982)1804.

a percentage of the photons in state $|1\rangle$. When both polarizers are in the "parallel" setting, there is perfect correlation between the two photons of each pair, i.e., exactly as in Bell's Experiment I. In the photon experiment, this correlation was checked for 50 millions photons every second for about 12 000 seconds.

Bell's experiments II–IV have been carried out. Common sense indicates that even if the two photons in a pair have random polarizations (perfectly correlated though always the same – like the bars), they still have *some* polarizations, i.e., maybe unknown but definite (as in the case of the bars, i.e., what E, P, and R believed happens). *Hence, the results of the photon experiments would have to fulfill the Bell inequality.* However, the photon experiments have shown that the Bell inequality is violated, but still the results are in accordance with the prediction of quantum mechanics.

There are therefore only two possibilities (compare the two points at the end of the previous section):

a) either the measurement on a photon carried out at polarizer A (B) results in some *instantaneous interaction* with the photon at polarizer B (A), and/or

b) the polarization of any of these photons is completely *indefinite* (even if the polarizations of the two photons are fully correlated, i.e., the same) and only the *measurement on one of the photons at A (B) determines its polarization*, which results in the automatic determination of the polarization of the second photon at B (A), even if they are separated by millions of light years.

Both possibilities are sensational. The first assumes a strange form of communication between the photons or the polarizers. This communication must be propagated with a velocity *exceeding* the speed of light, because an experiment was performed in which the polarizers were switched (this took something like 10 nanoseconds) *after* the photons started (their flight took about 40 nanoseconds). Despite this, communication between the photons did exist.[74] Possibility (b) as a matter of fact represents Bohr's interpretation of quantum mechanics: elementary particles do not have definite attributes (e.g., polarization).

> As a result there is dilemma: either the world is "nonreal" (in the sense that the properties of particles are not determined before measurement) and/or there is instantaneous (i.e., faster than light) communication between particles which operates independently of how far apart they are ("nonlocality").

This dilemma may make everybody's metaphysics shiver.

---

[74] This again is the problem of delayed choice. It seems that when starting the photons have a knowledge of the *future* setting of the apparatus (the two polarizers).

1.11 Teleportation

The idea of teleportation comes from science fiction and means:

- *acquisition of full information* about an object located at A,
- *its transmission* to B,
- *creation* (materialization) of an identical object at B, and
- *at the same time, the disappearance* of the object at A.

At first sight *it seems* that this contradicts quantum mechanics. The Heisenberg uncertainty principle says that it is not possible to prepare a perfect copy of the object, because, in case of mechanical quantities with noncommuting operators (like positions and momenta), there is *no way to have them measured exactly*, in order to rebuild the system elsewhere with the same values of the quantities.

The trick is, however, that the quantum teleportation we are going to describe *will not violate* the Heisenberg principle, because the mechanical quantities needed *will not be measured and the copy, based on their values, will not be made*.

The teleportation protocol was proposed by Bennett and coworkers[75] and applied by Anton Zeilinger's group.[76] The latter used the entangled states (EPR effect) of two photons described above.[77]

Assume that photon A (number 1) from the entangled state belongs to Alice, and photon B (number 2) to Bob. Alice and Bob know this and introduce a common fixed coordinate system. Both photons have *identical* polarizations in this coordinate system. Let us assume that the state of the two photons is the following superposition[78]: $|00\rangle + |11\rangle$, where the first position in every ket pertains to Alice's photon, the second to Bob's.

Now, Alice wants to *carry out teleportation of her additional photon (number 3)*, which is in an unknown quantum state

$$\phi_{\text{tele}} = a|0\rangle + b|1\rangle$$

---

[75] C.H. Benneth, G. Brassard, C. Crépeau, R. Josza, A. Peres, W.K. Wootters, *Phys. Rev. Letters*, 70(1993)1895.

[76] Then at the University of Innsbruck (Austria).

[77] A UV laser beam hits a barium borate crystal (known for its birefringence). Photons with parallel polarization move along the surface of a cone (with the origin at the beam–surface collision point), the photons with perpendicular polarization move on another cone, the two cones intersecting. From time to time a single UV photon splits into two equal energy photons of different polarizations. Two such photons when running along the intersection lines of the two cones, *and therefore not having a definite polarization* (i.e., being in a superposition state composed of both polarizations), represent the two entangled photons.

[78] The teleportation result does not depend on the state.

(known as *qubit*), where a and b stand for *unknown* coefficients[79] satisfying the normalization condition $a^2 + b^2 = 1$. Therefore, the state of three photons (Alice's: the first and the third position in the three-photon ket, Bob's: the second position) will be

$$\psi_{\text{start}} = [|00\rangle + |11\rangle]\,[a|0\rangle + b|1\rangle] = a|000\rangle + b|001\rangle + a|110\rangle + b|111\rangle. \tag{1.27}$$

Note that Alice might decide to measure the polarization of her two photons (1 and 3) and send her results by telephone to Bob. This, however, will not help Bob to construct the state ϕ_{tele} of his photon. Indeed, if Alice's result were $|00\rangle$ or $|01\rangle$, Bob would only know that his photon's state is $|0\rangle$. If Alice communicated $|10\rangle$ or $|11\rangle$ Bob's photon state would be $|1\rangle$. He knew that sometimes he may get $|0\rangle$ and sometimes $|1\rangle$ even without any phone call, even from his girlfriend.

Alice is however a very smart girl. She prepares herself for teleportation of the qubit ϕ_{tele} (corresponding to her photon 3). *The crux of the procedure will be to use such experimental optical devices, which transform the initial three-photon state* ψ_{start} *to a very special final state* ψ_{final}. *Alice's measurement results (in the state* ψ_{final}*) will be linked unambiguously to Bob's knowledge of what he is supposed to do with his photon (using some optical devices) in order to get the* ϕ_{tele} *state, which means teleportation*[80] *of* ϕ_{tele}.

To this end Alice prepares first a device called the *XOR gate*.[81]

Table 1.1. The XOR gate changes the state of the steered photon, only when the steering photon is on.

Steering	Steered before XOR	Steered after XOR
$\|0\rangle$	$\|0\rangle$	$\|0\rangle$
$\|0\rangle$	$\|1\rangle$	$\|1\rangle$
$\|1\rangle$	$\|0\rangle$	$\|1\rangle$
$\|1\rangle$	$\|1\rangle$	$\|0\rangle$

What is the XOR gate? The device manipulates two photons; one is treated as the steering photon, the second as the steered photon (see Table 1.1). The device operates (the operator $\hat{A}_{\text{XOR}}$) as follows: if the steering photon is in state $|0\rangle$, then no change is introduced for the

---

[79] Neither Alice nor Bob will know these coefficients up to the end of the teleportation procedure, but still Alice will be able to send her qubit to Bob!

[80] At the same time, Alice's abovementioned measurements will cause the collapse of her ϕ_{tele}, which means the disappearance of ϕ_{tele}, cf. p. 28.

[81] Abbreviation of "*eXclusive OR*."

state of the steered photon. If, however, the steering photon is in state $|1\rangle$, the steered photon will be switched over, i.e., it will be changed from 0 to 1 or from 1 to 0.

Alice chooses the photon in the state ϕ_{tele} as her steering photon 3, and photon 1 as her steered photon. After the XOR gate is applied, the state of the three photons will be as follows:

$$\hat{\text{A}}_{\text{XOR}}\psi_{\text{start}} = a|000\rangle + b|101\rangle + a|110\rangle + b|011\rangle. \tag{1.28}$$

Alice continues her preparation by using another optical device, called the *Hadamard gate*, which operates ($\hat{\text{A}}_{\text{Had}}$) on a single photon and does the following:

$$|0\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle),$$
$$|1\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

Alice applies the Hadamard operation to her photon 3, and after this the three-photon state is changed to the following:

$$\begin{aligned}
&\psi_{\text{final}} = \hat{\text{A}}_{\text{Had}}\hat{\text{A}}_{\text{XOR}}\psi_{\text{start}} = \\
&\frac{1}{\sqrt{2}}[a|000\rangle + a|001\rangle + b|100\rangle - b|101\rangle + a|110\rangle + a|111\rangle + b|010\rangle - b|011\rangle] = \\
&\frac{1}{\sqrt{2}}[|0\ (a|0\rangle + b|1\rangle)\ 0\rangle + |0\ (a|0\rangle - b|1\rangle)\ 1\rangle - |1\ (a|1\rangle + b|0\rangle)\ 0\rangle + |1\ (a|1\rangle - b|0\rangle)\ 1\rangle].
\end{aligned} \tag{1.29}$$

We have a superposition of four three-photon states in the last row. Each state shows the state of Bob's photon (number 2 in the ket), at any given state of Alice's two photons. Finally, Alice carries out the measurement of the polarization states of her photons (1 and 3). This inevitably causes her to get (for each of the photons) either $|0\rangle$ or $|1\rangle$. This causes her to know the state of Bob's photon from the three-photon superposition (1.29):

- Alice's photons 00, i.e., Bob has his photon in state $(a|0\rangle + b|1\rangle) = \phi_{\text{tele}}$,
- Alice's photons 01, i.e., Bob has his photon in state $(a|0\rangle - b|1\rangle)$,
- Alice's photons 10, i.e., Bob has his photon in state $(a|1\rangle + b|0\rangle)$,
- Alice's photons 11, i.e., Bob has his photon in state $(a|1\rangle - b|0\rangle)$.

Then Alice calls Bob and tells him the result of her measurements of the polarization of her two photons.

Bob knows therefore, that if Alice sends him $|00\rangle$ this means that the teleportation is over: he already has his photon in state ϕ_{tele}! If Alice sends him one of the remaining possibilities, he would know exactly what to do with his photon to prepare it in state ϕ_{tele} and he does this with his equipment. The *teleportation is over: Bob has the teleported state* ϕ_{tele}*, Alice has lost her photon state* ϕ_{tele} *when performing her measurement (wave function collapse).*

Note that to carry out the successful teleportation of a photon state Alice had to communicate something to Bob by a classical channel (like telephone).

1.12 Quantum computing

Richard Feynman pointed out that contemporary computers are based on the "all" or "nothing" philosophy (two bits: $|0\rangle$ *or* $|1\rangle$), while in quantum mechanics one may also use a linear combination (superposition) of these two states with arbitrary coefficients a and b: $a|0\rangle + b|1\rangle$. Would a quantum computer based on such superpositions be better than a traditional one? The hope associated with quantum computers relies on a multitude of quantum states (those obtained using variable coefficients $a, b, c, \ldots$) and the possibility of working with many of them using a single processor. It was (theoretically) proved in 1994 that quantum computers could factorize natural numbers much faster than traditional computers. This sparked intensive research on the concept of quantum computation, which uses the idea of entangled states. According to many researchers, any entangled state (a superposition) is extremely sensitive to the slightest interaction with the environment, and as a result decoherence takes place very easily, which is devastating for quantum computing.[82] First attempts at constructing quantum computers were based on protecting the entangled states, but, after a few simple operations, decoherence took place.

In 1997 Neil Gershenfeld and Isaac Chuang realized that any routine nuclear magnetic resonance measurement represents nothing but a simple quantum computation. The breakthrough was recognizing that a qubit may be also represented by the huge number of molecules in a liquid.[83] The nuclear spin angular momentum (say, corresponding to $s = \frac{1}{2}$) is associated with a magnetic dipole moment (Chapter V2-4) and those magnetic dipole moments interact with an external magnetic field and with themselves. An isolated magnetic dipole moment has

---

[82] It pertains to an entangled state of (already) distant particles. When the particles interact strongly the state is more stable. The wave function for H_2 also represents an entangled state of two electrons with opposite spins, yet decoherence does not take place even at short internuclear distances. As we will see, entangled states can also be obtained in liquids.

[83] Interaction of the molecules with the environment does not necessarily result in decoherence.

two states in a magnetic field: a lower energy state corresponding to the antiparallel configuration (state $|0\rangle$) and a higher energy state related to the parallel configuration (state $|1\rangle$). By exposing a sample to a carefully tailored nanosecond radio wave impulse one obtains a rotation of the nuclear magnetic dipoles, which corresponds to their state being a superposition $a|0\rangle + b|1\rangle$.

Here we present a prototype of the XOR gate. Take chloroform,[84] $^{13}CHCl_3$. Due to the interaction of the magnetic dipoles of the proton and of the carbon nucleus (both either in parallel or antiparallel configurations with respect to the external magnetic field) a radio wave impulse of a certain frequency causes the carbon nuclear spin magnetic dipole to rotate by 180° *provided* the proton spin dipole moment is parallel to that of the carbon.[85] Similarly, one may conceive other logical gates. The spins change their orientations according to a sequence of impulses, which play the role of a computer program. There are many technical problems to overcome in the quantum computers "in liquid": the magnetic interaction of distant nuclei is very weak, and decoherence remains a worry and for the time being limits the number of operations to several hundred. However, this is only the beginning of a new computer technology. It is most important that chemists know the future computers well – they are simply molecules.

Summary

Classical mechanics was unable to explain certain phenomena: black body radiation, the photoelectric effect, and the stability of atoms and molecules as well as their spectra. Quantum mechanics, created mainly by Werner Heisenberg and Erwin Schrödinger, explained these effects. The new mechanics was based on six postulates:

- Postulate I says that all information about the system follows from the wave function ψ. The quantity $|\psi|^2$ represents the probability density of finding particular values of the coordinates of the particles the system is composed of.
- Postulate II allows mechanical quantities (e.g., energy) to be ascribed to *operators*. One obtains the operators by writing down the classical expression for the corresponding quantity, and replacing momenta (e.g., p_x) by momenta operators (here, $\hat{p}_x = -i\hbar\frac{\partial}{\partial x}$).
- Postulate III gives the time evolution equation for the wave function ψ (time-dependent Schrödinger equation, $\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}$), using the energy operator (*Hamiltonian* $\hat{H}$). For time-independent $\hat{H}$ one obtains the time-independent Schrödinger equation $\hat{H}\psi = E\psi$ for the stationary states.

---

[84] The nuclear magnetic resonance operations on spins pertain in practice to a tiny fraction of the nuclei of the sample (of the order of 1:1 000 000).

[85] This is indeed consistent with the XOR gate logical table, because for the parallel spins (entries: 00 and 11) the output is 0 (meaning: transition), while for the opposite spins (entries: 01 and 10) the output is 1 (meaning: no transition).

- Postulate IV pertains to ideal measurements. When making a measurement of a quantity A, one can obtain only *an eigenvalue of the corresponding operator* $\hat{A}$. If the wave function ψ represents an eigenfunction of $\hat{A}$, i.e., ($\hat{A}\psi = a\psi$), then one obtains always as a result of the measurement the eigenvalue corresponding to ψ (i.e., a). If, however, the system is described by a wave function, which *does not represent any eigenfunction of* $\hat{A}$, then one obtains also an eigenvalue of $\hat{A}$, but there is no way to predict which eigenvalue. The only thing one can predict is the mean value of many measurements, which may be computed as $\bar{a} = \frac{\left\langle \psi|\hat{A}\psi \right\rangle}{\langle \psi|\psi \rangle}$.
- Postulate V says that an elementary particle has an internal angular momentum (spin). One can measure only two quantities: the square of the spin length $s(s+1)\hbar^2$ and one of its components $m_s\hbar$, where $m_s = -s, -s+1, ..., +s$, with spin quantum number $s \geq 0$ characteristic for the type of particle (integer for bosons, half-integer for fermions). The spin magnetic quantum number m_s takes $2s+1$ values, related to the $2s+1$ values of the (granular) spin coordinate σ.
- Postulate VI has to do with the symmetry of the wave function with respect to different labeling of identical particles. If one exchanges the labels of two identical particles (we sometimes call this the exchange of all the coordinates of the two particles), then for two identical fermions the wave function has to change its sign (antisymmetric), while for two identical bosons the function does not change (symmetry). As a consequence, two identical fermions with the same spin coordinate cannot occupy the same point in space.

Quantum mechanics is one of the most peculiar theories. It gives numerical results that agree extremely well with experiments, but on the other hand the relation of these results to our everyday experience sometimes seems shocking. For example, it turned out that a particle or even a molecule may somehow exist in two locations (they pass through two slits simultaneously), but when one checks that out they are always in one place. It also turned out that

- either a particle has no definite properties ("*the world is unreal*"), and the measurement fixes them somehow, or/and
- there is instantaneous communication between particles however distant they are from each other ("*nonlocality of interactions*").

It turned out that in the Bohr–Einstein controversy, Bohr was right. The Einstein–Podolsky–Rosen paradox resulted (in agreement with Bohr's view) in the concept of entangled states. These states have been used to experimentally teleport a photon state without violating the Heisenberg uncertainty principle. Also the entangled states stand behind the idea of quantum computing: with a superposition of two states (qubit) $a|0\rangle + b|1\rangle$ instead of $|0\rangle$ and $|1\rangle$ as information states.

Main concepts, new terms

antisymmetric function (p. 39)
axis of quantization (p. 32)
Bell inequality (p. 54)
bilocation (p. 52)
decoherence (p. 52)
delayed choice (p. 53)
Dirac notation (p. 23)
eigenfunction (p. 19)
eigenvalue problem (p. 26)
entangled states (p. 47)

EPR effect (p. 47)
experiment of Aspect (p. 57)
Gedankenexperiment (p. 44)
Heisenberg uncertainty principle (p. 44)
Hilbert space (p. 598)
interference of particles (p. 52)
locality of the world (p. 54)
logical gate (p. 59)
mean value of an operator (p. 29)
measurement (p. 26)
normalization (p. 21)
operator of a quantity (p. 22)
Pauli exclusion principle (p. 40)
Pauli matrices (p. 29)
quantum eraser (p. 53)
qubit (p. 59)
"reality of the world" (p. 54)
Schrödinger's cat (p. 49)
singlet state (p. 37)
spin angular momentum (p. 29)
spin coordinate (p. 29)
stationary state (p. 26)
symmetric function (p. 39)
symmetry of wave function (p. 39)
teleportation (p. 59)
time evolution equation (p. 24)
triplet state (p. 37)
wave function (p. 19)
wave function collapse (p. 28)

From the research front

Until recently, the puzzling foundations of quantum mechanics could not be verified directly by experiment. As a result of enormous technological advances in quantum electronics and quantum optics it became possible to carry out experiments on single atoms, molecules, photons, etc. It is possible nowadays to carry out teleportation across the Danube river. Even molecules such as fullerene were subjected to successful interference experiments. Quantum computer science is just beginning to prove that its principles are correct.

Ad futurum

Quantum mechanics has been proved in the past to give excellent results, but its foundations are still unclear.[86] There is no successful theory of decoherence that would explain why and how a delocalized state becomes localized after the measurement. It is possible to make fullerene interfere, and it may be that in the near future we will be able to do this with a virus.[87] It is interesting that fullerene passes instantaneously through two slits with its whole complex electronic structure as well as nuclear framework, although the de Broglie wave length is quite different for the electrons and for the nuclei. Visibly the "overweighted" electrons interfere differently from free ones. After the fullerene passes the slits, one sees it in a single spot on the screen (decoherence). It seems that there are cases when even strong interaction does not make decoherence necessary. Sławomir Szymański presented his theoretical and experimental results[88] and showed that the functional group $-CD_3$ exhibits a delocalized state (which

---

[86] A pragmatic viewpoint is shared by the vast majority: "*do not wiseacre, just compute!*"

[87] As announced by Anton Zeilinger.

[88] S. Szymański, *J. Chem. Phys.*, 111(1999)288.

corresponds to its rotation instantaneously in both directions, i.e., coherence) and, which makes the thing more peculiar, interaction with the environment *not only does not destroy the coherence, but makes it more robust*. This type of phenomenon might fuel investigations towards future quantum computer architectures.

Additional literature

"The Ghost in the atom: a discussion of the mysteries of quantum physics," P.C.W. Davies and J.R. Brown, eds, Cambridge University Press, (1986).

Two BBC journalists interviewed eight outstanding physicists: Alain Aspect (photon experiments), John Bell (Bell inequalities), John Wheeler (Feynman's PhD supervisor), Rudolf Peierls ("Peierls metal-semiconductor transition"), John Taylor ("black holes"), David Bohm ("hidden parameters"), and Basil Hiley ("mathematical foundations of quantum physics"). It is most striking that all these physicists give *very* different theoretical interpretations of quantum mechanics (partially covered in the present chapter).

R. Feynman, "QED – the Strange Theory of Light and Matter," Princeton University Press, Princeton, (1985).

Excellent popular presentation of quantum electrodynamics written by one of the outstanding physicists of the 20th century.

A. Zeilinger, "Quantum teleportation," *Scientific American*, 282(2000)50.

The leader in teleportation describes this new domain.

N. Gershenfeld, I.L. Chuang, "Quantum computing with Molecules," *Scientific American*, 278(1998)66.

First-hand information about nuclear magnetic resonance computing.

Ch. H. Bennett, "Quantum Information and Computation," *Physics Today*, 48(1995)24.

Another first-hand description.

F. Wilczek, "A Beautiful Question. Finding Nature's Deep Design," Penguin Press, London, 2015.

An attempt to sketch a generalized pattern of *all* interactions. In my opinion, the value of this book relies on two features. First, the author is a cofounder of chromodynamics (Nobel Prize in 2004) – the deepest theory of what we see around. Second, he makes an enormous effort to describe the topic in the simplest way possible, in addition discovering an amazing beauty of the (up to now) deepest pattern of Nature. One feels also a kind of a charming mystic-like enchantment by the internal beauty of Nature.

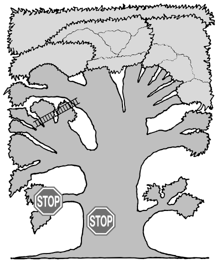
Questions

1. In order to calculate the expectation value $\bar{A}$ of a mechanical quantity A for a system described by the wave function ψ, one has to:
   a. express A in terms of coordinates and momenta, and then insert into A the most probable values of those.
   b. express A in terms of coordinates and momenta, and then replace in A the momenta by their operators.
   c. express A in terms of coordinates and momenta, replace in A the momenta by their operators thus producing operator $\hat{A}$, and then compute the integral $\bar{A} = \left\langle \psi | \hat{A}\psi \right\rangle$.
   d. impossible, if ψ does not satisfy the time-independent Schrödinger equation.
2. For a Hermitian operator $\hat{A}$:
   a. $\left\langle f | \hat{A} f \right\rangle$ is real for any function f.
   b. $\left\langle \phi | \hat{A}\psi \right\rangle = \left\langle \psi | \hat{A}\phi \right\rangle$ for any functions ψ, ϕ of class Q.
   c. $\left\langle \phi | \hat{A}\psi \right\rangle = \left\langle \psi | \hat{A}\phi \right\rangle$ for any functions ψ, ϕ.
   d. its eigenfunctions form a complete set of functions.
3. For a Hermitian operator $\hat{A}$ with eigenfunctions ψ_n, $n = 1, 2, ...,$ and eigenvalues a_n $(\hat{A}\psi_n = a_n\psi_n)$:
   a. a_n is real.
   b. ψ_n cannot take the value 0.
   c. ψ_n may be a real function.
   d. it may happen that $\psi_n \equiv 0$.
4. For a Hermitian operator $\hat{A}$ with eigenfunctions ψ_n, $n = 1, 2, ...,$ and eigenvalues a_n $(\hat{A}\psi_n = a_n\psi_n)$:
   a. ψ_n must be real.
   b. must be $\langle \psi_n | \psi_m \rangle = \delta_{mn}$.
   c. if ψ_1 and ψ_2 are some eigenfunctions of $\hat{A}$, $\phi = c_1\psi_1 + c_2\psi_2$ is necessarily an eigenfunction of $\hat{A}$.
   d. one can orthonormalize the set of eigenfunctions and in this way create such a set $\{\psi_n\}$, for which $\langle \psi_n | \psi_m \rangle = \delta_{mn}$.
5. For a Hermitian operator $\hat{A}$ with the orthonormalized eigenfunctions ψ_n, $n = 1, 2, ...,$ any function ϕ of class Q can be expanded into the series $\phi = \sum_n c_n\psi_n$, where c_n are some complex numbers:
   a. ϕ must be normalized.
   b. the above expansion $\phi = \sum_n c_n\psi_n$ cannot be finite or infinite.
   c. $c_n = \langle \psi_n | \phi \rangle$.
   d. $c_n = \langle \phi | \psi_n \rangle$.
6. $\hat{H}$ represents the time-independent Hamiltonian of the system, E stands for energy:
   a. the time-independent Schrödinger equation $\hat{H}\psi_n = E_n\psi_n$ has zero, a finite number, or an infinite number of solutions ψ_n, depending on the system under study.
   b. the solutions $\{\psi_n\}$ of the Schrödinger equation $\hat{H}\psi_n = E_n\psi_n$ form a complete set of the basis functions.
   c. the solutions $\{\psi_n\}$ of the Schrödinger equation $\hat{H}\psi_n = E_n\psi_n$ form a complete set of the basis functions and are orthonormal.

   d. the solutions $\{\psi_n\}$ of the Schrödinger equation $\hat{H}\psi_n = E_n\psi_n$ form a complete set of the basis functions, which may be chosen to be orthonormal.
7. $\hat{H}$ represents the Hamiltonian of the system, x, t stand for the spatial (x) and time (t) coordinates, respectively, and E stands for energy:
   a. $\hat{H}\Psi(x,t) = i\hbar\frac{\partial\Psi}{\partial t}$ represents the evolution in time of the wave function Ψ (the equation is known as the time-dependent Schrödinger equation).
   b. If $\hat{H}$ is time-independent, the solution of the time-dependent Schrödinger equation reads as $\Psi(x,t) = \sum_{n=1}^{\infty} c_n\Psi_n(x,t)$, where $\Psi_n(x,t)$ are the stationary states and c_n are constant complex numbers.
   c. Any stationary state has the following time dependence: $\Psi_n(x,t) = \psi_n(x)\exp(-i\frac{E_n}{\hbar}t)$, where $\psi_n(x)$ satisfies the time-independent Schrödinger equation $\hat{H}\psi_n = E_n\psi_n$.
   d. The probability density ρ corresponding to the wave function $\Psi_n(x,t) = \psi_n(x)\exp(-i\frac{E_n}{\hbar}t)$ is time-independent.
8. For two electrons:
   a. any two-electron wave function has to be symmetric with respect to the exchange of the space and spin coordinates of these two electrons.
   b. the angle between two electron spins in the singlet state is equal to 180°.
   c. the electron spins in the triplet state are parallel.
   d. there are three triplet states that differ by total spin projection on the quantization axis.
9. In the Heisenberg uncertainty principle, ΔA, ΔB stand for the standard deviation determined for the measured quantities A and B:
   a. if $\hat{A}$ and $\hat{B}$ commute, the errors made when measuring A and B may be arbitrarily small.
   b. if $\hat{A}\hat{B} \neq \hat{B}\hat{A}$, then $\Delta A\,\Delta B \geqslant \frac{\hbar}{2}$.
   c. if $\hat{A}$ and $\hat{B}$ commute, $\Delta A\,\Delta B \geqslant 0$.
   d. one cannot make measurements with the calculated standard deviations ΔA and ΔB both equal to zero.
10. For the space (x, y, z) and spin (σ) coordinates for an electron:
   a. real numbers x, y, z determine the position of the electron in space, σ takes only two real values.
   b. there are two quantities that are measurable for a particle's spin: its length being a constant equal to $\sqrt{\frac{3}{4}}\hbar$ and one of the coordinates (chosen as z), which may take one of two values: $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$.
   c. the measurable quantities for a particle's spin $\mathbf{S}$ are its S_x, S_y, S_z components, which may take one of the two values: $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$.
   d. $\sum_\sigma \alpha(\sigma)\beta(\sigma) = 0$.

Answers

1c, 2abd, 3ac, 4d, 5bc, 6abd, 7abcd, 8bd, 9abc, 10abd



CHAPTER 2

The Schrödinger Equation

> *When the solution is simple, God is answering.*
> ***Albert Einstein***

Where are we?

The postulates constitute the foundation of quantum mechanics (and also the base of the TREE trunk). One of their consequences is the Schrödinger equation for stationary states. Thus we begin our itinerary on the TREE. The second part of this chapter is devoted to the time-dependent Schrödinger equation, which, from a pragmatic point of view, is outside the main theme of this book (this is why it is a side branch on the left side of the TREE).

An example

A string, a fundamental constituent of many musical instruments, vibrates after it is excited, creating a series of *standing* sound waves of well-defined frequencies (related to the string length) which then propagate in the air. Physicists of the beginning of the 20th century were puzzled by a similar phenomenon exhibited by atoms. However, the atoms (also molecules) emitted electromagnetic waves but also of well-defined frequencies (known as their spectra, see Fig. 2.1), and nobody knew how this might happen and why. It turned out that only quantum mechanics (not classical mechanics) was able to solve this puzzle; the key is to solve the all important Schrödinger equation – the subject of the present chapter. The Schrödinger equation for atoms and molecules describes *standing waves*, which is why it is also known as the wave equation.[1] Thus, atoms and molecules may be viewed as the musical instruments of the Universe.

What is it all about?

---

[1] From theory of music.

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00010-0

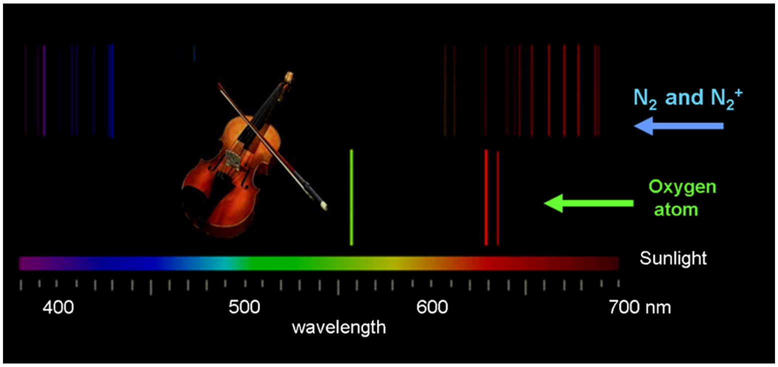


Fig. 2.1. Have you ever seen the fascinating aurora borealis? It is mostly green, but also pink and red, changing all the time. Only quantum mechanics is able to explain this beauty. It turned out that the corresponding emission spectrum comes mainly from the nitrogen molecule, the nitrogen molecule cation (upper panel, lower boreal altitude, pink and red colors of the aurora), and the oxygen atom (mostly green, lower panel). The striking feature of this picture is that the emitting objects exhibit a certain number of sharp monochromatic light lines of various intensities. The lines correspond to different wave lengths. Why do N_2, N_2^+, and O behave similarly to musical instruments like, e.g., the violin, which also emits waves (in this case acoustic) with some well-defined values of wave lengths? The colors of the aurora are explained in detail when one solves the Schrödinger equation for N_2, N_2^+, and O. The equation is also known as the wave equation, the same that describes the violin string vibrations.

- Meditation at a water spring
- Linearity

Evolution after switching a perturbation (Ⓢ) **p. 104**

- Time-independent perturbation – the two-state model
- Oscillating perturbation – the two-state model
- Short-time perturbation – the first-order approach
- Time-independent perturbation and the Fermi Golden Rule
- Oscillating perturbation and the Fermi Golden Rule

The *time-independent* Schrödinger equation is the one place where stationary states can be produced as solutions of the equation. The *time-dependent* Schrödinger equation plays the role of the "equation of motion," describing the evolution of a given wave function as time passes. As always for an equation of motion, one has to provide an initial state (starting point), i.e., the wave function for $t = 0$. Both the stationary states and the evolution of the nonstationary states depend on the total energy operator (Hamiltonian). If one finds some symmetry of the Hamiltonian, this will influence the symmetry of the wave functions. At the end of this chapter we will be interested in the evolution of a wave function after applying a perturbation.

Why is this important?

The wave function is a central notion in quantum mechanics, and is obtained as a solution of the Schrödinger equation. Hence this chapter is necessary for understanding quantum chemistry.

What is needed?

- Postulates of quantum mechanics, Chapter 1 (necessary),
- matrix algebra, Appendix A, p. 589 (advised),
- center-of-mass separation, Appendix J, p. 691 (necessary),
- translation versus momentum and rotation versus angular momentum, Appendix F, p. 665 (necessary),
- Dirac notation, p. 23 (necessary),
- two-state model, Appendix D, p. 655 (necessary),
- Dirac delta, Appendix E, p. 659 (necessary).

Classical works

A paper by the mathematician Emmy Noether, "*Invariante Variationsprobleme*," published in *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, 1918, p. 235–257, established a monumental link between the conservation laws for the energy, momentum and angular momentum and the symmetry properties of space and time. ★ Four papers by Erwin Schrödinger, which turned out to cause an "earthquake" in science, i.e., *Annalen der Physik*, 79(1926)361, *ibid.* 79(1926)489, *ibid.* 80(1926)437, and *ibid.* 81(1926)109, all under the title "*Quantisierung als Eigenwertproblem*," presented quantum mechanics as an eigenvalue problem (known from the developed differential equation theory), instead of an abstract Heisenberg algebra. Schrödinger proved the equivalence of both theories, gave the solution for the hydrogen atom, and introduced the variational principle. ★ The time-dependent perturbation theory described in this chapter was developed by Paul Dirac in 1926. Twenty years later, Enrico Fermi, lecturing at the University of Chicago, coined the term "The Golden Rule" for these results. Since then, they are known as the Fermi Golden Rule.

* * *

2.1 Symmetry of the nonrelativistic Hamiltonian and the conservation laws

> From classical mechanics it follows that for an isolated system (and assuming the forces to be central and obeying the action-reaction principle), its *energy*, *momentum*, and *angular momentum* are conserved.

Emmy Noether (1882–1935), German mathematician, informally professor, formally only the assistant of David Hilbert at the University of Göttingen (in the first quarter of the 20th century, women were not allowed to be professors in Germany). Her outstanding achievements in mathematics meant nothing to the Nazis, because Noether was Jewish (people should be reminded of such problems) and in 1933 Noether was forced to emigrate to the USA (Institute for Advanced Study in Princeton).

Imagine a well-isolated space ship observed in a space-fixed coordinate system. Its *energy is preserved*, its center of mass moves along a straight line with constant velocity (*the total, or center-of-mass, momentum vector is preserved*), and it preserves its *total angular momentum*.[2] The same is true for a molecule or atom, but the conservation laws have to be formulated in the language of quantum mechanics.

---

[2] That is, its *length and direction*. Think of a skater performing a spin: extending the arms sideways slows down the rotation, while stretching them along the axis of rotation results in faster rotation. *But all the time the total angular momentum vector is the same*. Well, what happens to the angular momentum, when the dancer finally stops rotating due to the friction? The angular momentum attains zero? No way. This is simply impossible. When the dancer increases its angular velocity, the Earth's axis changes a bit its direction to preserve the previous angular momentum of the total system (the Earth plus the dancer). When the dancer stops, the Earth's axis comes back towards its previous position, but not completely, because a part of the angular momentum is hidden in the rotation of molecules caused by the friction. Whatever happens, the total angular momentum has to be preserved! If the space ship captain wanted to stop the rotation of the ship which is making the crew sick, he could either throw something (e.g., gas from a steering jet) away from the ship, or spin a well-oriented body fast inside the ship. But even the captain is unable to change the *total* angular momentum.

Where do the conservation laws come from? Emmy Noether proved that they follow from the symmetry operations, with respect to which the equation of motion is invariant.[3]

> Thus, it turned out that invariance of the equation of motion with respect to an arbitrary:
> – *translation in time* (time homogeneity) results in the *energy conservation principle*;
> – *translation in space* (space homogeneity) gives the *total momentum conservation principle*;
> – *rotation in space* (space isotropy) implies the *total angular momentum conservation principle*.

These may be regarded as the foundations of science.[4] The homogeneity of time allows one to expect that repeating experiments give the same results. The homogeneity of space makes it possible to compare the results of the same experiments carried out in two different laboratories. Finally, the isotropy of space allows one to reject any suspicion that a different orientation of our laboratory bench changes the result.

Conservation laws represent most precious information about our system. It is not important what happens to the *isolated* system, what it is composed of, how complex the processes taking place in it are, whether they are slow or violent, whether there are people in the system or not, or whether they think about how to cheat the conservation laws; *nothing can violate the conservation of the energy, momentum, or angular momentum.*

Now, let us try to incorporate this into quantum mechanics.

---

[3] In case of a one-parameter family of operations $\hat{S}_\alpha \hat{S}_\beta = \hat{S}_{\alpha+\beta}$, e.g., translation ($\alpha$, β stand for the translation vectors), rotation (α, β are rotational angles), etc. Some other operations may not form such families and then the Noether theorem is no longer valid. This was an important discovery. Symmetry of a theory is much more fundamental than symmetry of an object. *The symmetry of a theory means that phenomena are described by the same equations no matter what laboratory coordinate system is chosen.*

[4] Well, to some extent. For example, the Universe does not show exact isotropy, because the matter there does not show spherical symmetry. Moreover, even if only one object were in the Universe, this very object would itself destroy the anisotropy of the Universe. We should rather think of this as a kind of idealization (approximation of the reality).

All symmetry operations (e.g., translation, rotation, reflection in a plane) are isometric, i.e., $\hat{U}^\dagger = \hat{U}^{-1}$ and $\hat{U}$ does not change distances between points of the transformed object (Figs. 2.2 and 2.3).

The operator $\hat{U}$ acting in three-dimensional Cartesian space corresponds to the operator $\hat{\mathcal{U}}$ acting in the Hilbert space; cf. Eq. (C.2), p. 608. Thus the function $f(\mathbf{r})$ transforms to $f' = \hat{\mathcal{U}} f = f(\hat{U}^{-1}\mathbf{r})$, while the operator $\hat{A}$ transforms to $\hat{A}' = \hat{\mathcal{U}}\hat{A}\hat{\mathcal{U}}^{-1}$ (Fig. 2.3). The formula for $\hat{A}'$ *differs* in general from $\hat{A}$, but *when it does not*, i.e., $\hat{A}' = \hat{A}$, then $\hat{\mathcal{U}}$ commutes with $\hat{A}$.

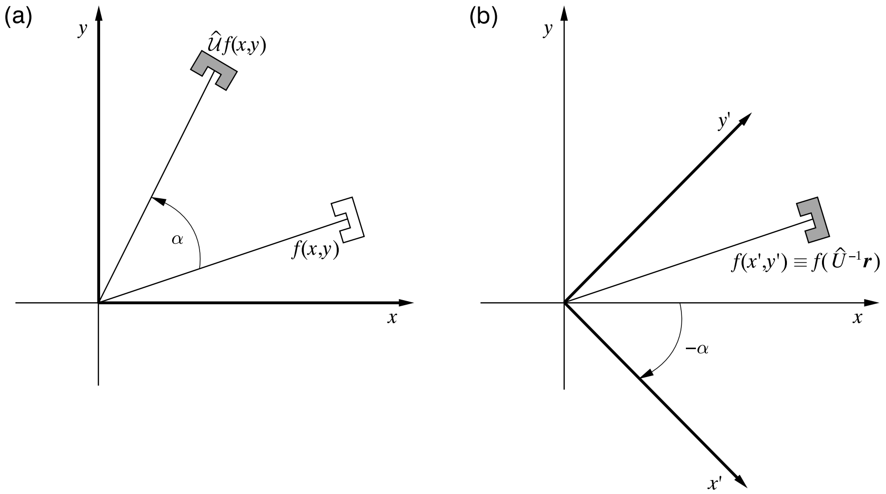


Fig. 2.2. (a) An *object* is rotated by angle α. (b) The *coordinate system* is rotated by angle $-\alpha$. The new position of the object in the old coordinate system (a) is the same as the initial position of the object in the new coordinate system (b).

Indeed, then $\hat{A} = \hat{\mathcal{U}}\hat{A}\hat{\mathcal{U}}^{-1}$, i.e., one has the commutation relation $\hat{A}\hat{\mathcal{U}} = \hat{\mathcal{U}}\hat{A}$, which means that $\hat{\mathcal{U}}$ and $\hat{A}$ share their eigenfunctions (Appendix B, p. 595).

Let us take the Hamiltonian $\hat{H}$ as the operator $\hat{A}$. Before writing it down let us introduce *atomic units*. Their justification comes from something similar to laziness. The quantities one calculates in quantum mechanics are stuffed up by some constants: $\hbar = \frac{h}{2\pi}$, where h is the Planck constant, the electron charge $-e$, the electron (rest) mass m_0, etc. These constants appear in clumsy formulae with various powers, in the nominator and denominator (see Table of units, p. 763). One always knows, however, that the quantity one com-

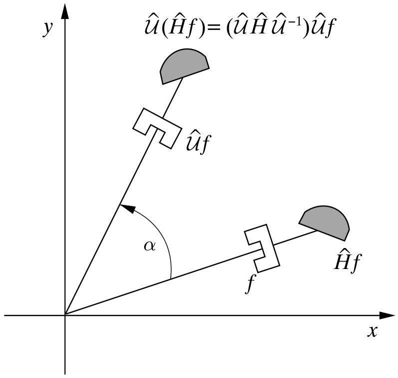


Fig. 2.3. A schematic view from the Hilbert space of functions $f(x, y)$. The f and $\hat{H}f$ represent, in general, different functions. Rotation (by α) of function $\hat{H}f$ gives function $\hat{\mathcal{U}}(\hat{H}f)$ and, in consequence, is bound to denote the *rotation of* f (i.e., $\hat{U}f$) and the transformation $\hat{U}\hat{H}\hat{U}^{-1}$ *of the operator* $\hat{H}$. Indeed, only then $\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1}$ when acting on the *rotated* function, i.e., $\hat{\mathcal{U}}f$ gives $\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1}(\hat{\mathcal{U}}f) = \hat{\mathcal{U}}(\hat{H}f)$, i.e., the rotated result of the Hamiltonian action. Because of $\hat{\mathcal{U}}(\hat{H}f) = (\hat{\mathcal{U}}\hat{H})(\hat{\mathcal{U}}f)$, when verifying the invariance of $\hat{H}$ with respect to transformation $\hat{\mathcal{U}}$, it is sufficient to check whether $\hat{\mathcal{U}}\hat{H}$ has the same formula as $\hat{H}$, but expressed in the new coordinates. Only this $\hat{\mathcal{U}}\hat{H}$ will fit to f expressed in the new coordinates, i.e., to $\hat{\mathcal{U}}f$. This is how we will proceed shortly.

putes is energy, length, time, etc., and how energy, length, etc., are expressed in $\hbar$, e, m_0, etc.

ATOMIC UNITS

If one inserts $\hbar = 1$, $e = 1$, $m_0 = 1$, this gives a dramatic simplification of the formulae. One has to remember though that these units have been introduced and, whenever needed, one can evaluate the result in other units (see Table of conversion coefficients, p. 764).

The Hamiltonian for a system of M nuclei (with charges Z_I and mass m_I, $I = 1, ..., M$) and N electrons, in the nonrelativistic approximation and assuming point-like particles without any internal structure,[5] takes in atomic units (a.u.) the following form (see p. 22):

---

[5] No internal structure of the electron has yet been discovered. The electron is treated as a point-like particle. Contrary to this, nuclei have a rich internal structure and nonzero dimensions.
A clear multilevel-like structure appears (which has to a large extent forced a similar structure on the corresponding scientific methodologies):

- Level I. A nucleon (neutron, proton) consists of three (the valence) quarks, clearly seen on the scattering image obtained for the proton. Nobody has yet observed a free quark.

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}, \tag{2.1}$$

where the kinetic energy operators for the nuclei and electrons (in a.u.) read as

$$\hat{T}_n = -\frac{1}{2}\sum_{I=1}^{M} \frac{1}{m_I}\Delta_I, \tag{2.2}$$

$$\hat{T}_e = -\frac{1}{2}\sum_{i=1}^{N} \Delta_i, \tag{2.3}$$

where the Laplacians are

$$\Delta_I = \frac{\partial^2}{\partial X_I^2} + \frac{\partial^2}{\partial Y_I^2} + \frac{\partial^2}{\partial Z_I^2},$$

$$\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2},$$

and x, y, z stand for the Cartesian coordinates of the nuclei and electrons indicated by vectors $\mathbf{R}_I = (X_I, Y_I, Z_I)$ and $\mathbf{r}_i = (x_i, y_i, z_i)$, respectively.

- Level II. The strong forces acting among nucleons have a range of about 1–2 fm (1 fm $= 10^{-15}$ m). Above 0.4–0.5 fm they are attractive, at shorter distances they correspond to repulsion. One does not need to consider their quark structure when computing the forces among nucleons, but they may be treated as particles *without internal structure*. The attractive forces between nucleons practically do not depend on the nucleon's charge and are so strong that they may overcome the Coulomb repulsion of protons. Thus nuclei composed of many nucleons (various chemical elements) may be formed, which in mean-field theory exhibit a shell structure (analogous to the electronic structure, cf. Chapter 8) related to the packing of the nucleons. The motion of the nucleons is strongly correlated. A nucleus may have various energy states (ground and excited), may be distorted, may undergo splitting, etc. About 2000 nuclei are known, of which only 270 are stable. The smallest nucleus is the proton, the largest known so far is ^{209}Bi (209 nucleons). The largest observed number of protons in a nucleus is 118. Even the largest nuclei have diameters about 100 000 times smaller than the electronic shells of the atom. Even for an atom with atomic number 118, the first Bohr radius is equal to $\frac{1}{118}$ a.u., or $5 \cdot 10^{-13}$ m, still about 100 times larger than the nucleus.
- Level III. Chemists *can* neglect the internal structure of nuclei. A nucleus *can* be treated as a structureless point-like particle and using the theory described in this book, one is able to predict extremely precisely virtually all the chemical properties of atoms and molecules. Some interesting exceptions will be given at the end of Chapter 6.

The operator $\hat{V}$ corresponds to the electrostatic interaction of all the particles (nucleus–nucleus, nucleus–electron, electron–electron)[6]:

$$\hat{V}=\sum_{I=1}^{M}\sum_{J>I}^{M}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|}-\sum_{I=1}^{M}\sum_{i=1}^{N}\frac{Z_I}{|\mathbf{r}_i-\mathbf{R}_I|}+\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}, \tag{2.4}$$

or, in a simplified form,

$$\hat{V}=\sum_{I=1}^{M}\sum_{J>I}^{M}\frac{Z_I Z_J}{R_{IJ}}-\sum_{I=1}^{M}\sum_{i=1}^{N}\frac{Z_I}{r_{iI}}+\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}}. \tag{2.5}$$

If the Hamiltonian turned out to be invariant with respect to a symmetry operation $\hat{U}$ (translation, rotation, etc.), this would imply the commutation of $\hat{U}$ and $\hat{H}$. We will check this in more detail below.

Note that the distances R_{IJ}, r_{iI}, and r_{ij} in the Coulombic potential energy (2.5) witness about assumption of instantaneous interactions in nonrelativistic theory (infinite travel speed of the interaction through space).

2.1.1 Invariance with respect to translation

Translation by vector $\mathbf{T}$ of function $f(\mathbf{r})$ in space means the function $\hat{\mathcal{U}} f(\mathbf{r}) = f\left(\hat{U}^{-1}\mathbf{r}\right) = f(\mathbf{r}-\mathbf{T})$, i.e., an opposite (by vector $-\mathbf{T}$) translation of the coordinate system (Fig. 2.4).

The transformation $\mathbf{r}' = \mathbf{r}+\mathbf{T}$ does not change the Hamiltonian. This is evident for the potential energy $\hat{V}$, because the translations $\mathbf{T}$ are canceled out, leaving the interparticle distances unchanged. For the kinetic energy one obtains $\frac{\partial}{\partial x'} = \sum_{\sigma=x,y,z}\frac{\partial\sigma}{\partial x'}\frac{\partial}{\partial\sigma} = \frac{\partial x}{\partial x'}\frac{\partial}{\partial x} = \frac{\partial}{\partial x}$, and all the kinetic energy operators (Eqs. (2.2) and (2.3)) are composed of the operators having this form.

> The Hamiltonian is therefore invariant with respect to any translation of the coordinate system.

There are two main consequences of translational symmetry:

---

[6] We do not include in this Hamiltonian tiny magnetic interactions of electrons and nuclei coming from their spin and orbital angular momenta, because they are of relativistic nature (see Chapter 3). In Chapter V2-4 we will be interested just in such small magnetic effects and the Hamiltonian will have to be generalized to include these interactions.

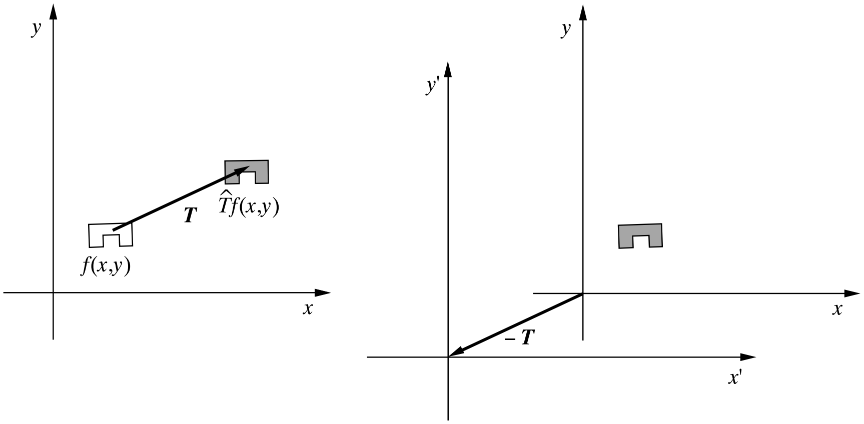


Fig. 2.4. A *function* f shifted by vector $\mathbf{T}$ (symmetry operation $\hat{T}$), i.e., $\hat{T} f(x, y)$ in the coordinate system (x, y) is the same as function $f\left(x', y'\right)$ in the *coordinate system* (x', y') shifted by $-\mathbf{T}$.

- No matter whether the coordinate system used is fixed in Trafalgar Square or in the center of mass of the system, one has to solve the same mathematical problem.
- The solution to the Schrödinger equation corresponding to the space-fixed coordinate system located in Trafalgar Square is Ψ_{pN}, whereas Ψ_{0N} is calculated in the body-fixed coordinate system (see Appendix J) located in the center of mass at $\mathbf{R}_{CM}$ that moves in the space-fixed coordinate system with the (total) momentum $\mathbf{p}_{CM}$. These two solutions are related by[7] $\Psi_{pN} = \Psi_{0N} \exp(i\mathbf{p}_{CM} \cdot \mathbf{R}_{CM})$. The number $N = 0, 1, 2, \ldots$ counts the energy states after the center-of-mass motion is separated.

> This means that the energy spectrum represents a continuum, because the center of mass may have any (nonnegative) kinetic energy $p_{CM}^2/(2m)$. If, however, one assumes that $p_{CM} = const$, then the energy spectrum is *discrete* for low-energy eigenvalues for atoms and molecules (see Eq. (1.13)).

This spectrum corresponds to the bound states, i.e., those states which do not correspond to any kind of dissociation (including ionization). Higher energy states lead to dissociation of the molecule, and the fragments may have any kinetic energy. Therefore, above the discrete spectrum one has a continuum of states. The states Ψ_{0N} will be called *spectroscopic states*. The bound states Ψ_{0N} are square integrable, as opposed to Ψ_{pN}, which are not because of function $exp(i\mathbf{p}\mathbf{R}_{CM})$, which describes the free motion of the center of mass.

---

[7] This follows from the separation of the center-of-mass motion (Appendix J) and noting that $\exp(i\mathbf{p}_{CM} \cdot \mathbf{R}_{CM})$ represents a solution for the motion of a free particle (Chapter 4).

2.1.2 Invariance with respect to rotation

The Hamiltonian is also invariant with respect to any rotation in space $\hat{U}$ of the coordinate system about a fixed axis. The rotation is carried out by applying an orthogonal matrix transformation[8] $\mathbf{U}$ of vector $\mathbf{r} = (x, y, z)^T$ that describes any particle of coordinates x, y, z. Therefore, all the particles undergo the same rotation and the new coordinates are $\mathbf{r}' = \hat{U}\mathbf{r} = \mathbf{U}\mathbf{r}$. Again, there is no problem with the potential energy, because a rotation does not change the interparticle distances. What about the Laplacians in the kinetic energy operators? Let us see. We have

$$\Delta = \sum_{k=1}^{3} \frac{\partial^2}{\partial x_k^2} = \sum_{k=1}^{3} \frac{\partial}{\partial x_k}\frac{\partial}{\partial x_k} = \sum_{k=1}^{3} \left(\sum_{i=1}^{3} \frac{\partial}{\partial x_i'}\frac{\partial x_i'}{\partial x_k} \right) \left(\sum_{i=1}^{3} \frac{\partial}{\partial x_i'}\frac{\partial x_i'}{\partial x_k} \right) =$$

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \left(\frac{\partial}{\partial x_i'}\frac{\partial x_i'}{\partial x_k} \right) \left(\frac{\partial}{\partial x_j'}\frac{\partial x_j'}{\partial x_k} \right) =$$

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \left(\frac{\partial}{\partial x_i'} U_{ik} \right) \left(\frac{\partial}{\partial x_j'} U_{jk} \right) = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} \left(\frac{\partial}{\partial x_i'} U_{ik} \right) \left(\frac{\partial}{\partial x_j'} U_{kj}^{\dagger} \right) =$$

$$\sum_{i=1}^{3}\sum_{j=1}^{3} \left(\frac{\partial}{\partial x_i'} \right) \left(\frac{\partial}{\partial x_j'} \right) \sum_{k=1}^{3} U_{ik} U_{kj}^{\dagger} =$$

$$\sum_{i=1}^{3}\sum_{j=1}^{3} \left(\frac{\partial}{\partial x_i'} \right) \left(\frac{\partial}{\partial x_j'} \right) \delta_{ij} = \sum_{k=1}^{3} \frac{\partial^2}{\partial \left(x_k'\right)^2} = \Delta'.$$

Thus, one has invariance of the Hamiltonian with respect to any rotation about the origin of the coordinate system. This means (see p. 665) that the Hamiltonian and the operator of the square of the total angular momentum $\hat{J}^2$ (as well as of one of its components, denoted by $\hat{J}_z$) commute. One is able, therefore, to measure simultaneously the energy, the square of the total angular momentum as well as one of the components of the total angular momentum, and (as will be shown in Section 4.9) one has ($\mathbf{r}$ and $\mathbf{R}$ denote the electronic and the nuclear coordinates, respectively)

$$\hat{J}^2 \Psi_N (\mathbf{r}, \mathbf{R}) = J(J+1)\hbar^2 \Psi_N (\mathbf{r}, \mathbf{R}), \tag{2.6}$$

$$\hat{J}_z \Psi_N (\mathbf{r}, \mathbf{R}) = M_J \hbar \Psi_N (\mathbf{r}, \mathbf{R}), \tag{2.7}$$

where $J = 0, 1, 2...$ and $M_J = -J, -J+1, ... + J$.

---

[8] $\mathbf{U}^T = \mathbf{U}^{-1}$.

Any rotation matrix may be shown as a product of "elementary" rotations, each about the x, y, or z axis. For example, rotation about the y axis by angle θ corresponds to the matrix $\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$. The pattern of such matrices is simple: one has to put in some places sines, cosines, zeros, and ones with the proper signs.[9] This matrix is orthogonal,[10] i.e., $\mathbf{U}^T = \mathbf{U}^{-1}$, which you may easily check. The product of two orthogonal matrices represents an orthogonal matrix; therefore, any rotation corresponds to an orthogonal matrix.

2.1.3 Invariance with respect to permutation of identical particles (fermions and bosons)

The Hamiltonian also has permutational symmetry. This means that if someone exchanged labels numbering the identical particles, independently of how it was done, he/she would always obtain the identical mathematical expression for the Hamiltonian. This implies that any wave function has to be either symmetric (for bosons) or antisymmetric (for fermions) with respect to the exchange of labels between two identical particles (cf. p. 39).

2.1.4 Invariance of the total charge

In addition to the energy, momentum, and angular momentum, strict conservation laws exist exclusively for total electric charge and the baryon and lepton numbers (a given particle contributes $+1$, the corresponding antiparticle -1).[11] The charge conservation (whatever happens) follows from the gauge symmetry: total electric charge conservation follows from the fact that a description of the system has to be invariant with respect to the mixing of the particle and antiparticle states, which is analogous to rotation.

---

[9] Clockwise and counterclockwise rotations and two possible signs at sines cause a problem with memorizing the right combination. In order to choose the correct one, one may use the following trick. First, we decide that what moves is an *object* (e.g., a function, not the coordinate system). Then, you have to have this book in your pocket. In Fig. 2.2a one sees that the rotation of the point with coordinates $(1, 0)$ by angle $\theta = 90°$ should give the point $(0, 1)$, and this is ensured only by the rotation matrix: $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

[10] And therefore also unitary (cf. Appendix A, p. 589).

[11] For example, in the Hamiltonian (2.1) it is assumed that whatever might happen to our system, the numbers of the nucleons and electrons will remain constant.

2.1.5 Fundamental and less fundamental invariances

The conservation laws described are of a fundamental character, because they are related to the homogeneity of space and time, the isotropy of space, and the nondistinguishability of identical particles.

Besides these strict conservation laws (energy, momentum, angular momentum, permutation of identical particles, charge, and baryon and lepton numbers), there are also some approximate laws. Two of these, parity and charge conjugation, will be discussed below. They are rooted in these strict laws, but are valid only in some conditions. For example, in most experiments, not only the baryon number, but also the number of nuclei of each kind is conserved. Despite the importance of this law in chemical reaction equations, this does not represent any strict conservation law, as shown by radioactive transmutations of elements.

Some other approximate conservation laws will soon be discussed.

2.1.6 Invariance with respect to inversion – parity

There are orthogonal transformations which are not equivalent to any rotation, e.g., the matrix of inversion $\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$, which corresponds to changing $\mathbf{r}$ to $-\mathbf{r}$ for all the particles, does not represent any rotation. If one performs such a symmetry operation, the Hamiltonian remains invariant and $|\Psi_{0N}(-\mathbf{r}, -\mathbf{R})|^2 = |\Psi_{0N}(\mathbf{r}, \mathbf{R})|^2$. This is evident, both for $\hat{V}$ (the interparticle distances do not change) and for the Laplacian (single differentiation changes sign, double does not). Two consecutive inversions mean identity operation: $\Psi_{0N}(\mathbf{r}, \mathbf{R}) = \exp(i\alpha)\Psi_{0N}(-\mathbf{r}, -\mathbf{R}) = [\exp(i\alpha)]^2\Psi_{0N}(\mathbf{r}, \mathbf{R})$. Hence, $[\exp(i\alpha)]^2 = 1$, $\exp(i\alpha) = \pm 1$ and one has

$$\Psi_{0N}(-\mathbf{r}, -\mathbf{R}) = \Pi\Psi_{0N}(\mathbf{r}, \mathbf{R}), \quad \text{where } \Pi \in \{1, -1\}.$$

Therefore,

> the wave function of a stationary state represents an eigenfunction of the inversion operator, and the eigenvalue can be either $\Pi = 1$ or $\Pi = -1$. This property is known as parity (P).

Now the reader will be taken by surprise. From what we have said, it follows that *no molecule has a nonzero dipole moment*. Indeed, the dipole moment is calculated as the mean value of the dipole moment operator (as usual in quantum mechanics, see Postulate IV in Chapter 1), i.e.,

$\boldsymbol{\mu} = \langle \Psi_{0N} | \hat{\boldsymbol{\mu}} \Psi_{0N} \rangle = \langle \Psi_{0N} | \left(\sum_i q_i \mathbf{r}_i \right) \Psi_{0N} \rangle$, where q's represent particles' charges. This integral will be calculated very easily: the integrand is antisymmetric with respect to inversion[12] and therefore $\boldsymbol{\mu} = \mathbf{0}$.

Is therefore the very meaning of the dipole moment, a quantity often used in chemistry and physics, a fairy tale? If HCl has no dipole moment, then it is more understandable that also H_2 does not. All this seems absurd.

Let us stress that our conclusion pertains to the *total* wave function, which has to reflect the space isotropy leading to the zero dipole moment, because all orientations in space are equally probable. If one applied the transformation $\mathbf{r} \to -\mathbf{r}$ only to *some* particles in the molecule (e.g., electrons) and not to others (e.g., the nuclei), the wave function will show no parity (it would be neither symmetric nor antisymmetric). We will introduce the Hamiltonian in Chapter 6, which corresponds to immobilizing the nuclei (the adiabatic or Born–Oppenheimer approximation) in certain positions in space, and in such a case the wave function depends on the electronic coordinates only. This wave function may be neither symmetric nor antisymmetric with respect to the *partial* inversion transformation $\mathbf{r} \to -\mathbf{r}$ (for the electrons only). To give an example, let us imagine an HF molecule in a coordinate system, its origin being in the middle between the H and F nuclei. Consider a particular configuration of the 10 electrons of the molecule; all close to the fluorine nucleus in some well-defined points. One may compute the value of the wave function for this configuration of electrons. Its square gives us the probability density of finding this particular configuration of electrons. Now, imagine the (*partial*) inversion $\mathbf{r} \to -\mathbf{r}$ applied to all the electrons. Now they will all be close to the proton. If one computes the probability density for the new situation, one would obtain a completely different value (much, much smaller, because the electrons prefer the fluorine, not the hydrogen), so we have no symmetry or antisymmetry. No wonder therefore that if one computed $\boldsymbol{\mu} = \langle \Psi_{0N} | \hat{\boldsymbol{\mu}} \Psi_{0N} \rangle$ with such a function (integration is over the electronic coordinates only), the result would differ from zero. This is why chemists believe the HF molecule has a nonzero dipole moment.[13] On the other hand, if the molecule taken as the example were B_2 (also ten electrons), then the two values have to be equal, because they describe the same physical situation. This corresponds, therefore, to a wave function with definite parity (symmetric or antisymmetric), and therefore, in this case $\boldsymbol{\mu} = \mathbf{0}$. This is why chemists believe molecules such as H_2, B_2, and O_2 have no dipole moment.

---

[12] Ψ_{0N} may be symmetric or antisymmetric, but $|\Psi_{0N}|^2$ is bound to be symmetric. Therefore, since $\sum_i q_i \mathbf{r}_i$ is antisymmetric, indeed, the integrand is antisymmetric (while the integration limits are symmetric).

[13] What therefore they do measure? The answer will be given in Chapter V2-4.

Product of inversion and rotation

The Hamiltonian is also invariant with respect to some other symmetry operations, like changing the sign of the x coordinates of all particles, or similar operations which are products of inversion and rotation. If one changed the sign of all the x coordinates, it would correspond to a mirror reflection. Since rotational symmetry stems from space isotropy (which we will treat as “trivial”), the mirror reflection may be identified with parity P.

Enantiomers

A consequence of inversion symmetry is that the wave functions have to be eigenfunctions of the inversion operator with eigenvalues $\Pi = 1$, i.e., the wave function is symmetric, or $\Pi = -1$, i.e., the wave function is antisymmetric. Any asymmetric wave function corresponding to a stationary state is therefore excluded (“illegal”). However, two optical isomers (enantiomers), corresponding to an object and its mirror image, do exist (Fig. 2.5).[14]

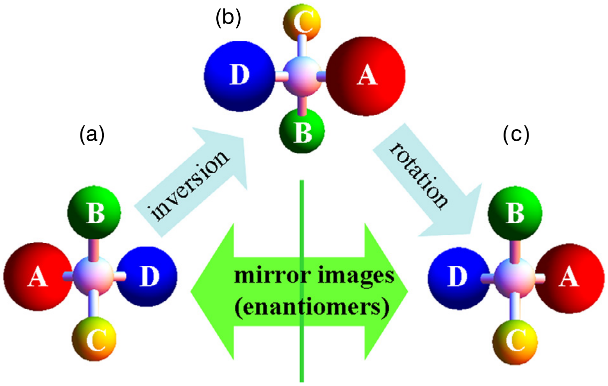


Fig. 2.5. Example of chirality. (a) A molecule’s central atom with four different side groups (A, B, C, and D) in a nonplanar configuration. (b) The same molecule after applying the inversion operation (with respect to the central atom). (c) An attempt at superposing the initial and the transformed molecules by rotation, to get matching, fails. (a) and (c) as well as (a) and (b) represent an example of a pair of enantiomers. Each of these isomers, after reflection in a mirror (like, e.g., the central atom-BC plane), becomes identical to its partner.

---

[14] The property that distinguishes them is known as chirality (your hands are an example of chiral objects). The chiral molecules (enantiomers) exhibit optical activity, i.e., polarized light passing through a solution of one of the enantiomers undergoes a rotation of the polarization plane *always in the same direction* (which may be easily seen by reversing the direction of the light beam). The enantiomeric molecules have the same properties, provided one is checking this by using nonchiral objects. If the probe were chiral, one of the enantiomers would interact with it differently (for purely sterical reasons). Enantiomers (e.g., molecular associates) may be formed from chiral or nonchiral subunits.

We ask in a pharmacy for D-glucose. Strangely enough, the pharmacist is fully cooperative and does not make trouble. We pay a small sum and he gives us something which should not exist[15] – a substance with a single enantiomer. We should obtain a substance composed of molecules in their stationary states, which therefore have to have definite parity, either as a sum of the wave functions for the two enantiomers D and L ($\Pi = 1$, cf. Appendix D on p. 655, Case I), i.e., $\psi_+ = \psi_D + \psi_L$, or as the difference ($\Pi = -1$), i.e., $\psi_- = \psi_D - \psi_L$. The energies corresponding to ψ_+ and ψ_- differ, but the difference is extremely small (quasidegeneracy). The brave shopkeeper has given us something with the wave function $\psi = N(\psi_+ + \psi_-) = \psi_D$ (as result of decoherence), which therefore describes a nonstationary state.[16] As we will see in a moment (p. 107), the approximate lifetime of the state is proportional to the inverse of the integral $\left\langle \psi_D | \hat{H} \psi_L \right\rangle$. If one computed this integral, one would obtain an extremely small number.[17] It would turn out that the pharmacy could safely keep the stock of glucose for millions of years. Maybe the reason for decoherence is interaction with the rest of the Universe, maybe even interaction with a vacuum. The very existence of enantiomers, or even prevailing of one of them on Earth, does not mean breaking parity symmetry. This would happen if one of the enantiomers corresponded to a lower energy than the other.[18]

---

[15] More exactly, should be unstable.

[16] Only ψ_+ and ψ_- are stationary states.

[17] This is seen even after attempting to overlap two molecular models physically (Fig. 2.5), they do not match. The overlap of the wave functions will be small for the same reasons (the wave functions decay exponentially with distance).

[18] This is what happens in reality, although the energy difference is extremely small. Experiments with β-decay have shown that Nature breaks parity in weak interactions. The parity conservation law therefore has an approximate character.

With no preference for any of the enantiomers, one of them may *spontaneously* increase its concentration until 100%, i.e., with a complete elimination of the other one (0%)! Is something like that possible at all? It seems it contradicts common sense, since one of the enantiomers won, while the other lost, whereas their chances were exactly equal!

This phenomenon occurs in reality if autocatalysis is involved. The key information is the following: in such a system a *large random and self-augmenting fluctuation* is possible. Indeed, let us imagine just for simplicity 50 molecules of D and 50 molecules of L together with a certain number of molecules N (let us call them "neutral"), with the following possible reactions (giving equal chances to D and L):

$$\mathrm{D} + \mathrm{L} \rightarrow 2\mathrm{N}$$
$$\mathrm{D} + \mathrm{N} \rightarrow 2\mathrm{D}$$
$$\mathrm{L} + \mathrm{N} \rightarrow 2\mathrm{L}.$$

The last two reactions (of autocatalytic character) represent in fact inducing an enantiomer ("forcing chirality") through interaction of N with a molecule of a given enantiomer. No preference of D or L is assumed in this induction. Nevertheless one of the enantiomers will defeat the other one! To explain this, let us assume that the elementary reactions 1, 2 or 3 of the individual molecules form a random chain in time, like, for example, 1, 2, 2, 3, 1, 3, ... If the chain starts by the first reaction, we have a situation similar to the starting one (49 : 49 instead of 50 : 50, in number of D:L molecules). If, however, the chain starts by chance from the second reaction,

Chen Ning Yang (b. 1922) and Tsung Dao Lee (b. 1926), American physicists, professors at the Advanced Study Institute in Princeton, predicted in 1956 parity breaking in the weak interactions, which a few months later was confirmed experimentally by Madam Wu. In 1957 Yang and Lee received the Nobel Prize "*for their penetrating investigation of parity laws, which led to important discoveries regarding elementary particles*."

2.1.7 Invariance with respect to charge conjugation

If one changed the signs of the charges of all particles, the Hamiltonian would not change.

---

there will be a 51 : 50 preference of D, while when the third reaction takes place one will have 50 : 51 preference of L. In the last two cases we have a deviation from the equal chances of two enantiomers (fluctuation). Suppose we have the second case, i.e., 51 : 50 preference of D. Note that *now, when continuing the reaction chain at random, the chance for reaction 2 to happen is greater than the chance of reaction 3*. It is seen therefore that any fluctuation from an "equal-chances situation" has a tendency to self-augment, although there is always a chance that the fluctuation disappears. The racemic mixture (with the important N molecules though) represents therefore an instable system. Sooner or later there will be a transition (first mild but then more and more violent) from racemate to the absolute prevailing of one of the enantiomers. Their chances were and are equal, but just by chance one of them won.

Sometimes random processes may lead to large fluctuations. For example, Louis Pasteur was able to crystallize the crystals of one enantiomer and the crystals of the other enantiomer from a racemic solution. This is possible not only without Louis Pasteur, but also without any human, just spontaneously. Suppose that two such pieces of crystals were created where the primordial life was to be started, and that afterwards a volcano explosion took place, in which the volcanic lava covered (or even destroyed) one of the crystals, eliminating it from anything that happened later on (e.g., the chiral induction during biological evolution). This simple argument, proposed by Leszek Stolarczyk, is worth to be considered in discussions on how it comes that in living nature there is a strong preference for one chirality.

This therefore corresponds to exchanging particles and antiparticles.[19] Such a symmetry operation is called the charge conjugation and is denoted as C symmetry. This symmetry will be not marked in the wave function symbol (because, as a rule, we have to do with matter, not antimatter), but we will remember it. Sometimes it may turn out unexpectedly to be useful (see Chapter V2-5, p. V2-367). After Wu's experiment, physicists tried to save the hypothesis that what is conserved is the CP symmetry, i.e., the product of charge conjugation and inversion. However, analysis of experiments with decay of K mesons has shown that even this symmetry is approximate (although the deviation is extremely small).

2.1.8 Invariance with respect to the symmetry of the nuclear framework

In many applications the positions of the nuclei are fixed (clamped nuclei approximation, Chapter 6), often in a high-symmetry configuration (cf. Appendix C, p. 605). For example, the benzene molecule in its ground state (after minimizing the energy with respect to the positions of the nuclei) has the symmetry of a regular hexagon. In such cases the electronic Hamiltonian additionally exhibits invariance with respect to some symmetry operations and therefore the wave functions are the eigenstates of these molecular symmetry operations. Therefore, any wave function may have an additional label: the symbol of the irreducible representation[20] it belongs to.

2.1.9 Conservation of total spin

In an isolated system the *total* angular momentum **J** is conserved. However, $\mathbf{J} = \mathbf{L} + \mathbf{S}$, where **L** and **S** stand for the orbital and spin angular momenta (sum over all particles), respectively. The spin angular momentum **S**, being a sum over all particles, is not conserved.

However, the (nonrelativistic) Hamiltonian does not contain any spin variables. This means that it commutes with the operator of the square of the total spin as well as with the operator of one of the spin components (by convention the z component). Therefore, in the nonrelativistic approximation one can simultaneously measure the energy E, the square of the spin S^2, and one of its components: S_z.

---

[19] Somebody thought he had carried out computations for benzene, but he also computed antibenzene. The wave functions for benzene and antibenzene are the same.

[20] Of the symmetry group composed of the symmetry operations mentioned above, see Appendix C.

2.1.10 Indices of spectroscopic states

In summary, assumptions about the homogeneity of space and time, isotropy of space, and parity conservation lead to the following quantum numbers (indices) for the spectroscopic states:

- N quantizes energy,
- J quantizes the length of total angular momentum,
- M quantizes the z component of total angular momentum,
- Π determines parity:

$$\boxed{\Psi_{N,J,M,\Pi}(\mathbf{r},\mathbf{R}).}$$

Besides these indices following from the fundamental laws (in the case of parity it is a little too exaggerated), there may be also some indices related to less fundamental conservation laws:

- the irreducible representation index of the symmetry group of the clamped nuclei Hamiltonian (Appendix C),
- the values of S^2 (traditionally one gives the *multiplicity* $2S+1$) and S_z.

2.2 Schrödinger equation for stationary states

It may be instructive to see how Erwin Schrödinger invented his famous equation (1.13) for *stationary states* ψ of energy E ($\hat{H}$ denotes the Hamiltonian of the system),

$$\hat{H}\psi = E\psi. \tag{2.8}$$

Schrödinger surprised the contemporary quantum elite (associated mainly with Copenhagen and Göttingen) by a clear formulation of quantum mechanics as wave mechanics. January 27, 1926, when Schrödinger submitted a paper entitled "*Quantisierung als Eigenwertproblem*"[21] to *Annalen der Physik,* may be regarded as the birthday of wave mechanics.

Most probably, Schrödinger's reasoning was as follows. De Broglie discovered that what people called a particle also had a wave nature (Chapter 1). That is really puzzling. If a wave is involved, then according to Debye's suggestion at the November seminar in Zurich, it might be possible to write the standing wave equation with $\psi(x)$ as its amplitude at position x, i.e.,

$$v^2\frac{d^2\psi}{dx^2} + \omega^2\psi = 0, \tag{2.9}$$

---

[21] Quantization as an eigenproblem. Well, once upon a time quantum mechanics was discussed in German. Some traces of that period remain in the nomenclature. One is the "*eigen*value problem or *eigen*problem," which is a German-English hybrid.

Erwin Schrödinger (1887–1961), Austrian physicist, professor at Jena, Stuttgart, Graz, Breslau, Zurich, Berlin, and Vienna.

In later years Schrödinger recalled the Zurich period most warmly, in particular, discussions with the mathematician Hermann Weyl and the physicist Peter Debye. In 1927 Schrödinger succeeded Max Planck as the Chair of Theoretical Physics at the University of Berlin, and in 1933 he received the Nobel Prize "*for the discovery of new productive forms of atomic theory*." Hating the Nazi regime, he left Germany in 1933 and moved to the University of Oxford. However, homesick for his native Austria he went back in 1936 and took a professorship at the University of Graz. Meanwhile, Hitler carried out his *Anschluss* with Austria in 1938, and Schrödinger, even though not a Jew, could have been an easy target as he had fled Germany because of the Nazis. He emigrated to the USA (Princeton), and then to Ireland (Institute for Advanced Studies in Dublin), worked there till 1956, and then returned to Austria, where he worked at Vienna University until his death. In his scientific work as well as in his personal life Schrödinger did not strive for big goals; he worked by himself. Maybe what characterizes him best is that he was always ready to leave having belongings ready in his rucksack. Among the goals listed in the preface of this textbook there is no demoralization of youth. This is why I will stop here, limit myself to the carefully selected information given above and refrain from describing the circumstances in which quantum mechanics was born. For those students who read the recommendations in the Additional Literature, I provide some useful references: W. Moore, "*Schrödinger: Life and Thought*," Cambridge University Press, 1989, and the comments on the book given by P.W. Atkins, *Nature*, 341(1989), also http://www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Schrodinger.html.

Schrödinger's curriculum vitae found in Breslau (now Wrocław):

"*Erwin Schrödinger, born on Aug., 12, 1887 in Vienna, the son of the merchant Rudolf Schrödinger and his wife née Bauer. The family of my father comes from the Upper Palatinate and Wirtemberg region, and the family of my mother from German Hungary and (from the maternal side) from England. I attended a so-called "academic" high school (once part of the university) in my native town. Then during 1906–1910 I studied physics at Vienna University, where I graduated in 1910 as a doctor of physics. I owe my main inspiration to my respectable teacher Fritz Hasenöhrl, who by an unlucky fate was torn from his diligent students – he fell gloriously as an attack commander on the battlefield of Vielgereuth. As well as Hasenöhrl, I owe my mathematical education to Professors Franz Mertens and Wilhelm Wirtinger, and a certain knowledge of experimental physics to my principal of many years (1911–1920) Professor Franz Exner and my intimate friend R.M.F. Rohrmuth. A lack of experimental and some mathematical skills oriented me basically towards theory. Presumably the spirit of Ludwig Boltzmann (deceased in 1906), operating especially intensively in Vienna, directed me first towards probability theory in physics. Then, (...) a closer contact with experimental works of Exner and Rohrmuth* *oriented me to the physiological theory of colors, in which I tried to confirm and develop the achievements of Helmholtz. In 1911–1920 I was a laboratory assistant under Franz Exner in Vienna, of course, with a $4\frac{1}{2}$-year-long pause caused by war. I have obtained my habilitation in 1914 at the University of Vienna, and in 1920 I accepted an offer from Max Wien and become his assistant professor at the new theoretical physics department in Jena. This lasted, unfortunately, only one semester, because I could not refuse a professorship at the Technical University in Stuttgart. I was there also only one semester, because in April 1921 I came to the University of Hessen in succession to Klemens Schrafer. I am almost ashamed to confess that at the moment I sign the present curriculum vitae I am no longer a professor at the University of Breslau, because on Oct. 15 I received my nomination to the University of Zurich. My instability may be recognized exclusively as a sign of my ingratitude!"*

Breslau, Oct. 5, 1921.
Dr Erwin Schrödinger

(Found in the archives of the University of Wrocław (Breslau) by Professor Zdzisław Latajka and Professor Andrzej Sokalski, translated by Andrzej Kaim and the Author. Since the manuscript was hardly legible due to Schrödinger's difficult hand-writing, some names may have been misspelled.)

where v stands for the (phase) velocity of the wave and ω represents its angular frequency ($\omega = 2\pi\nu$, where ν is the usual frequency) which is related to the wave length λ by the well-known formula[22]

$$\omega/v = \frac{2\pi}{\lambda}. \tag{2.10}$$

Besides, Schrödinger knew from de Broglie, who had lectured in Zurich about this, that the wave length λ is related to a particle's momentum p through $\lambda = h/p$, where $h = 2\pi\hbar$ is the Planck constant. This equation is the most famous achievement of de Broglie, and relates the corpuscular (p) character and the wave (λ) character of any particle.

On the other hand the momentum p is related to the total energy (E) and the potential energy (V) of the particle through $p^2 = 2m(E - V)$, which follows from the expression for the kinetic energy $T = \frac{mv^2}{2} = p^2/(2m)$ and $E = T + V$. Therefore, Eq. (2.9) can be rewritten as

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + \frac{1}{\hbar^2}[2m(E - V)]\psi = 0. \tag{2.11}$$

The most important step towards the great discovery was the transfer of the term with E to the right-hand side. Let us see what Schrödinger obtained:

$$[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V]\psi = E\psi. \tag{2.12}$$

---

[22] In other words, $\nu = \frac{v}{\lambda}$ or $\lambda = vT$ (i.e., wave length is equal to the velocity times the period). Eq. (2.9) represents an oscillating function $\psi(x)$. Indeed, it means that $\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}$ and ψ differ by sign, i.e., if ψ is above the x axis, then it curves down, while if it is below the x axis, then it curves up.

This was certainly a good moment for a discovery. Schrödinger obtained a kind of *eigenvalue equation* (1.13), recalling his experience with eigenvalue equations in the theory of liquids.[23] What is striking in (2.12) is the odd fact that an operator $-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}$ amazingly plays the role of the kinetic energy. Indeed, keeping calm we see the following: *something* plus potential energy, all that multiplied by ψ, equals total energy times ψ. Therefore, clearly this *something* must be the kinetic energy! But wait a minute, the kinetic energy is equal to $\frac{p^2}{2m}$. From this it follows that in the equation obtained, instead of p there is a certain *operator* $i\hbar\frac{d}{dx}$ or $-i\hbar\frac{d}{dx}$, because only then does the squaring give the right answer.

Hermann Weyl (1885–1955), German mathematician, professor at ETH Zurich, then the University of Göttingen and the Institute for Advanced Studies at Princeton (USA), expert in the theory of orthogonal series, group theory, and differential equations. Weyl adored Schrödinger's wife, was a friend of the family, and provided an ideal partner for Schrödinger in conversations about the eigenvalue problem.

Would the key to the puzzle be simply taking the classical expression for total energy and inserting the above operators instead of the momenta? What was the excited Schrödinger supposed to do? The best choice is always to begin with the simplest toys, such as the free particle, the particle in a box, the harmonic oscillator, the rigid rotator, or the hydrogen atom. Nothing is known about whether Schrödinger himself had a sufficiently deep knowledge of mathematics to be able to solve the (sometimes nontrivial) equations related to these problems, or whether he had advice from a friend versed in mathematics, such as Hermann Weyl.

It turned out that instead of p, $-i\hbar\frac{d}{dx}$ had to be inserted, and not $i\hbar\frac{d}{dx}$ (Postulate II, Chapter 1).

2.2.1 Wave functions of class Q

The postulates of quantum mechanics, especially the probabilistic interpretation of the wave function given by Max Born, limit the class of functions allowed ("class Q", or "quantum").

Any wave function

- cannot be zero everywhere (Fig. 2.6a), because the system *is somewhere* in space;

---

[23] Very interesting coincidence: Heisenberg was also involved in fluid dynamics. At the beginning, Schrödinger did not use operators. They appeared after he established closer contacts with the University of Göttingen.

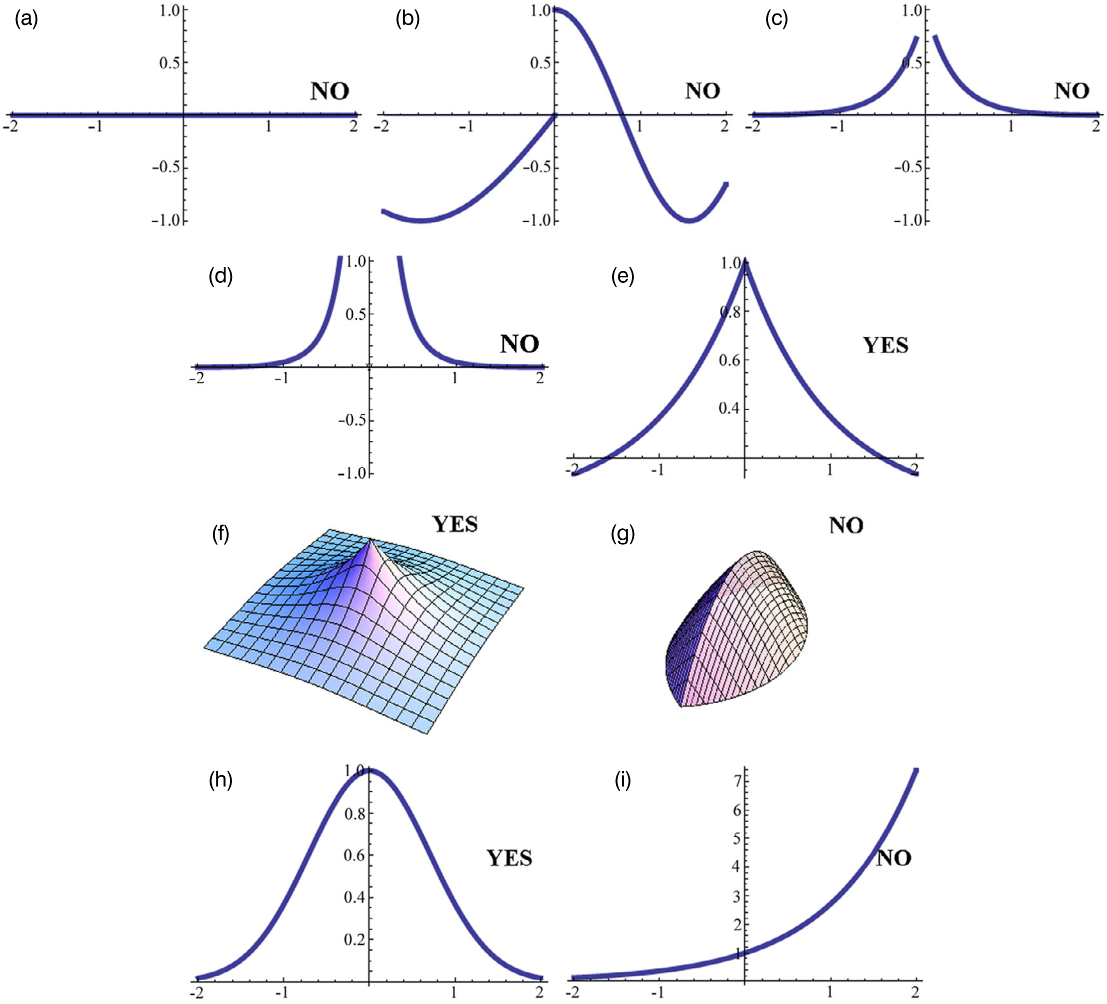


Fig. 2.6. Functions of class Q (i.e., wave functions allowed in quantum mechanics) – examples and counterexamples. A wave function (a) must not be zero everywhere in space; (b) has to be continuous; (c) cannot tend to infinity even at a single point; and (d) cannot tend to infinity. (e,f,g) Its first derivative cannot be discontinuous for an infinite number of points. (h,i) It must be square integrable, and (j,k,l,m) it has to be defined uniquely in space (for angular variable θ).

- has to be continuous, (Fig. 2.6b); this also means it cannot take infinite values at any point in space[24] (Fig. 2.6c,d);

---

[24] If this happened in any nonzero volume of space (Fig. 2.6d) the probability would tend to infinity (which is prohibited). However, the requirement is stronger than that: a wave function cannot take an infinite value even at a single point (Fig. 2.6c). Sometimes such functions appear among the solutions of the Schrödinger equation, and they have to be rejected. The formal argument is that, if not excluded from the domain of the Hamiltonian,

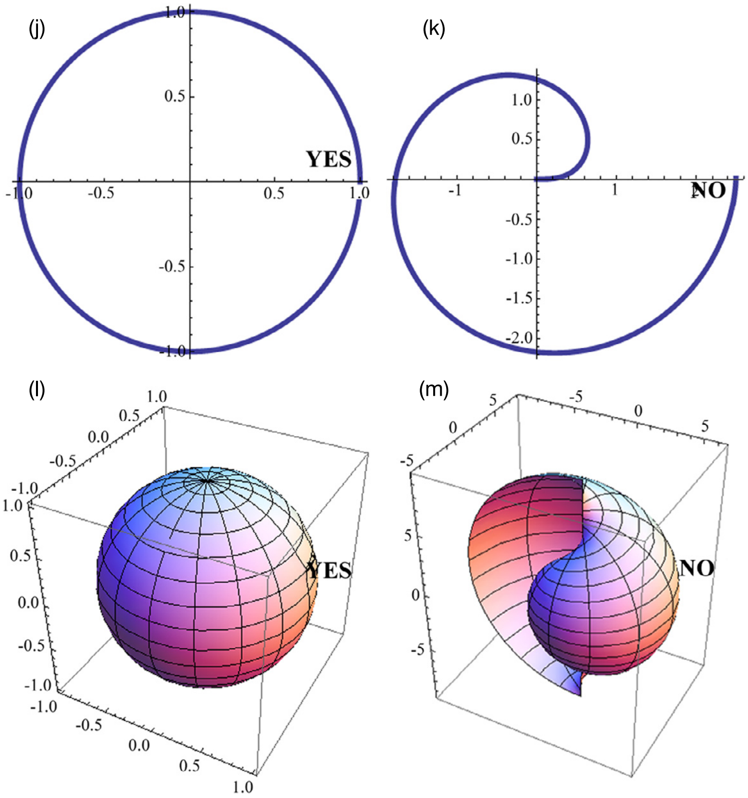


Fig. 2.6. (*continued*)

- has to have a continuous first derivative as well (everywhere in space except isolated points (Fig. 2.6e,f,g), where the potential energy tends to $-\infty$), because the Schrödinger equation is a second-order differential equation and the second derivative must be defined;
- has to have a uniquely defined value in space[25] (Fig. 2.6j,k,l,m);
- for bound states has to tend to zero at infinite values of any of the coordinates (Fig. 2.6h,i,k,m), because such a system is compact and does not disintegrate in space; in consequence (from the probabilistic interpretation), the wave function is square integrable, i.e., $\langle\Psi|\Psi\rangle < \infty$.

---

the latter would be non-Hermitian when such a function was involved in $\left\langle f|\hat{H}g\right\rangle = \left\langle \hat{H}f|g\right\rangle$. A non-Hermitian Hamiltonian might lead to complex energy eigenvalues, which is prohibited.

[25] At any point in space the function has to have a single value. This plays a role only if we have an angular variable, say ϕ. Then, ϕ and $\phi + 2\pi$ have to give the same value of the wave function. We will encounter this problem in the solution for the rigid rotator.

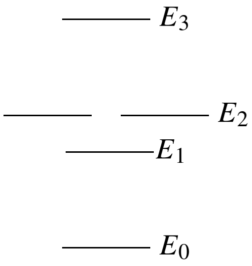
2.2.2 Boundary conditions

The Schrödinger equation is a differential equation. In order to obtain a special solution to such equations, one has to insert the particular boundary conditions to be fulfilled. Such conditions follow from the physics of the problem, i.e., with which kind of experiment we are going to compare the theoretical results. For example:

- for the *bound states* (i.e., square integrable states) we put the condition that the wave function has to vanish at infinity, i.e., if any of the coordinates tends to infinity, $\psi(x=\infty)=\psi(x=-\infty)=0$;
- for cyclic systems of circumference L, the natural conditions will be $\psi(x)=\psi(x+L)$ and $\psi'(x)=\psi'(x+L)$, because they ensure a smooth matching of the wave function for $x<0$ and of the wave function for $x>0$ at $x=0$;
- for scattering states (not discussed here) the boundary conditions are more complex.[26]

There is a countable number of bound states. Each state corresponds to eigenvalue E.

An energy level may be *degenerate*, that is, more than one wave function may correspond to it, all the wave functions being linearly independent (their number is the *degree of degeneracy*). The eigenvalue spectrum is usually represented by putting a single horizontal section (in the energy scale) for each wave function:



An analogy

Let us imagine all the stable positions of a chair on the floor (Fig. 2.7).

Consider a simple chair, very uncomfortable for sitting, but very convenient for a mathematical exercise. Each of the four legs represents a rod of length a, the "seat" is simply a square built of identical rods, and the back consists of three such rods making a C shape. The potential energy of the chair (in position i) in gravitational field equals mgh_i, where m stands for the mass of the chair, g is gravitational acceleration, and h_i denotes the height of the center of mass with respect to the floor. We obtain the following energies, E_i, of the stationary states (in units of mga):

---

[26] J.R. Taylor, "*Scattering Theory*," Wiley, New York, 1972 is an excellent reference.

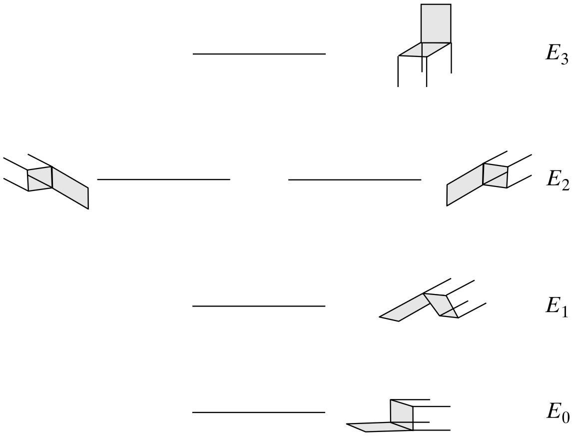


Fig. 2.7. The stable positions of a chair on the floor. In everyday life we most often use the third excited state.

- The chair is lying on the support: $E_0 = \frac{4}{11}$.
- The chair is lying inclined, so the support and the seat touch the floor: $E_1 = \frac{7\sqrt{2}}{22}$.
- The chair is lying on the side: $E_2 = \frac{1}{2}$. Note, however, that we have *two* sides. The energy is the same for the chair lying on the first and second side (because the chair is symmetric), but these are *two* states of the chair, not one. The degree of degeneracy equals *two,* and therefore on the energy diagram we have two horizontal sections. Note how naturally the problem of degeneracy has appeared. The degeneracy of the energy eigenstates of molecules results from their symmetry, exactly as in the case of the chair. In some cases, one may obtain an *accidental degeneracy* (cf. p. 233), which does not follow from the symmetry of an object like a chair, but from the properties of the potential field, and is called *dynamic symmetry*.[27]
- The chair is in the normal position: $E_3 = 1$.

There are no more stable states of the chair and there are only four energy levels (Fig. 2.7). The stable states of the chair are analogs of the stationary quantum states of Fig. 1.8a,b, on p. 27, while unstable states of the chair on the floor are analogs of the nonstationary states of Fig. 1.8c,d. Of course, there are plenty of unstable positions of the chair with respect to the floor. The stationary states of the chair have more in common with chemistry than we might think. A chair-like molecule (organic chemists have already synthesized much more complex molecules) interacting with a crystal surface would very probably have similar stationary states (Fig. 2.8).

---

[27] See the original works C. Runge, "*Vektoranalysis*", vol. I, p. 70, ed. S. Hirzel, Leipzig, 1919, W. Lenz, *Zeit. Physik,* 24(1924)197, and L.I. Schiff, "*Quantum Mechanics*," McGraw Hill, 1968.

Note that a chair made by a very impractical cabinet maker, a fan of surrealism (in our analogy a strange chair-like molecule: just another pattern of chemical bonds, i.e., another electronic state), will result in a different set of energy levels for such a weird chair. Thus, we see that the vibrational levels depend in general on the electronic state.

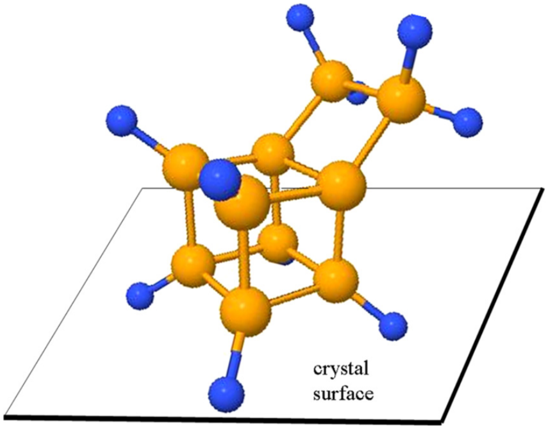


Fig. 2.8. A quantum mechanical analogy of the stable positions of a chair on the floor. A stiff molecule $C_{10}H_{10}$ with the shape shown above, when interacting with a crystal surface, would acquire several stable positions similar to those of the chair on the floor. They would correspond to some vibrational states (the molecule would vibrate about these positions) of a given electronic state ("the same bond pattern"), which in this analogy would correspond to the fixed structure of the chair.

2.2.2.1 Mathematical and physical solutions

It is worth noting that not all solutions of the Schrödinger equation are physically acceptable.

For example, for bound states, all other solutions than those of class Q (see p. 595) must be rejected. Also, the solution ψ, which does not exhibit the proper symmetry, even if $|\psi|^2$ does, has to be rejected as well. In particular, such illegal, nonacceptable functions are *asymmetric* with respect to the label exchange for electrons (e.g., symmetric for some pairs and antisymmetric for others). Also, a fully symmetric function would also be such a nonphysical (purely mathematical) solution. They are called mathematical, but nonphysical, solutions to the Schrödinger equation. Sometimes such mathematical solutions correspond to a *lower* energy than any physically acceptable energy (in such a case they are called the *underground states*).

2.3 The time-dependent Schrödinger equation

What would happen if one prepared the system in a given state ψ, which does not represent a stationary state? For example, one may deform a molecule by using an electric field and then switch the field off.[28] The molecule will turn out to be suddenly in state ψ, which is not its stationary state. Then, according to quantum mechanics, the state of the molecule will start to change according to the time evolution equation (time-dependent Schrödinger equation)

$$\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}. \tag{2.13}$$

The equation plays a role analogous to Newton's equation of motion in classical mechanics. In Newton's equation the position and momentum of a particle evolve. In the time-dependent Schrödinger equation the evolution proceeds in a completely different space – in the space of states or the Hilbert space (cf. Appendix B, p. 595).

Therefore, in quantum mechanics one has absolute determinism, but in the state space. Indeterminism begins only in our space, when one asks about the coordinates of a particle.

2.3.1 Evolution in time

As seen from Eq. (2.13) knowledge of the Hamiltonian and of the wave function ψ at a given time $t = 0$ (left-hand side) represents sufficient information to determine the time derivative of the wave function (right-hand side). This means that we may compute the wave function after an infinitesimal time dt as follows:

$$\psi + \frac{\partial\psi}{\partial t}\mathrm{d}t = \psi - \frac{i}{\hbar}\hat{H}\psi\mathrm{d}t = [1 + (-i\frac{t}{N\hbar})\hat{H}]\psi,$$

where we have set $\mathrm{d}t = t/N$ with N (natural number) very large. Thus, the new wave function results from action of the operator $[1 + (-i\frac{t}{N\hbar})\hat{H}]$ on the old wave function. Now, we may pretend that we did not change any function and apply the operator again and again. We assume that $\hat{H}$ is time-independent. The total operation is nothing but the action of the operator

$$\lim_{N\to\infty}[1 + (-i\frac{t}{N\hbar})\hat{H}]^N.$$

Please recall that $e^x = \lim_{N\to\infty}[1 + \frac{x}{N}]^N$.

---

[28] We neglect the influence of the magnetic field that accompanies any change of electric field.

Hence, the time evolution corresponds to action on the initial ψ of the operator $\exp(-\frac{it}{\hbar}\hat{H})$,

$$\psi' = \exp(-\frac{it}{\hbar}\hat{H})\psi. \tag{2.14}$$

Quantity $exp\hat{A}$ is defined through the Taylor expansion $e^{\hat{A}} = 1 + \hat{A} + \hat{A}^2/2 + \ldots.$

Our result satisfies the time-dependent Schrödinger equation[29] if $\hat{H}$ does not depend on time (as we assumed when constructing ψ').

Inserting the spectral resolution of the identity[30] (cf. Postulate II in Chapter 1) one obtains[31]

$$\psi(t) = \exp(-i\frac{t}{\hbar}\hat{H})1\psi(t=0) = \exp(-i\frac{t}{\hbar}\hat{H})\sum_n |\psi_n\rangle\langle\psi_n|\psi(0)\rangle = \tag{2.15}$$

$$= \sum_n c_n \exp(-i\frac{t}{\hbar}E_n)|\psi_n\rangle, \tag{2.16}$$

where $c_n = \langle\psi_n|\psi(0)\rangle$ are some time-independent coefficients. This is how the state ψ (known as *wave packet*) evolves. Note that *the only time dependence is in the* $\exp(-i\frac{t}{\hbar}E_n)$ *factors*. Function ψ will be similar to one or another stationary state ψ_n, more often to those ψ_n which overlap significantly with the starting function $[\psi(0)]$ and/or correspond to low energy (low frequency). If the overlap $\langle\psi_n|\psi(0)\rangle$ of the starting function $\psi(0)$ with a stationary state ψ_n is zero, then during the evolution no admixture of the ψ_n state will be seen in $\psi(t)$, i.e., *only those stationary states that constitute the starting wave function* $\psi(0)$ *contribute to the evolution of* $\psi(t)$.

2.3.2 Time dependence of mechanical quantities

Let us take a mechanical quantity A and the corresponding (Hermitian) operator $\hat{A}$. Let us check whether the computed mean value (the normalization of the wave function ψ is assumed)

---

[29] One may verify inserting ψ' into the Schrödinger equation. Differentiating ψ' with respect to t, the left-hand side is obtained.

[30] The use of the spectral resolution of the identity in this form is not fully justified. A sudden cut of the electric field may leave the molecule with a nonzero translational energy. However, in the above spectral resolution one has the stationary states computed in the center-of-mass coordinate system, and therefore translation is not taken into account.

[31] We used here the property of an analytical function f that for any eigenfunction ψ_n of the operator $\hat{H}$ one has $f(\hat{H})\psi_n = f(E_n)\psi_n$. This follows from the Taylor expansion of $f(\hat{H})$ acting on eigenfunction ψ_n.

$\bar{A} \equiv \left\langle \hat{A} \right\rangle \equiv \left\langle \psi\left(t\right) | \hat{A}\left(t\right) \psi\left(t\right) \right\rangle$ depends on time. The time derivative of $\left\langle \hat{A} \right\rangle$ reads as (we use the time-dependent Schrödinger equation $\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}$)

$$\begin{gathered}
\frac{\mathrm{d}\left\langle \hat{A}\right\rangle}{\mathrm{d}t} = \left\langle \frac{\partial}{\partial t}\psi | \hat{A}\psi \right\rangle + \left\langle \psi | \left[\frac{\partial}{\partial t}\hat{A}\right]\psi \right\rangle + \left\langle \psi | \hat{A}\frac{\partial}{\partial t}\psi \right\rangle = \\
\left\langle -\frac{i}{\hbar}\hat{H}\psi | \hat{A}\psi \right\rangle + \left\langle \psi | \left[\frac{\partial}{\partial t}\hat{A}\right]\psi \right\rangle + \left\langle \psi | \hat{A}(-\frac{i}{\hbar})\hat{H}\psi \right\rangle = \\
\frac{i}{\hbar}\left\langle \hat{H}\psi | \hat{A}\psi \right\rangle + \left\langle \psi | \left[\frac{\partial}{\partial t}\hat{A}\right]\psi \right\rangle - \frac{i}{\hbar}\left\langle \psi | \hat{A}\hat{H}\psi \right\rangle = \\
\frac{i}{\hbar}\left\langle \psi | \hat{H}\hat{A}\psi \right\rangle + \left\langle \psi | \left[\frac{\partial}{\partial t}\hat{A}\right]\psi \right\rangle - \frac{i}{\hbar}\left\langle \psi | \hat{A}\hat{H}\psi \right\rangle = \\
\frac{i}{\hbar}\left\langle \psi | \left[\hat{H},\hat{A}\right]\psi \right\rangle + \left\langle \psi | \left[\frac{\partial}{\partial t}\hat{A}\right]\psi \right\rangle = \frac{i}{\hbar}\left\langle \left[\hat{H},\hat{A}\right]\right\rangle + \left\langle \frac{\partial}{\partial t}\hat{A}\right\rangle.
\end{gathered}$$

It is seen that

the mean value of a mechanical quantity in general depends on time through two components: the first contains the mean value of the commutator $\left[\hat{H},\hat{A}\right] = \hat{H}\hat{A} - \hat{A}\hat{H}$, the second one represents the mean value of the time derivative of the operator (the Ehrenfest theorem):

$$\frac{\mathrm{d}\left\langle \hat{A}\right\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle \left[\hat{H},\hat{A}\right]\right\rangle + \left\langle \frac{\partial}{\partial t}\hat{A}\right\rangle. \tag{2.17}$$

Thus, even if $\hat{A}$ does not depend on time explicitly ($\frac{\partial}{\partial t}\hat{A} = 0$), but does not commute with $\hat{H}$, the expected value of A, i.e., $\left\langle \hat{A}\right\rangle$, is time-dependent.[32]

---

[32] Eq. (2.17) looks a bit suspicious. The quantity $\left\langle \frac{\partial}{\partial t}\hat{A}\right\rangle$ is certainly a real function as the mean value of a Hermitian operator, but what about $\frac{i}{\hbar}\left\langle \left[\hat{H},\hat{A}\right]\right\rangle$ with this imaginary unit i? Well, everything is all right, because the operator $\left[\hat{H},\hat{A}\right]$ is anti-Hermitian, i.e., $\left\langle \psi | \left[\hat{H},\hat{A}\right]\psi \right\rangle = -\left\langle \left[\hat{H},\hat{A}\right]\psi | \psi \right\rangle$. This, however, means that $\left\langle \psi | \left[\hat{H},\hat{A}\right]\psi \right\rangle = -\left\langle \psi | \left[\hat{H},\hat{A}\right]\psi \right\rangle^*$ and, therefore, for the complex number $z = \left\langle \psi | \left[\hat{H},\hat{A}\right]\psi \right\rangle = \left\langle \left[\hat{H},\hat{A}\right]\right\rangle$ we have $z + z^* = 0$. Therefore, $\left\langle \left[\hat{H},\hat{A}\right]\right\rangle$ is necessarily an imaginary number of the type ib, with real b, and $\frac{i}{\hbar}\left\langle \left[\hat{H},\hat{A}\right]\right\rangle = \frac{i}{\hbar}ib = -\frac{b}{\hbar}$ is for sure a real number.

2.3.3 Mean energy is conserved

For any isolated system $\hat{H} \neq f(t)$, and when we take $\hat{A} = \hat{H}$, both terms are equal to zero and we get $\frac{\mathrm{d}\langle \hat{H} \rangle}{\mathrm{d}t} = 0$.

The mean value of the Hamiltonian is conserved during evolution.

2.3.4 Symmetry is conserved

The time-dependent Schrödinger equation says the following: we have the wave function at time $t = 0$ (i.e., $\psi(x, 0)$). If you want to see how it will look like at time t, you just have to apply to function $\psi(x, 0)$ an evolution operator $\exp(-i\frac{\hat{H}}{\hbar}t) \equiv \hat{U}(t)$ and you get the answer: $\hat{U}(t)\psi(x, 0) = \psi(x, t)$. That is it!

There remains, however, a small problem: how will the function $\psi(x, t)$ be related to $\psi(x, 0)$? What kind of question is this? This will be in general just another function (well, it will, however, preserve the normalization condition and the mean value of the energy). It is as if somebody asked about evolution of an Arabian horse. It will preserve its weight (analog of the normalization), and it will move with the same kinetic energy (analog of conserving total energy). Let us, however, consider some more subtle features. For example, is it possible that at $t = 0$, the horse is in a symmetric state,[33] while after a while (action of $\hat{U}(t)$) we get an *asymmetric* state of the horse, like such one which has a tendency to bend its head to the right?

Let us consider a symmetry operator $\hat{R}$ that commutes with $\hat{H} : \hat{R}\hat{H} = \hat{H}\hat{R}$. Let us assume that the initial state $\psi(x, 0)$ *exhibits* a symmetry, i.e., satisfies the following equation:

$$\hat{R}\psi(x, 0) = \exp(i\alpha)\psi(x, 0), \tag{2.18}$$

where α represents a certain real number.[34] The symmetry is guaranteed in such a case, because the modulus of the transformed function (which decides about probability) does not change:

---

[33] This means that the horse in its movements does not prefer its right (left) side with respect to its left (right) side. This does not mean that taking a picture of such a horse results in a perfect symmetry on the picture. Such a state means only that after taking very many such pictures and after superposing all of them to get one picture, we see a perfectly symmetric creature.

[34] If this symmetry operation means

- an arbitrary time-independent *translation* of the coordinate system, say along the x axis, we have $\alpha \sim p_x$ (p_x means the x component of the momentum, see Appendix F on p. 665) and p_x represents a constant of motion, i.e., does not change;

$|\hat{R}\psi(x,0)| = |\exp(i\alpha)\psi(x,0)| = |\exp(i\alpha)| \cdot |\psi(x,0)| = 1 \cdot |\psi(x,0)| = |\psi(x,0)|$. This symmetry is characterized by the value of α. What can we say about the symmetry of the final state? Let us see.[35] We have

$$\hat{R}\psi(x,t) = \hat{R}\hat{U}(t)\psi(x,0) = \hat{R}\exp(-i\frac{\hat{H}}{\hbar}t)\psi(x,0) = \exp(-i\frac{\hat{H}}{\hbar}t)\hat{R}\psi(x,0) =$$
$$\exp(-i\frac{\hat{H}}{\hbar}t)\exp(i\alpha)\psi(x,0) = \exp(i\alpha)\exp(-i\frac{\hat{H}}{\hbar}t)\psi(x,0) = \exp(i\alpha)\psi(x,t).$$

Thus, the final state exhibits at any time t the same symmetry (due to the same value of α), i.e.,

$$\hat{R}\psi(x,t) = \exp(i\alpha)\psi(x,t). \tag{2.19}$$

> Evolution in time can make the symmetry of the wave function neither appear nor disappear. The symmetry is conserved.

2.3.5 *Energy-time uncertainty principle*

Besides the Heisenberg uncertainty principle (p. 41) there is another relation of this kind connected with the time evolution of the wave function described by the time-dependent Schrödinger equation. This new inequality is known as the energy-time uncertainty. It will be derived below following Mandelshtam and Tamm.[36]

The first ingredient will be the general Eq. (1.25) derived on p. 43,

$$\Delta A \cdot \Delta B \geq \frac{1}{2}|\langle\Psi|[\hat{A},\hat{B}]\Psi\rangle|,$$

which pertains to two standard deviations ΔA and ΔB (errors of measurements of quantities A and B), whereas $\hat{A}$ and $\hat{B}$ stand for the quantum mechanical operators corresponding to A and B.

- an arbitrary *translation on the time* axis, we have $\alpha \sim E$ (E stands for the total energy) and E is also a constant of motion;
- for an arbitrary *rotation* of the coordinate system, say, about the z axis, we have $\alpha \sim J_z$ (J_z denotes the z component of the angular momentum, see Appendix F on p. 665) with J_z as a constant of motion.

[35] Note that from $\hat{R}\hat{H} = \hat{H}\hat{R}$ also follows that $\hat{R}\exp(-i\frac{\hat{H}}{\hbar}t) = \exp(-i\frac{\hat{H}}{\hbar}t)\hat{R}$ (please recall the definition of $\exp(-i\frac{\hat{H}}{\hbar}t)$ through the Taylor expansion).

[36] L.I. Mandelshtam, I.E. Tamm, *J. Phys. (USSR)* 9(1945)249.

Now, we choose as A the total energy E (i.e., $\hat{A}$ is identical to the Hamiltonian, $\hat{A} \equiv \hat{H}$), and we assume that $\hat{H}$ does not depend on time. However, even if $\hat{H}$ does not depend on time, ΔE may not be zero. This is because of Eq. (2.15), p. 97, where we have evolution of ψ in time:

$$\psi(t) = \sum_n c_n \exp(-i\frac{t}{\hbar}E_n)|\psi_n\rangle$$

with $c_n \equiv \langle\psi_n|\psi(0)\rangle$, and we see that the mean value of the energy is[37]

$$\bar{E} = \left\langle \hat{H} \right\rangle = \left\langle \psi(t)|\hat{H}\psi(t) \right\rangle = \sum_n \sum_m c_n c_m^* \exp[-i\frac{t}{\hbar}(E_n - E_m)]E_n \langle\psi_m|\psi_n\rangle = \sum_n |c_n|^2 E_n.$$

Therefore, after noting that the coefficients c_n do not depend on time, we see that *in quantum mechanics, for an arbitrary quantum state* ψ, *only the mean value of energy (not energy* E *itself) remains constant,*[38] *so* $\bar{E} = \text{const}$. In other words, what we measure as energy will be in general one of the E_n values (according to Postulate IV, p. 29), with the probability of such a measurement equal to $|c_n|^2$ (note that these probabilities are time-independent).

Now, what about the time dependence of the mean values $\left\langle \hat{B} \right\rangle$ of other quantities B, that do not depend on time explicitly, i.e., $\frac{\partial}{\partial t}\hat{B} = 0$? Their mean values $\left\langle \hat{B} \right\rangle$ in general change in time, since from Eq. (2.17) we have

$$\frac{\mathrm{d}\left\langle \hat{B} \right\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle\left[\hat{H}, \hat{B}\right]\right\rangle + \left\langle \frac{\partial}{\partial t}\hat{B} \right\rangle = \frac{i}{\hbar}\left\langle\left[\hat{H}, \hat{B}\right]\right\rangle + 0 = \frac{i}{\hbar}\left\langle\left[\hat{H}, \hat{B}\right]\right\rangle. \tag{2.20}$$

Therefore, we have two cases:

- $\left[\hat{H}, \hat{B}\right] = 0$, i.e., $\hat{H}$ and $\hat{B}$ commute; therefore $\frac{\mathrm{d}\left\langle \hat{B} \right\rangle}{\mathrm{d}t} = 0$ and $\left\langle \hat{B} \right\rangle = \text{const}$, i.e., the corresponding mean values stay constant;
- $\left[\hat{H}, \hat{B}\right] \neq 0$, i.e., $\hat{H}$ and $\hat{B}$ do not commute; therefore $\frac{\mathrm{d}\left\langle \hat{B} \right\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle\left[\hat{H}, \hat{B}\right]\right\rangle \neq 0$, therefore $\left\langle \hat{B} \right\rangle$ vary in time.

In the first case $\Delta E \cdot \Delta B \geq 0$, which means the measurement errors of energy E and B can be made arbitrarily small.

---

[37] We have chosen $\{\psi_n\}$ as the orthonormalized complete set and ψ is ensured to be normalized.

[38] The same is seen also from Eq. (2.17). Indeed, we have $\frac{\mathrm{d}\left\langle \hat{H} \right\rangle}{\mathrm{d}t} = \frac{\mathrm{d}\bar{E}}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle\left[\hat{H}, \hat{H}\right]\right\rangle + \left\langle \frac{\partial}{\partial t}\hat{H} \right\rangle = 0 + 0 = 0$, which means again $\bar{E} = \text{const}$.

In the second case, using Eq. (1.25) for $\hat{A} = \hat{H}$ we have

$$\Delta E \cdot \Delta B \geq \frac{1}{2}\left|\left\langle[\hat{H}, \hat{B}]\right\rangle\right| = \frac{\hbar}{2}\frac{\mathrm{d}\left\langle\hat{B}\right\rangle}{\mathrm{d}t}.$$

Dividing (for the second case) both sides by $\frac{\mathrm{d}\left\langle\hat{B}\right\rangle}{\mathrm{d}t}$ one gets

$$\Delta E \cdot \frac{\Delta B}{\frac{\mathrm{d}\left\langle\hat{B}\right\rangle}{\mathrm{d}t}} \geq \frac{\hbar}{2}.$$

The quantity $\frac{\Delta B}{\frac{\mathrm{d}\left\langle\hat{B}\right\rangle}{\mathrm{d}t}}$ has the dimension of time. It is the time that elapsed till the change of $\left\langle\hat{B}\right\rangle$ reaches its single standard deviation (ΔB). We have

$$\Delta E \cdot \Delta t_B \geq \frac{\hbar}{2}. \tag{2.21}$$

Note that here Δt_B does not represent any error in determining time. It is rather a time associated with recognizing that $\left\langle\hat{B}\right\rangle$ changes, because its change goes beyond its error bar ΔB. The faster $\left\langle\hat{B}\right\rangle$ is changing, the shorter Δt_B needs to reach the change that amounts to ΔB. Among all B, we have to choose the shortest Δt_B (let us denote it by Δt). Even for such choice of time we should have

$$\Delta E \cdot \Delta t \geq \frac{\hbar}{2}, \tag{2.22}$$

which is known as the energy-time uncertainty principle. Its meaning, however, is fundamentally different from the canonical position-momentum uncertainty of Heisenberg (Δt is not any standard deviation of time).

Practical applications of the energy-time uncertainty are related to several distinct situations. Their general feature is connected to time-dependent states ψ that evolve among smaller or larger numbers of stationary states of different discrete energies. If their number is large/small, also ΔE is supposed to be large/small and, from the energy-time uncertainty principle, it follows that Δt becomes small/large. A short Δt means a substantial departure from the initial state going on. In this spirit one links a long or short lifetime Δt of a system to its sharp or diffuse energy distribution (ΔE). The expression (2.22), treated as equality, may be used as the relation between the energy uncertainty ΔE and lifetime Δt of the system (Fig. 2.9).

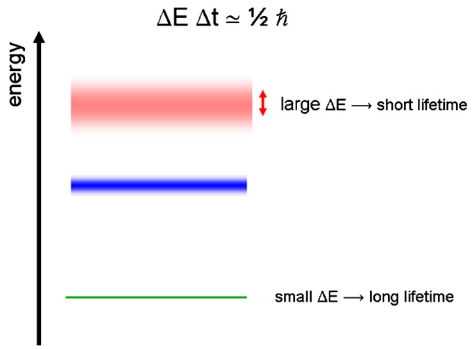


Fig. 2.9. The energy-time uncertainty principle.

2.3.6 *Meditations at water spring*

In chemistry one assumes (tacitly) that two molecules, say, a water molecule created in a chemical reaction a millisecond ago and a water molecule from the Oligocene (i.e., created, say, more than 23 millions years ago), represent identical objects. How could we know this?

In liquid water the molecules are subject to intermolecular interactions, which complicates things. Let us consider the same molecules, but isolated in outer space. It is “generally believed” that even a molecule created a millisecond ago (not speaking about one from the Oligocene) had enough time to achieve the ground state via emission of photons. If this is true, we can consider them as described by identical ground-state wave functions.

What about 1 femtosecond (10^{-15} s) instead of one millisecond? Well, very probably the first molecule would be in a nonstationary state[39] and it would have no time for emitting photons. These two molecules would be different (distinguishable).

Let us consider larger molecules, e.g., two molecules of hemoglobin in the interstellar space, created by two different methods a femtosecond ago.[40] With the probability very close to 1, these two molecules would be created in two different conformational states. Now, both states evolve in time. Even if they would lower their energies and reach the same ground state by emitting photons, this would take virtually an infinite amount of time due to the plethora of kinetic traps (metastable conformations) on their trajectories within the configurational space. These metastable conformations are separated by quite important energy barriers, difficult to overcome. The hemoglobin molecules, unlike the water molecule, will in general have a long

---

[39] Single vibration in molecule is a matter of femtoseconds.

[40] This example looks surrealistic, but science relies on questions of the type “what if?”.

memory of their initial states. Thus, we see that all systems evolve, but the evolution time spans an incredibly large time scale.[41]

Well, let us come back to water aggregates (treated as large molecules) formed by a net of hydrogen bonds, like those in liquid water. As shown by Margarita Rodnikova the hydrogen bond network in liquid water exhibits a kind of elastic properties,[42] i.e., behaves like molecular aggregates with some stability. What about the lifetime of such aggregates[43]? A contemporary approach to this problem is just ignoring it or saying arbitrarily that liquid water has no memory. It would be certainly more appropriate to leave the answer to experiment.[44]

2.3.7 Linearity

The most mysterious feature of the Schrödinger equation is its linear character. The world is nonlinear, because the effect is never strictly proportional to its cause. However, if $\psi_1(x,t)$ and $\psi_2(x,t)$ satisfy the time-dependent Schrödinger equation, then their arbitrary linear combination also represents a solution.[45]

2.4 Evolution after switching a perturbation

Let us suppose that we have a system with the Hamiltonian $\hat{H}^{(0)}$ and its stationary states $\psi_k^{(0)}$, which form the orthonormal complete set[46]

$$\hat{H}^{(0)}\psi_k^{(0)} = E_k^{(0)}\psi_k^{(0)}, \tag{2.23}$$

$$\psi_k^{(0)}(x,t) = \phi_k^{(0)}(x)\exp\left(-i\frac{E_k^{(0)}}{\hbar}t\right), \tag{2.24}$$

where x represents the coordinates, and t denotes time.

---

[41] Think of a shell visible in a rock, or the rock structure itself. These structures have been created many millions years ago, but evolve so slowly that we see them today.

[42] M.N. Rodnikova, *J. Phys. Chem. (Russ.)*, **67**(1993)275.

[43] One has to define somehow the lifetime ("memory"). It could be, e.g., the relaxation time τ, after which the root mean square deviation from the starting structure (in atomic resolution) exceeds, say, 1 Å. If the structure is stable, τ is large and the memory of the molecular aggregate is also large.

[44] Some water aggregates bound by the hydrogen bonds have for sure very large lifetimes. How can I know this? Simply I have seen my footprints on the snow. They certainly represented nothing else but structures existing only because of the hydrogen bonds. These structures did not disappear in a femtosecond, but were there for many hours. Interestingly, when the temperature was raised by a few degrees (above 0°C) my footprints disappeared in an hour... Did they disappear instantaneously upon melting?

[45] Indeed, $\hat{H}(c_1\psi_1 + c_2\psi_2) = c_1\hat{H}\psi_1 + c_2\hat{H}\psi_2 = c_1 i\hbar\frac{\partial\psi_1}{\partial t} + c_2 i\hbar\frac{\partial\psi_2}{\partial t} = i\hbar\frac{\partial(c_1\psi_1 + c_2\psi_2)}{\partial t}$.

[46] This *can* always be ensured (by suitable orthogonalization and normalization) and follows from the Hermitian character of the operator $\hat{H}^{(0)}$.

Let us assume that at time $t = 0$ the system is in the stationary state $\psi_m^{(0)}$.

> At $t = 0$ a drama begins: one switches on the perturbation $V(x, t)$, which in general depends on all the coordinates (x) and time (t), and after time τ the perturbation is switched off. Now we ask a question about the probability of finding the system in the stationary state $\psi_k^{(0)}$.

After the perturbation is switched on, wave function $\psi_m^{(0)}$ is no longer stationary and begins to evolve in time according to the time-dependent Schrödinger equation $\left(\hat{H}^{(0)} + \hat{V}\right)\psi = i\hbar\frac{\partial\psi}{\partial t}$. This is a differential equation with partial derivatives with the boundary condition $\psi(x, t = 0) = \phi_m^{(0)}(x)$. The functions $\left\{\psi_n^{(0)}\right\}$ form a complete set and therefore the wave function that fulfills the Schrödinger equation $\psi\,(x, t)$ at any time can be represented as a linear combination with time-dependent coefficients c, i.e.,

$$\psi\,(x, t) = \sum_{n=0}^{\infty} c_n\,(t)\,\psi_n^{(0)}\,(x, t)\,. \tag{2.25}$$

Inserting this into the left-hand side of the time-dependent Schrödinger equation one obtains

$$\left(\hat{H}^{(0)} + \hat{V}\right)\psi = \sum_n c_n\left(\hat{H}^{(0)} + \hat{V}\right)\psi_n^{(0)} = \sum_n c_n\left(E_n^{(0)} + V\right)\psi_n^{(0)},$$

whereas its right-hand side gives

$$i\hbar\frac{\partial\psi}{\partial t} = i\hbar\sum_n\left[\psi_n^{(0)}\frac{\partial c_n}{\partial t} + c_n\frac{\partial\psi_n^{(0)}}{\partial t}\right] = i\hbar\sum_n\left[\psi_n^{(0)}\frac{\partial c_n}{\partial t} + c_n\left(-\frac{i}{\hbar}E_n^{(0)}\right)\psi_n^{(0)}\right] =$$
$$\sum_n\left[i\hbar\psi_n^{(0)}\frac{\partial c_n}{\partial t} + c_n E_n^{(0)}\psi_n^{(0)}\right].$$

Both sides give

$$\sum_n c_n\hat{V}\psi_n^{(0)} = \sum_n\left(i\hbar\frac{\partial c_n}{\partial t}\right)\psi_n^{(0)}.$$

Multiplying the left-hand side by $\psi_k^{(0)*}$ and integrating results in

$$\sum_{n}^{\infty} c_n V_{kn} = i\hbar \frac{\partial c_k}{\partial t}, \tag{2.26}$$

for $k = 1, 2, ...$, where

$$V_{kn} = \left\langle \psi_k^{(0)} | \hat{V} \psi_n^{(0)} \right\rangle. \tag{2.27}$$

The formulae obtained are equivalent to the Schrödinger equation. These are differential equations, which we would generally like, provided the summation is not infinite.[47] In practice, however, one has to keep the summation finite.[48] If the assumed number of terms in the summation is not too large, then in the computer era solution of the problem is feasible.

2.4.1 *Time-independent perturbation – the two-state model*

For the sake of simplicity let us take the *two-state model* (cf. Appendix D, p. 655) with two orthonormal eigenfunctions $\left|\phi_1^{(0)}\right\rangle = |1\rangle$ and $\left|\phi_2^{(0)}\right\rangle = |2\rangle$ of the Hamiltonian $\hat{H}^{(0)}$

$$\hat{H}^{(0)} = E_1^{(0)} |1\rangle \langle 1| + E_2^{(0)} |2\rangle \langle 2|$$

with the perturbation (to ensure $\hat{V}$ is Hermitian)

$$\hat{V} = v |1\rangle \langle 2| + v^* |2\rangle \langle 1|, \tag{2.28}$$

with the corresponding matrix $\mathbf{V} = \begin{pmatrix} 0 & v \\ v^* & 0 \end{pmatrix}$.

This model has an exact solution (even for a large perturbation $\hat{V}$). One may introduce various time dependencies of $\hat{V}$, including various regimes for switching on the perturbation.

The differential equations (2.26) for the coefficients $c_1(t)$ and $c_2(t)$ are (in a.u., $\omega_{21} = E_2^{(0)} - E_1^{(0)}$)

$$c_2 v \exp(-i\omega_{21} t) = i \frac{\partial c_1}{\partial t},$$
$$c_1 v \exp(i\omega_{21} t) = i \frac{\partial c_2}{\partial t}.$$

---

[47] In fact only then the equivalence to the Schrödinger equation is ensured.

[48] This is typical for expansions into the complete set of functions (the so-called *algebraic approximation*).

Now we assume that v is time-independent and the initial wave function is $|1\rangle$, i.e., $c_1(0)=1$, $c_2(0)=0$. In such a case one obtains[49]

$$c_1(t)=\exp(-i\frac{1}{2}\omega_{21}t)\left[\cos(avt)+i\frac{\omega_{21}}{2av}\sin(avt)\right], \tag{2.29}$$

$$c_2(t)=-\frac{i}{a}\exp(i\frac{1}{2}\omega_{21}t)\sin(avt), \tag{2.30}$$

where $a=\sqrt{1+\left(\frac{\omega_{21}}{2v}\right)^2}$. Fig. 2.10 shows how the weights $|c_1(t)|^2$ and $|c_2(t)|^2$ change in time for various time-independent perturbation strengths. One may conclude, that a time-independent perturbation is unable to switch completely the two states that differ in energy. The switching is possible for the case of degeneracy only.

Two states – degeneracy

One of the most important cases corresponds to the degeneracy $\omega_{21}=E_2^{(0)}-E_1^{(0)}=0$, (Fig. 2.10a). One obtains $a=1$ and

$$c_1(t)=\cos(vt),$$
$$c_2(t)=-i\sin(vt).$$

A very interesting result. After applying symmetric orthogonalization (Appendix K), the functions $|1\rangle$ and $|2\rangle$ may be identified with ψ_D and ψ_L for the D and L enantiomers (cf. p. 84) or, with the wave functions 1s, centered on the two nuclei in the H_2^+ molecule. As one can see from the last two equations, the two wave functions oscillate transforming one to the other with an oscillation period $T=\frac{2\pi}{v}$. If v were very small (as in the case of D- and L-glucose), then the oscillation period would be very large. This happens to D- and L-enantiomers of glucose, where changing the nuclear configuration from one to the other enantiomer means breaking a chemical bond (a high and wide energy barrier to overcome). This is why the drugstore keeper can safely stock a single enantiomer for a very long time.[50] This may not be true for other enantiomers. For example, imagine a pair of enantiomers that represent some intermolecular complexes where a small change in the nuclear framework may cause one of them to transform into the other. In such a case, the oscillation period may be much smaller than the lifetime of the Universe, e.g., it may be comparable to the time of an experiment. In such a case one could observe the oscillation between the two enantiomers.

---

[49] Use, e.g., Mathematica software. Let us check the conservation of normalization (i.e., its time independence): $|c_1(t)|^2+|c_2(t)|^2=\cos^2(avt)+\frac{\omega^2}{(2av)^2}\sin^2(avt)+\frac{1}{a^2}\sin^2(avt)=\cos^2(avt)+\frac{\omega^2+4v^2}{(2av)^2}\sin^2(avt)=\cos^2(avt)+\sin^2(avt)=1$.

[50] No longer, however, than the expiration date.

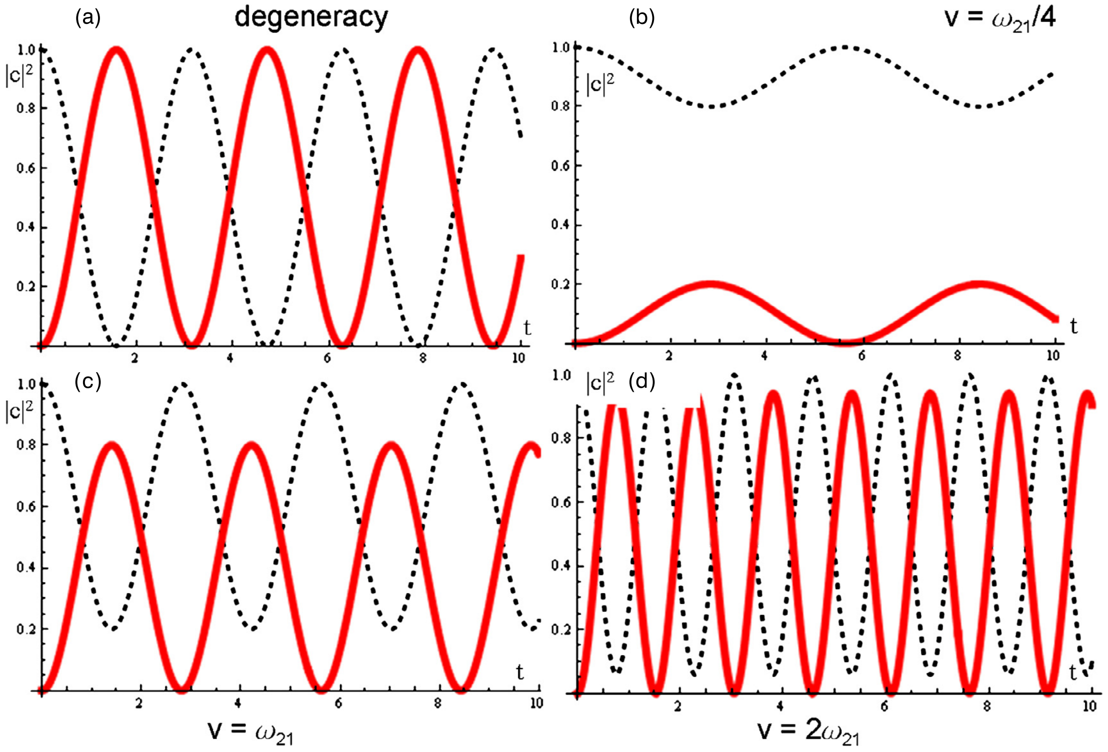


Fig. 2.10. The two-state model with the time-independent perturbation v. All panels show the time dependence of $|c_1(t)|^2$ (dotted line) and $|c_2(t)|^2$ (thick line) according to Eqs. (2.29) and (2.30) obtained for the initial conditions $c_1(0) = 1$ and $c_2(0) = 0$. Panel (a) corresponds to the double degeneracy of the energy level. Panel (b) is for the perturbation v much smaller than the energy gap between the two energy levels ($\hbar\omega_{21} = \omega_{21}$ in a.u.). Panel (c) corresponds to perturbation v being equal to the energy gap. No transition to state 2 takes place during the time evolution. Panel (d) shows the same as panel (c), but for a larger perturbation ($v = 2\omega_{21}$). As one can see, the gap does not count much for a large perturbation v (one has to do with quasidegeneracy) and, therefore, we have oscillations that practically (not exactly) resemble those shown in panel (a).

2.4.2 Oscillating perturbation – the two-state model

It is interesting what happens to the wave function as time goes on, when a perturbation varying with time as $\exp(i\omega t)$ is switched on. Eq. (2.28) has to be modified to (the second term has an $\exp(-i\omega t)$ time dependence to ensure the operator (2.31) is Hermitian)

$$\hat{V} = v[\exp(i\omega t)\,|1\rangle\,\langle 2| + \exp(-i\omega t)\,|2\rangle\,\langle 1|], \tag{2.31}$$

with $|1\rangle \equiv \psi_1^{(0)}$ and $|2\rangle \equiv \psi_2^{(0)}$.

Inserting such a perturbation into Eqs. (2.26) results once again in Eqs. (2.29) and (2.30), but this time with replacing $\omega_{21} \to \omega_{21} - \omega$. Assuming function $|1\rangle$ as the starting one means $c_1(0) = 1$, $c_2(0) = 0$ and therefore solution (2.32) takes the form

$$c_1(t) = \exp\left[-i\frac{1}{2}(\omega_{21} - \omega)\,t\right]\left[\cos(avt) + i\frac{(\omega_{21} - \omega)}{2av}\sin(avt)\right],$$
$$c_2(t) = -\frac{i}{a}\exp\left[i\frac{1}{2}(\omega_{21} - \omega)\,t\right]\sin(avt),$$

where this time $a = \sqrt{1 + \left(\frac{\omega_{21} - \omega}{2v}\right)^2}$. A numerical example is demonstrated in Fig. 2.11.

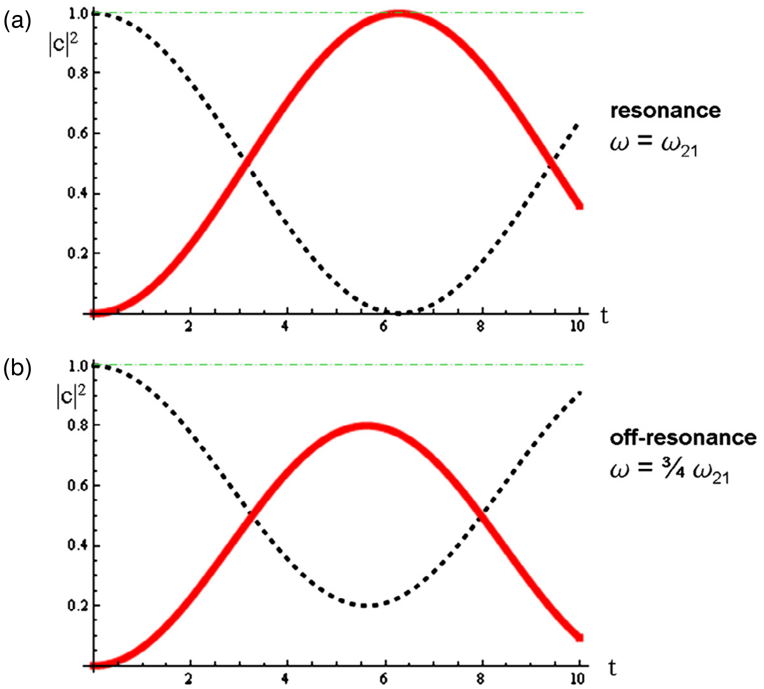


Fig. 2.11. The two-state model with the (time-dependent) oscillating perturbation of frequency ω, other parameters are as in Fig. 2.10b for a time-independent perturbation. At the start time $t = 0$ the coefficients $c_1(0) = 1$ and $c_2(0) = 0$ are assumed. The panels (a) and (b) show the weights $|c_1(t)|^2$ (dashed line) and $|c_2(t)|^2$ (solid line) as functions of time. Panel (a) pertains to the resonance ($\omega = \omega_{21}$) case, while panel (b) shows what happens for an off-resonance situation (in this case $\omega = 0.75\,\omega_{21}$). Note that only in the resonance case one obtains a full exchange of the two states during oscillations.

Two states – resonance case

For $\omega = \omega_{21}$ (the energy of the photon matches the energy level difference) we obtain

$$\begin{aligned} c_1(t) &= \cos(vt), \\ c_2(t) &= -i\sin(vt), \end{aligned} \tag{2.32}$$

i.e., the system oscillates between state $\psi_1^{(0)}$ and state $\psi_2^{(0)}$, with period $\frac{2\pi}{v}$ and none of these states is privileged[51] (Fig. 2.11a).

> It is intriguing to see that for oscillating perturbation in the case of two levels of different energies we got exactly the same behavior as in the case of ...degenerate levels! It looks (Fig. 2.12) as if the two levels became indeed of equal energy (degeneracy) after the resonance photon energy is counted to one of them depending on absorption or emission.

2.4.3 Short-time perturbation – the first-order approach

If one is to apply first-order perturbation theory, two things have to be ensured: the perturbation $\hat{V}$ has to be small and the time of interest has to be small (switching the perturbation on corresponds to $t = 0$). This is what we are going to assume from now on. At $t = 0$ one starts from the m-th state and therefore $c_m = 1$, while other coefficients $c_n = 0$. Let us assume that to the first approximation the domination of the m-th state continues even after switching the perturbation on, and we will be interested in detecting the most important tendencies in time evolution of c_n for $n \neq m$. These assumptions (they give first-order perturbation theory[52]) lead to a considerable simplification of Eqs. (2.26):

$$V_{km} = i\hbar\frac{\partial c_k}{\partial t} \quad \text{for } k = 1, 2, \ldots N.$$

---

[51] Such oscillations necessarily mean that the energy of the system changes periodically: after time $\tau = \frac{\pi}{2v}$ the system absorbs a photon of energy $\hbar\omega_{12}$ from the electromagnetic field, and then, again after τ, emits the same photon, then again absorbs, emits, etc.; this scenario repeats periodically (excitations and deexcitations). Such behavior is possible only because of a continuous supply of photons from the field given by Eq. (2.31). If the system interacted with a single photon instead of the field (2.31), the excited state would change to the ground state and the photon in the form of a spherical wave expanding to infinity (equal probability of detecting the photon in any spot on a sphere). If we had only our system in the Universe and the Universe were limited by a mirror, the photon would finally come back to the system, causing its excitation, then again emission, etc., similarly as in our solution. However, our Universe is not limited by a mirror and the photon spherical wave would expand to infinity without finding any obstacle. There would be no chance for the photon to come back and the atom would not be excited.

[52] For the sake of simplicity we will not introduce a new notation for the coefficients c corresponding to the first-order procedure. If the above simplified equation were introduced to the left-hand side of Eq. (2.26), then its solution would give c accurate up to the second order, etc.

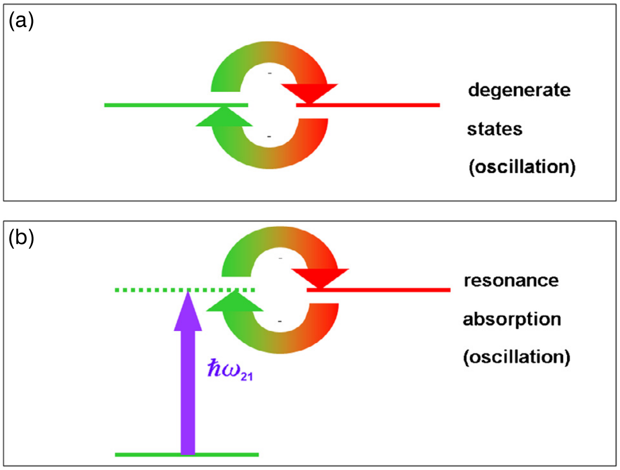


Fig. 2.12. The two-state model in the time-dependent Schrödinger equation. In panel (a) it is shown what happens when the two states correspond to the same energy (degeneracy) and the perturbation is time-independent. It turns out (see the text) that the wave function oscillates in time between two component states with the oscillation period $\frac{2\pi}{v}$, where v represents the two-states coupling constant. Panel (b) shows what happens for the wave function when the two states differ in energy (by $\hbar\omega_{21}$), and the coupling between the two states oscillates with the frequency ω (it is of the form $v\exp(i\omega t)$). For $\omega = \omega_{21}$ (resonance) one has a very special behavior: the solution of the time-dependent Schrödinger equation looks exactly as in the case of degeneracy (related to panel (a)). It looks as if a degeneracy of the two states is achieved, after the lower state absorbs a photon of the resonance energy $\hbar\omega_{21}$.

In this, and the further equations of this chapter, the coefficients c_k will depend implicitly on the initial state m. The change of $c_k(t)$ is therefore proportional to V_{km}. A strong coupling for the expansion function $\psi_k^{(0)}$ (i.e., the system becomes a bit similar to that described by $\psi_k^{(0)}$) corresponds to large values of the coupling coefficient V_{km}, which happens when function $\psi_k^{(0)}$ resembles function $\hat{V}\psi_m^{(0)}$. This represents a strong constraint both for $\psi_k^{(0)}$ and $\hat{V}$; only some special perturbations $\hat{V}$ are able to couple effectively two states,[53] such as $\psi_k^{(0)}$ and $\psi_m^{(0)}$.

---

[53] First of all, there must be something in $\hat{V}$ which influences the particles of the system, like, e.g., electric field interaction with electrons of an atom. If the atom stays spherically symmetric, there is no coupling with the field. Only making a shift of electrons we get some interaction, proportional to this shift. Therefore, roughly speaking, $\hat{V}$ is in this case proportional to the shift x of the most weakly bound electron. Let us assume that we start from $\psi_m^{(0)}$ (describing this electron, we neglect the other electrons), which is a spherically symmetric function. Therefore, $\hat{V}\psi_m^{(0)} \sim x \cdot$ (spherically symmetric) and the most promising function $\psi_k^{(0)}$ would be of the p_x type, since only then the integral V_{km} would have a chance to be nonzero. But still, if $\psi_m^{(0)}$ and $\psi_k^{(0)}$ correspond to different energies, they have their phase factors multiplied in the integrand, which results in their product oscillating in time with the frequency $\omega_{km} = (E_k^{(0)} - E_m^{(0)})/\hbar$. The only way to damp these oscillations

The quantity V_{km} depends on time for two or even three reasons: firstly and secondly, the stationary states $\psi_m^{(0)}$ and $\psi_k^{(0)}$ do have a time dependence, and thirdly, in addition the perturbation $\hat{V}$ may also depend on time. Let us highlight the time dependence of the wave functions by introducing the frequency

$$\omega_{km} = \frac{E_k^{(0)} - E_m^{(0)}}{\hbar}$$

and the definition

$$v_{km} \equiv \left\langle \phi_k^{(0)} | \hat{V} \phi_m^{(0)} \right\rangle.$$

One obtains

$$-\frac{i}{\hbar} v_{km}\, e^{i\omega_{km}t} = \frac{\partial c_k}{\partial t}.$$

Subsequent integration with the boundary condition $c_k(\tau = 0) = 0$ for $k \neq m$ gives

$$c_k(\tau) = -\frac{i}{\hbar} \int_0^\tau dt\; v_{km}(t)\, e^{i\omega_{km}t}. \tag{2.33}$$

The square of $c_k(\tau)$ represents (to the accuracy of first-order perturbation theory) the probability that at time τ the system will be found in state $\psi_k^{(0)}$. Let us calculate this probability for a few important cases of perturbation $\hat{V}$.

2.4.4 Time-independent perturbation and the Fermi Golden Rule

From Eq. (2.33) one has

$$c_k(\tau) = -\frac{i}{\hbar} v_{km} \int_0^\tau dt\; e^{i\omega_{km}t} = -\frac{i}{\hbar} v_{km} \frac{e^{i\omega_{km}\tau} - 1}{i\omega_{km}} = -v_{km} \frac{e^{i\omega_{km}\tau} - 1}{\hbar\omega_{km}}. \tag{2.34}$$

(making V_{km} having a substantial value for any t) is to use $\hat{V}$, and therefore x, oscillating itself, preferably with the same frequency to damp effectively. It is like to damp oscillations $\exp(i\omega_{km}t)$ we have to have $\hat{V}$ oscillating as $\exp(-i\omega_{km}t)$, because these factors when multiplied give no oscillations. The last conclusion is independent of the nature of $\hat{V}$. Therefore, even from such a simple reasoning, to change quantum states we have to use light frequencies which match the difference of the corresponding energy levels.

Now let us calculate the probability density $P_m^k = |c_k|^2$ that at time τ the system will be in state k (the initial state is m). We have

$$P_m^k(\tau) = |v_{km}|^2 \frac{(-1+\cos\omega_{km}\tau)^2 + \sin^2\omega_{km}\tau}{(\hbar\omega_{km})^2} = |v_{km}|^2 \frac{(2-2\cos\omega_{km}\tau)}{(\hbar\omega_{km})^2} =$$

$$|v_{km}|^2 \frac{\left(4\sin^2\frac{\omega_{km}\tau}{2}\right)}{(\hbar\omega_{km})^2} = |v_{km}|^2 \frac{1}{\hbar^2} \frac{\left(\sin^2\frac{\omega_{km}\tau}{2}\right)}{\left(\frac{\omega_{km}}{2}\right)^2}.$$

In order to undergo the transition from state m to state k one has to have a large v_{km}, i.e., a large coupling of the two states through perturbation $\hat{V}$. Note that probability P_m^k strongly depends on the time τ chosen; the probability oscillates as the square of the sine when τ increases, for some τ it is large, for others it is just zero. From Appendix E, p. 659 one can see that for large values of τ one may write the following approximation[54] to P_m^k:

$$P_m^k(\tau) \cong |v_{km}|^2 \pi \frac{\tau}{\hbar^2} \delta\left(\frac{\omega_{km}}{2}\right) = \frac{2\pi\tau}{\hbar^2} |v_{km}|^2 \delta(\omega_{km}) = \frac{2\pi\tau}{\hbar} |v_{km}|^2 \delta\left(E_k^{(0)} - E_m^{(0)}\right),$$

where we have used twice the Dirac delta function property that $\delta(ax) = \frac{\delta(x)}{|a|}$.

As one can see, P_m^k is proportional to time, which makes sense only because time τ has to be relatively small (first-order perturbation theory has to be valid). Note that the Dirac delta function forces the energies of both states (the initial and the final) to be equal, because of the time independence of $\hat{V}$.

A time-independent perturbation is unable to change the state of the system when it corresponds to a change of its energy.

A very similar formula is systematically derived in several important cases. Probably this is why the probability per unit time is called, a little poetically, the Fermi Golden Rule[55]:

FERMI GOLDEN RULE

$$w_m^k \equiv \frac{P_m^k(\tau)}{\tau} = |v_{km}|^2 \frac{2\pi}{\hbar} \delta\left(E_k^{(0)} - E_m^{(0)}\right). \tag{2.35}$$

---

[54] Large when compared to $2\pi/\omega_{km}$, but not too large in order to keep the first-order perturbation theory valid.

[55] E. Fermi, "*Nuclear Physics*," University of Chicago Press, Chicago, 1950, p. 142.

2.4.5 Oscillating perturbation and the Fermi Golden Rule

Let us assume a time-dependent periodic perturbation

$$\hat{V}(x,t) = \hat{v}(x)\, e^{\pm i\omega t}.$$

Such a perturbation corresponds, e.g., to an oscillating electric field[56] of angular frequency ω.

Let us take a look at successive equations, which we obtained at the time-independent $\hat{V}$. The only change will be that V_{km} will have the form

$$V_{km} \equiv \left\langle \psi_k^{(0)} | \hat{V} \psi_m^{(0)} \right\rangle = v_{km} e^{i(\omega_{km} \pm \omega)t} \quad \text{instead of} \quad V_{km} \equiv \left\langle \psi_k^{(0)} | \hat{V} \psi_m^{(0)} \right\rangle = v_{km} e^{i\omega_{km} t}.$$

The whole derivation will therefore be identical, except that the constant ω_{km} will be replaced by $\omega_{km} \pm \omega$. Hence, we have a new form of the Fermi Golden Rule for the probability per unit time of transition from the m-th to the k-th state:

THE FERMI GOLDEN RULE

$$w_m^k \equiv \frac{P_m^k(\tau)}{\tau} = |v_{km}|^2 \frac{2\pi}{\hbar} \delta\left(E_k^{(0)} - E_m^{(0)} \pm \hbar\omega\right). \qquad (2.36)$$

Note that $\hat{V}$ with $\exp(+i\omega t)$ needs the equality $E_k^{(0)} + \hbar\omega = E_m^{(0)}$, which means that $E_k^{(0)} \le E_m^{(0)}$ and therefore one has emission from the m-th to the k-th state. On the other hand, $\hat{V}$ with $\exp(-i\omega t)$ forces the equation $E_k^{(0)} - \hbar\omega = E_m^{(0)}$, which corresponds to absorption from the m-th to the k-th state.

Therefore a periodic perturbation is able to make a transition between states of different energy.

---

[56] In the homogeneous field approximation, the field interacts with the dipole moment of the molecule (cf. Chapter V2-4)

$$V(x,t) = V(x)\, e^{\pm i\omega t} = -\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}} e^{\pm i\omega t},$$

where $\boldsymbol{\mathcal{E}}$ denotes the electric field intensity of the light wave and $\hat{\boldsymbol{\mu}}$ is the dipole moment operator.

Summary

The Hamiltonian of any isolated system is invariant with respect to the following transformations (operations):

- any translation in time (homogeneity of time);
- any translation of the coordinate system (space homogeneity);
- any rotation of the coordinate system (space isotropy);
- inversion ($r \rightarrow -r$);
- reversing all charges (charge conjugation);
- exchanging labels of identical particles.

This means that the wave function corresponding to a stationary state (the eigenfunction of the Hamiltonian) also has to be an eigenfunction of the:

- total momentum operator (due to the translational symmetry);
- total angular momentum operator and one of its components (due to the rotational symmetry);
- inversion operator;
- any permutation (of identical particles) operator (due to the nondistinguishability of identical particles);
- $\hat{S}^2$ and $\hat{S}_z$ operators (for the nonrelativistic Hamiltonian [p. 77] due to the absence of spin variables in it).

Such a wave function corresponds to the energy belonging to the energy continuum.[57] Only after separation of the center-of-mass motion one obtains the spectroscopic states (belonging to a discrete spectrum) $\Psi_{N,J,M,\Pi}(\mathbf{r},\mathbf{R})$, where $N = 0, 1, 2...$ denotes the quantum number of the electronic state, $J = 0, 1, 2, ...$ quantizes the total angular momentum, M_J, $-J \leq M_J \leq J$ quantizes its component along the z axis, and $\Pi = \pm 1$ represents the parity with respect to the inversion. As to the invariance with respect to permutations of identical particles, an acceptable wave function has to be antisymmetric with respect to the exchange of identical fermions, whereas it has to be symmetric when exchanging identical bosons.

The time-independent Schrödinger equation $\hat{H}\psi = E\psi$ has been "derived" from the wave equation and the de Broglie formula. Solving this equation results in the stationary states and their energies. This is the basic equation of quantum chemistry. The prevailing weight of research in this domain is concentrated on solving this equation for various systems.

The time-dependent Schrödinger equation $\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t}$ represents the time evolution of an arbitrary initial wave function. The assumption that translation in time is a unitary operator leads to preserving the normalization of the wave function and of the mean value of the Hamiltonian. If the Hamiltonian is time-independent, then one obtains the formal solution to the Schrödinger equation by applying the operator $\exp(-\frac{it}{\hbar}\hat{H})$ to the initial wave function. The time evolution of the stationary state $\phi_m^{(0)}$ is most interesting in the case of suddenly switching on the perturbation $\hat{V}$. The state is no longer stationary and the wave function begins to change as time passes. Two cases have been considered:

---

[57] Because the molecule as a whole (i.e., its center of mass) may have an arbitrary kinetic energy. Sometimes it is rewarding to introduce the notion of the *quasicontinuum* of states, which arises if the system is enclosed in a large box instead of considering it in infinite space. This simplifies the underlying mathematics.

- time-independent perturbation and
- periodic perturbation.

Main concepts, new terms

atomic units (p. 74)
algebraic approximation (p. 106)
baryon number (p. 80)
bound state (p. 93)
charge conjugation (p. 85)
dipole moment (p. 82)
dynamic symmetry (p. 94)
enantiomers (p. 81)
energy-time uncertainty (p. 100)
Fermi Golden Rule (p. 113)
first-order perturbation theory (p. 110)
functions of class Q (p. 90)
gauge symmetry (p. 80)
invariance of theory (p. 80)
inversion (p. 81)
lepton number (p. 80)
mathematical solution (p. 95)
mirror reflection (p. 81)
molecular symmetry (p. 86)
periodic perturbation (p. 114)
physical solutions (p. 95)
rotational symmetry (p. 79)
Schrödinger equation (p. 87)
space isotropy (p. 79)
space homogeneity (p. 77)
spectroscopic state (p. 78)
spin conservation (p. 86)
stationary state (p. 87)
symmetry C (p. 86)
symmetry of the Hamiltonian (p. 72)
symmetry P (p. 81)
time evolution operator (p. 97)
time-independent perturbation (p. 112)
translational symmetry (p. 77)
two-state model (p. 106)
wave function evolution (p. 96)
wave function "matching" (p. 93)

From the research front

The overwhelming majority of research in the domain of quantum chemistry is based on the solution of the time-independent Schrödinger equation. Without computers it was possible to solve (in an approximate way) the equation for H_2^+ by conducting a hall full of secretaries with primitive calculators for many hours (what a determination). Thanks to computers, such problems became easy as early as the 1960s. Despite enormous progress in computer science, until the end of the 1980s the molecules studied were rather small when compared to expectations of experimentalists. They could be treated only as models, because they usually contained some substituents theoreticians were forced to consider irrelevant. The last years of the 20th century were marked by the unprecedented delivery by theoreticians of powerful high-tech efficient tools of quantum chemistry to other specialists: chemists, physicists, etc., but also to those who were not active in these field. The software computes millions of integrals and uses sophisticated mathematics, literally the whole arsenal of quantum chemistry, but users need not know about it. It is sufficient to make a mouse click on a quantum chemistry method icon.[58] Despite such progress,

---

[58] I hope all students understand that a quantum chemist has to be equipped with something more than a strong forefinger for clicking.

the time-dependent Schrödinger equation is solved extremely rarely. For the time being, researchers are interested mainly in stationary states. The quality of results depends on the size of the molecules investigated. Very accurate computations (accuracy ~ 0.01 kcal/mol) are feasible for the smallest molecules containing a few electrons, less accurate ones use first principles (*ab initio* methods) and are feasible for hundreds of atoms (accuracy to a few kcals/mol). Semiempirical quantum calculations[59] of even poorer accuracy are applicable to thousands of atoms.

Ad futurum

The numerical results routinely obtained so far indicate that, for the vast majority of chemical problems (yet not all, cf. Chapter 3) there is no need to search for a better tool than the Schrödinger equation. Future progress will be based on more and more accurate solutions for larger and larger molecules. The appetite is unlimited here, but the numerical difficulties increase much faster than the size of the system. However, progress in computer science has systematically opened new possibilities, always many times larger than previous ones. Some simplified alternatives to the Schrödinger equation (e.g., such as described in Chapter V2-3) will also be more important.

Laser π pulse and the dawn of new chemistry

Undoubtedly methods based on the time-dependent Schrödinger equation will also be developed.

The perturbation given by Eq. (2.31) leads in the resonance case to periodic oscillations given by Eq. (2.32). This means that starting from the pure wave function $\psi(x,0) = \left|\psi_1^{(0)}\right\rangle = |1\rangle$, after time $\frac{T}{2} = \frac{\pi}{v}$, we will get $\psi(x, \frac{T}{2}) = \left|\psi_2^{(0)}\right\rangle = |2\rangle$, which is the pure upper state. The situation, however, continues to change and, after the next time section $\frac{T}{2}$, the system comes back to $\psi(x,T) = \left|\psi_1^{(0)}\right\rangle = |1\rangle$, etc. This coming back is, however, not always desirable.

Chemists are already aware (see Chapter V2-6, p. V2-495) that what some other chemists used to see as chemical reaction products, is nothing else but reactants' excited electronic state, which became stabilized by some displacement of the nuclei (as the new ground state). This means that the traditional chemistry might be replaced in the future by knowledge on how to excite the reactants by using laser pulses in such a clever way that at the end one gets 100% pure products. Thus, the light may serve as an additional reactant, applied in the form of a sum (superposition) of the precisely planned laser pulses. These laser pulses would differ in frequency, intensity, polarization, duration, starting time, and time profile. Their role will be to change the reactants' wave function to the products' wave function.

As a prototype of this kind of action, one may imagine such a laser pulse, that in the two-state model transforms the lower-energy state to the higher-energy state with 100% yield (as the final result). To this end let us construct a perturbation of Eq. (2.31) that provides a very special v as a function of time, instead of a constant. The new $v = v_0 s(t)$, where $s(t)$ is the time profile, makes from the laser infinite

---

[59] In such calculations many integrals are approximated by simple formulae (sometimes involving experimental data), the main goal of which is efficiency.

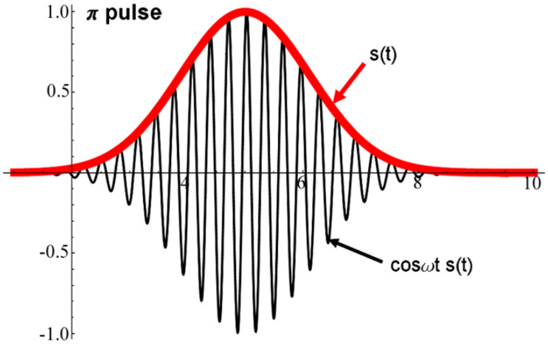


Fig. 2.13. A precisely tailored π pulse at the resonance frequency (transforming the initial stationary state of the system to the desired finite stationary state) will represent a basis for the future state-to-state chemistry.

wave (like $\cos \omega t$) a finite laser pulse of the electric field (multiplication of the oscillatory function by a bell-like profile[60] $\frown$, Fig. 2.13). The role of the time profile is to ensure that the wave function changes from the pure $\left|\psi_1^{(0)}\right\rangle = |1\rangle$ to the pure $\left|\psi_2^{(0)}\right\rangle = |2\rangle$ as it was before, but this time without any further oscillations. The weights $|c_1(t)|^2$ and $|c_2(t)|^2$ monotonically decrease and increase from 1 to 0 and from 0 to 1, respectively. Whatever the shape of s is, in order to have the 100% efficiency, one has to have[61] $\frac{|v_0|}{\hbar}\int_{-\infty}^{\infty} s(t)dt = \pi$, which explains the name "π pulse."

Of course, the π pulse procedure will fail (forbidden transition) when $v_0 = 0$. This may happen because of what is known in spectroscopy as selection rules[62] (cf. Appendix C). However, in many cases the π pulse procedure is allowed by the selection rules, thus creating a possibility for steering structural changes including chemical reactions, even chains of such reactions (by applying more complex laser pulses). Fig. 2.14 shows how such a steering could be achieved even without any quantum mechanical analysis (such an analysis made either *a priori* or *a posteriori* could be of great help anyway). The idea is to establish a fast feedback between the laser pulse series parameters and the internal parameters of

---

[60] The laser pulse time profile may have various shapes, e.g., it can be a single Gaussian function $\exp(-\frac{(t-t_0)^2}{\sigma^2})$ of width σ and centered at t_0 or $\sin^2(\pi\frac{t-t_0}{t_f})$ in the section $[t_0, t_f]$ (0 elsewhere; t_0 is the beginning and t_f the end of the pulse), etc.

[61] More generally, $\frac{|v_0|}{\hbar}\int_{-\infty}^{\infty} s(t)\mathrm{d}t = (2n+1)\pi$ for $n = 0, 1, 2, \ldots$ also does the job of 100% transformation; see P. Saalfrank, "*Theoretische Chemie III: Dynamik und Spektroskopie*," Universität Potsdam, 2008/2009, p. 83.

[62] The selection rules follow ultimately from the conservation laws, which have to be satisfied (for allowed transitions) and cannot be violated (as would be the case in transitions that are forbidden). Technically, for the forbidden transitions one gets a zero transition intensity, because of the symmetry properties of $\psi_1^{(0)}$ and $\psi_2^{(0)}$ and/or lack of coupling between, e.g., the electric field of the electromagnetic wave and the system dipole moment operator (cf. Appendix C, p. 605).

the neural network analyzer (or of another kind of optimization) in order to be able to keep improving the set of the laser parameters.[63]

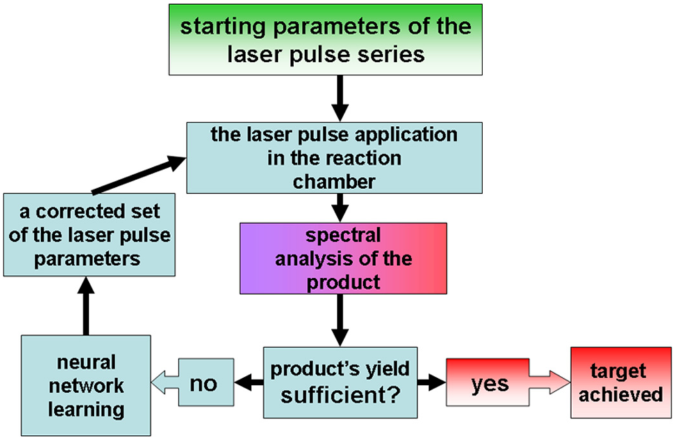


Fig. 2.14. Future laser chemistry with feedback.

How extensively could this future chemistry really be applied? Now I would like to ask my reader to give a chance to a bit of fantasy. Around us we see a spectacle of wonders: plants, animals, humans, and their functioning and interactions. This complicated world exists because plants are able to harvest photons coming from the Sun in a sophisticated coupling of π pulses, electronic rearrangements, vibrational and structural changes, and chemical reactions. What happens next is a masterpiece of Nature that leads to its incredible wealth we see around. My writing and your reading would not be possible without photons. Thus, no doubt, my friend, the steering by photons is feasible. We are only at the beginning of this adventure.

It seems that at the essence of science is the fundamental question "why?" and a clear answer to this question following from a deep understanding of Nature's machinery. We cannot tell a student, "well, this is what the computer says," because it implies that the computer understands, but what about you and me? Hence, interpretation of the results will be of crucial importance (a sort of Bader analysis, cf. Chapter V2-3). Progress here seems to be rather modest for the time being.

Additional literature

R. Feynman, "The Character of Physical Law, Cox and Wyman," Ltd, London, (1965).

The best recommendation is that the Feynman books need no recommendation.

---

[63] The neural network approach belongs to what is known as artificial intelligence. Its idea is to couple a certain number of input gates (of any nature) with a number of output gates (any input gate i connected with each output gate j with a link of weight w_{ij}). A "learning procedure" based on some teaching examples results in the successful set of w_{ij} (with the necessary nonlinearity of such a procedure). This set can then be used to predict the outputs for other examples.

J. Ciosłowski, in "Pauling's Legacy: Modern Modelling of the Chemical Bond," Elsevier, *Theor. Comput. Chem.*, 6(1999)1, eds. Z.B. Maksić, W.J. Orville-Thomas.

A concise presentation of the symmetry requirements.

P. Saalfrank, "Theoretische Chemie III: Dynamik und Spektroskopie," Universität Potsdam, 2008/2009.

A precise and clear presentation of molecular spectroscopy from the perspective of the time-dependent Schrödinger equation.

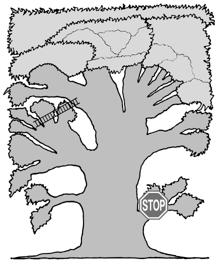
Questions

1. The conservation laws in physics:
   a. can be derived from invariance of the physical equations with respect to a symmetry operation.
   b. Emmy Noether derived the conservation of energy from the homogeneity of time.
   c. the conservation of angular momentum follows from the homogeneity of space.
   d. the conservation of total momentum follows from the homogeneity of space.
2. In atomic units:
   a. the Planck constant is assumed to be equal to 1.
   b. the electron rest mass $m_0 = 1$, the electron charge is equal to 1.
   c. the electron rest mass $m_0 = 1$, the electron charge is equal to $-e = -1$, $\frac{h}{2\pi} = 1$.
   d. the electron rest mass $m_0 = 1$, the electron charge is equal to $-e$, where $e = 1$, $\hbar = 1$.
3. For enantiomers A and B (A≠B):
   a. the interaction energy of AA is the same as BB and AB.
   b. the interaction energy of A with a chiral molecule C is the same as that of B with C.
   c. a nonchiral molecule when interacting with a chiral molecule may become chiral.
   d. one may create a chiral complex uniquely from some nonchiral molecules.
4. For functions of class Q:
   a. if $\psi(x) \in Q$, then $\int_{-\infty}^{+\infty} \psi^*(x)\,\psi(x)\,dx$ cannot be 0.
   b. if $\psi(x) \in Q$, then $\int_{-\infty}^{+\infty} \psi^*(x)\,\psi(x)\,dx = 1$.
   c. if $\psi(x) \in Q$, then $\int_{-\infty}^{+\infty} \psi^*(x)\,\psi(x)\,dx \leq \infty$.
   d. if $\psi(x) \in Q$, then $\int_{-\infty}^{+\infty} \psi(x)\,dx \leq 1$.
5. For the time-independent Schrödinger equation $\hat{H}\psi = E\psi$:
   a. $\hat{H}$ is the operator of the total energy represents a Hermitian operator.
   b. E and ψ must be real.
   c. E is real.
   d. $\int_{-\infty}^{+\infty} \psi^*(x)\,\psi(x)\,dx = 1$.
6. For the time-independent Schrödinger equation $\hat{H}\psi_n = E_n\psi_n$:
   a. $\int_{-\infty}^{+\infty} \psi_n^*(x)\,\psi_m(x)\,dx = \delta_{nm}$.
   b. one can choose such eigenfunctions of $\hat{H}$ that $\int_{-\infty}^{+\infty} \psi_n^*(x)\,\hat{H}\psi_m(x)\,dx = E_n\delta_{nm}$.
   c. one can choose such eigenfunctions of $\hat{H}$ that are orthonormal.
   d. if $E_n \neq E_m$, then $\langle\psi_n|\psi_m\rangle = 0$.

7. For the time-dependent Schrödinger equation $\hat{H}\Psi = i\hbar\frac{\partial\Psi}{\partial t}$:
   a. if $\hat{H}$ does not depend on time, its solution reads as $\psi(x,t) = \exp(-i\frac{\hat{H}}{\hbar}t)\psi(x,t=0)$.
   b. $\hat{H}\Psi = i\hbar\frac{\partial\Psi}{\partial t}$ is satisfied, only if $\hat{H}$ does not depend on time.
   c. any function $\psi(x,t) = \sum_n c_n\psi_n \exp(-i\frac{E_n}{\hbar}t)$, with the time-independent c_n and ψ_n satisfying $\hat{H}\psi_n = E_n\psi_n$, satisfies the time-dependent Schrödinger equation.
   d. for any mechanical quantity A one has $\frac{\mathrm{d}\langle\hat{A}\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle\left[\hat{H},\hat{A}\right]\right\rangle$.
8. In the two-state model (time-independent perturbation v, two orthonormal states with energy gap ω_{21}):
   a. a time-independent perturbation is unable to change the state of the system.
   b. when the energy level is doubly degenerate, the period of the oscillation is $T = \frac{2\pi}{|v|}$.
   c. when $\omega_{21} = 0$, one gets oscillation between the two states.
   d. for large and time-independent $|v|$ as compared to ω_{21}, one obtains a quasiresonance of the two energy levels.
9. In the two-state model (oscillating perturbation v with frequency ω, two orthonormal states with energy gap ω_{21}):
   a. in case of resonance one obtains the same coefficients $c_1(t)$ and $c_2(t)$ as in the case of degeneracy and the time-independent v.
   b. the resonance appears at $\omega = \omega_{21}$.
   c. if $\omega \neq \omega_{21}$ the starting wave function does not change in time.
   d. at the resonance $\omega = \omega_{21}$, the coefficients $c_1(t)$ and $c_2(t)$ do not change anymore.
10. The Fermi Golden Rule:
    a. has been derived by Dirac.
    b. pertains to the probability of changing the state due to perturbation applied.
    c. a time-independent perturbation enables transition from the state of higher energy to a state of lower energy.
    d. a time-dependent periodic perturbation may change a state to another state of different energy.

Answers

1abd, 2cd, 3cd, 4ac, 5ac, 6bcd, 7ac, 8bcd, 9ab, 10abd



CHAPTER 3

Beyond the Schrödinger Equation

> *The only reason for time is so that everything does not happen at once.*
> ***Albert Einstein***

Where are we?

The problems considered in the present chapter are shown as a small side branch at the base of the TREE.[1]

An example

Imagine yourself sitting with your friends in your brand-new luxurious red car made on your very special order with an extravagant feature: unlimited speed.[2] The manual says proudly, its exclusive quantum mechanical construction is based purely on the famous Schrödinger equation, in which $c = \infty$. Your colleagues are sure your car is much better than the corresponding cheap relativistic model for everybody. Well, it turned out recently[3] your nonrelativistic wonder car would be able to make only a unpleasant sound indicating dead battery. The reason would be a large relativistic effect in the electric potential difference between the lead electrode and the lead dioxide electrode. Your nonrelativistic battery would attain only something like 20% of the voltage the relativistic battery your colleagues have in their cars. Well, I presume you will enthusiastically agree with the author (together with millions of car drivers and passengers all over the world), that indeed there is a need to abandon the nonrelativistic theory and quest for a more accurate one.

Now, still another argument. Copper, silver and gold – many people would want to know everything about them (especially about the latter). The yellow shine of this metal has hypnotized humanity for centuries. Few people know that the color of gold, as calculated assuming infinite velocity of light, would be silver-like.[4]

---

[1] This chapter owes much to the presentation given by L. Pisani, J.-M. André, M.-C. André, E. Clementi, *J. Chem. Educ.*, 70(1993)894–901, as well as to the work of my friends J.-M. André, D.H. Mosley, M.-C. André, B. Champagne, E. Clementi, J.G. Fripiat, L. Leherte, L. Pisani, D. Vercauteren, M. Vracko, *Exploring Aspects of Computational Chemistry: vol. I, Concepts*, Presses Universitaires de Namur, p. 150–166 (1997), vol. II, Exercises, Presses Universitaires de Namur, p. 249–272 (1997).

[2] Of light.

[3] R. Ahuja, A. Blomqvist, P. Larsson, P. Pyykkö, P. Zaleski-Ejgierd, *Phys. Rev. Letters*, 106(2011)18301.

[4] P. Pyykkö, *Chem. Rev.*, 88(1988)563; also P. Pyykkö, *ibid.*, 97(1997)597.

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00011-2

The Schrödinger equation fails especially for heavy elements. Here is an example of three diatomics: Cu_2, Ag_2, Au_2 ($Z_{Cu} = 29$, $Z_{Ag} = 47$, $Z_{Au} = 79$).[5]

Bond length (Å)	Cu	Ag	Au
Nonrelativistic calculations	2.26	2.67	**2.90**
Relativistic calculations	2.24	2.52	2.44
Experimental results	2.22	2.48	**2.47**

The heavier the element, the larger is the error of the nonrelativistic approach. This is a huge discrepancy for such a quantity as bond length.

What is it all about?

---

[5] J.-M. André, M.-C. André, "*Une introduction à la théorie de la relativité classique et quantique à l'usage des chimistes*," Namur, 1999, p. 2.

Towards larger systems (▶ ◆) **p. 166**

Exploring beyond the Dirac equation... (▶ ◆✠) **p. 170**

- The Breit equation
- A few words about quantum electrodynamics

The greater the velocity of an object, the greater the errors in Newton dynamics. Electrons have greater velocity when close to nuclei of large electric charge.[6] This is why relativistic corrections may turn out to be important for heavy elements.

The Schrödinger equation is incompatible with special relativity theory. This has to be corrected somehow. This is far from being solved, but progress so far shows the Schrödinger equation, the spin of a particle, etc., in a new light.

Why is this important?

The subject of the present chapter addresses the very foundations of physics, and in principle has to be treated on an equal footing with the postulates of quantum mechanics. The Schrödinger equation of Chapter 2 does not fulfill (as will be shown in the present chapter) the requirements of relativity theory, and therefore is in principle "illegal." In the present chapter, Dirac's attempt to generalize the Schrödinger equation to adapt it to relativity theory will be described. If one assumes that particle velocities are small compared to that of light, then from this more general theory one obtains the Schrödinger equation. Also the notion of spin, which was introduced as a postulate in Chapter 1, follows as a natural consequence of the relativistic theory. One may draw the conclusion that the present chapter addresses to "the foundations of foundations" and therefore should occupy a prominent position in the TREE, instead of representing a small side branch (as it does now). However, the relativistic effects, even if visible in chemistry, do not play an important role in the case of the light elements (almost the whole of organic chemistry as well as almost the whole of biology). This is why I have chosen a rather pragmatic ("nonfundamentalist") way of presentation. This chapter is mainly for those readers who are interested in:

- "the foundations of foundations,"
- very accurate calculations for small atoms and molecules,
- calculations for systems containing heavy elements.

What is needed?

- The postulates of quantum mechanics (Chapter 1, necessary),

---

[6] This is easy to estimate. From Appendix H on p. 683 it follows that the mean value of the kinetic energy of an electron described by the $1s$ orbital in an atom of atomic number Z is equal to $\bar{T} = \frac{1}{2}Z^2$ (in a.u.). On the other hand, for a rough estimation of the electron velocity v, one may write $\bar{T} = \frac{mv^2}{2}$. This results in the expression $v = Z$ valid in a.u., while the velocity of light $c = 137.036$ a.u. The largest Z known hardly exceeds a hundred. It is seen, therefore, that if an atom with $Z > 137$ existed, then the $1s$ electrons would attain velocities exceeding the velocity of light. Even if this calculation is nothing but a rule of thumb, there is no doubt that when Z increases a certain critical Z value is approached (the so-called *relativistic mass effect*).

- operator algebra (Appendix A, p. 589, necessary),
- vector and scalar potentials (Appendix G, p. 673, necessary).

Classical works

In 1881 the American physicist Albert Michelson and in 1887 Albert Michelson with Edward Morley carried out some experiments showing that the speed of light is the same in the directions perpendicular and parallel to the Earth's orbit, i.e., the Earth's orbital velocity did not change the speed of light with respect to the Earth. The results were published in the *American Journal of Science*, 22(1881)120 under the title "*The Relative Motion of the Earth and the Luminiferous Aether*" and *ibid.*, 34(1887)333 (with the same title). ★ In 1889 the Irish physicist George Francis FitzGerald made the conjecture that if all moving objects were foreshortened in the direction of their motion, this would account for the strange results of the Michelson–Morley experiment. This was published in *Science*, 13(1889)390 with the title "*The Ether and the Earth's Atmosphere*." ★ The revolutionary special relativity theory (that explained this in detail) was developed by Albert Einstein in an article entitled "*Zur Elektrodynamik bewegter Körper*," published in *Annalen der Physik (Leipzig),* 17(1905)891. ★ The article is based largely on the ideas of the Dutchman Hendrik Antoon Lorentz, who independently of FitzGerald proposed the Lorentz transformation (of space and time) in 1904. The transformation accounted for the contraction of moving objects, as predicted by FitzGerald. The paper "*Electromagnetic Phenomena in a System Moving with any Velocity less than that of Light*" was published in *Proceedings of the Academy of Sciences of Amsterdam,* 6(1904)809. ★ The German mathematician Hermann Minkowski realized that the work of Lorentz and Einstein could best be understood using a non-Euclidean space of the space and time variables. His first paper on this subject was "*Die Grundgleichungen für die elektromagnetischen Vorgänge in bewegten Körper*," published in *Nachrichten der königlichen Gesellschaft der Wissenschaften zu Göttingen* (1908). ★ The Soviet physicist Vladimir A. Fock derived the first relativistic wave equation for a particle (published in *Zeitschrift für Physik*, 39(1926)226); then the German Walter Gordon did the same and also published in *Zeitschrift für Physik*, 40(1926)117. Finally, a similar theory was proposed independently by the Swede Oskar Klein in *Zeitschrift für Physik*, 41(1927)407. The Austrian Erwin Schrödinger also derived the same equation, and this is why it is sometimes called "*the equation with many fathers*." ★ A more advanced quantum mechanical theory (for a single particle) adapted to the principles of relativity was given by the British Paul Adrien Maurice Dirac in several articles in *Proceedings of the Royal Society A (London)* entitled "*The Fundamental Equations of Quantum Mechanics*," 109(1926)642, "*Quantum Mechanics and a Preliminary Investigation of the Hydrogen Atom*," *ibid.* 110(1926)561, "*The Quantum Theory of Radiation*," *ibid.* 114(1927)243, "*The Quantum Theory of the Electron*," *ibid.* 117(1928)610, and "*The Quantum Theory of the Electron. Part II*," *ibid.* 118(1928)351. ★ An extension of relativistic quantum theory to many-electron problems (still approximate) was published by the American Gregory Breit in *Physical Review* with the title "*The Effect of Retardation on the Interaction of Two Electrons*," 34(1929)553, and then in two other papers entitled "*Fine Structure of He as a Test of the Spin Interaction of Two Electrons*," *ibid.* 36(1930)383 and "*Dirac's Equation and the Spin-Spin Interactions of Two Electrons*," *ibid.* 39(1932)616. ★ In 1948 the Americans Richard Feynman and Julian Schwinger and the Japanese Shinichiro Tomonaga independently invented quantum electrodynamics (QED), which successfully combined quantum theory with the special theory of relativity and produced extremely accurate results.

* * *

3.1 A glimpse of classical relativity theory

3.1.1 The vanishing of apparent forces

The three principles of Newtonian[7] dynamics were taught us in school. The first principle, that a free body (with no acting force) moves uniformly along a straight line, seems to be particularly simple. It was not so simple for Ernest Mach though.

Mach wondered how one recognizes that no force is acting on a body. The contemporary meaning of the first principle of Newton dynamics is the following. First, we introduce a Cartesian coordinate system x, y, z to the Universe, and then remove from the Universe all objects except one, to avoid any interactions. Then, we measure equal time intervals using a spring clock and insert the corresponding positions of the body into the coordinate system (we are thus there with our clock and our ruler...). The first principle says that the positions of the body are along a straight line and equidistant. What a crazy procedure! The doubts and dilemmas of Mach were implanted in the mind of Albert Einstein.

Ernest Mach (1838–1916), Austrian physicist and philosopher, professor at the Universities of Graz, Prague, and Vienna, godfather of Wolfgang Pauli. Mach investigated supersonic flows. In recognition of his achievements the velocity of sound in air (1224 km/hour) is called Mach 1.

This Bern Patent Office employee also knew about the dramatic dilemmas of Lorentz, which we will discuss in a moment. Einstein recalls that there was a clock at a tram stop in Bern. Whenever his tram moved away from the stop, the modest patent office clerk asked himself what the clock would show if the tram had the velocity of light. Other passengers probably read their newspapers, but Einstein had questions which led humanity on new pathways.

Let us imagine two coordinate systems (each in 1D): O "at rest" (we assume it inertial[8]) while the coordinate system O' moves with respect to the first in a certain way (possibly very

---

[7] For Newton's biography see Chapter 7.

[8] That is, in which the Newton equation is satisfied. A coordinate system associated with an accelerating train is not inertial, because there is a nonzero force acting on everybody in the train, while the acceleration with respect to the train coordinate system is zero.

Albert Einstein (1879–1955), born in Ulm (Germany), studied at ETH, Zurich. He is considered by many as the greatest genius of all time. As a teenager and student, Einstein rejected many social conventions. This is why he was forced to begin his scientific career at a secondary position in the Federal Patent Office. Being afraid of his supervisor, he used to read books hidden in a drawer (he called the drawer the “Department of Physics”).

The year of his 26th birthday was particularly fruitful (“miraculous year” 1905). He published three fundamental papers: about relativity theory, about Brownian motion, and about the photoelectric effect. For the latter, Einstein received the Nobel Prize in 1921. After these publications he was appointed professor at the University of Zurich and then at the University of Prague. From 1914 Einstein headed the Physics Institute in Berlin, which was founded specially for him. He emigrated to the USA in 1933, because of menacing persecution due to his Jewish origin. Einstein worked at the Institute for Advanced Study in Princeton in the USA. He died there in 1955. According to his will, his ashes were dispersed over America from the air.

complicated). The position of the moving point may be measured in O giving the number x as the result, while in O' one gets the result x'. These numbers are related one to another (t is time) as follows:

$$x' = x + f(t). \tag{3.1}$$

If a scientist working in a lab associated with the coordinate system O would like to calculate the force acting on the abovementioned point body, he would get a result proportional to the acceleration, i.e., to $\frac{d^2x}{dt^2}$. If the same were done by another scientist working in a lab in O', then he would obtain *another force*, this time proportional to the acceleration computed as $\frac{d^2x'}{dt^2} = \frac{d^2x}{dt^2} + \frac{d^2f}{dt^2}$. The second term in this force is the *apparent* force. Such apparent forces (from the point of view of an observer on the ground) one encounters in lifts, on a carousel, etc.

Let us note an important consequence: if one postulates the same forces (and therefore the same dynamics) in two coordinate systems, $f(t)$ has to be a *linear* function (because its second derivative is equal to zero). This means that a family of all coordinate systems that moved uniformly with respect to one another would be characterized by the same description of phenomena because the forces computed would be the same (*inertial systems*).

Physics textbooks written in the two laboratories associated to O and O' would describe all the phenomena in the same way.

The linearity condition gives $x' = x + vt$. Let us take a fresh look of this equation: x' represents a linear combination of x and t, which means that *time and the linear coordinate mix together*. One has two coordinates: one in the O coordinate system and the other in the O' coordinate system. Wait a minute! Since the time and the coordinate are on an equal footing (they mix together), maybe one may also have the time (t) appropriate for (i.e., running in) the O and the time (t') running in the O' coordinate system?

Now follows a crucial step in the reasoning. Let us write *in a most general way* a linear transformation of coordinates and time (the forces computed in both coordinate systems are therefore the same):

$$x' = Ax + Bt,$$
$$t' = Cx + Dt.$$

First of all the corresponding transformation matrix *has to be* invertible (i.e., nonsingular), because inversion simply means exchanging the roles of the two coordinate systems and of the observers flying with them. Thus, one has

$$x = \bar{A}x' + \bar{B}t',$$
$$t = \bar{C}x' + \bar{D}t'.$$

Next, A has to be equal to $\bar{A}$, because the measurements of length in O and O', i.e., x and x', cannot depend on whether one looks at the O coordinate system from O', or at O' from O. If the opposite were true, then one of the coordinate systems would be privileged (treated in a special way). This, however, is impossible, because the two coordinate systems differ *only* in that O' flies from O with velocity v, while O flies from O' with velocity $-v$, but the space is isotropic. The same *has to happen* with the time measurements: on board O, i.e., t, and on board O', i.e., t', therefore $D = \bar{D}$. Since (from the inverse transformation matrix) $\bar{A} = \frac{D}{AD-BC}$ and $\bar{D} = \frac{A}{AD-BC}$, we have

$$\frac{D}{AD-BC} = A,$$
$$\frac{A}{AD-BC} = D.$$

From this $\frac{D}{A} = \frac{A}{D}$ follows, or

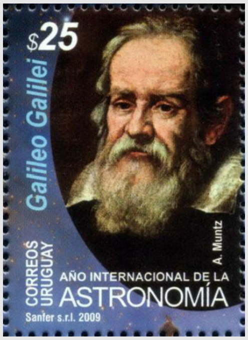
$$A^2 = D^2. \tag{3.2}$$

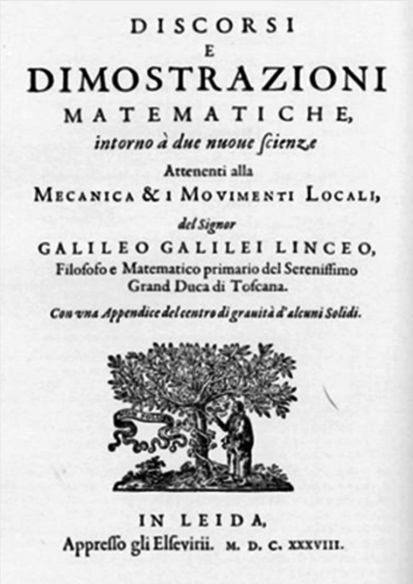
From the two solutions, $A = D$ and $A = -D$, one has to choose only $A = D$, because the second solution would mean that the times t and t' have opposite signs, i.e., when time run forwards in O it would run backwards in O'. Thus, we have

$$\boxed{A = D. \tag{3.3}}$$

3.1.2 The Galilean transformation

Galileo Galilei (1564–1642), Italian scientist, professor of mathematics at Pisa. Only those who have visited Pisa are able to appreciate the inspiration (for studying the free fall of bodies of different materials) from the *incredibly* leaning tower. Galileo's opus magnum (right-hand side) was published by Elsevier in 1638. Courtesy of Wikipedia (public domain).





DISCORSI
E
DIMOSTRAZIONI
MATEMATICHE,
intorno à due nuoue ſcienze
Attenenti alla
MECANICA & I MOVIMENTI LOCALI,
del Signor
GALILEO GALILEI LINCEO,
Filoſofo e Matematico primario del Sereniſſimo
Grand Duca di Toſcana.
Con vna Appendice del centro di grauità d'alcuni Solidi.

IN LEIDA,
Appreſſo gli Elſevirii. M. D. C. XXXVIII.

The equality condition $A = D$ is satisfied by the Galilean transformation, in which the two coefficients are equal to 1:

$$x' = x - vt$$
$$t' = t,$$

where position x and time t, say, of a passenger in a train, are measured in a fixed-platform coordinate system, while x' and t' are measured in a fixed-train coordinate system. There are no apparent forces in the two coordinate systems related by the Galilean transformation. Also, the Newtonian equation is consistent with our intuition, saying that time flows at the same pace in any coordinate system.

3.1.3 The Michelson–Morley experiment

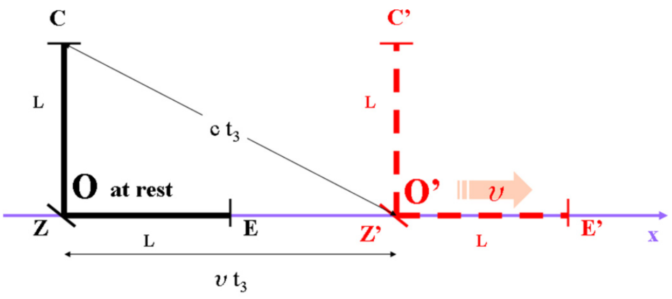
Hendrik Lorentz (1853–1928), Dutch scientist, professor at Leiden. Lorentz was very close to formulating the special theory of relativity. It was pointed out to Lorentz in 1894 that FitzGerald had published something similar. He wrote to FitzGerald, but the latter replied that indeed he had sent a half-page article to *Science*, but he did not know "*whether they ever published it*." Afterwards Lorentz took every opportunity to stress that FitzGerald was first to present the idea.

Hendrik Lorentz indicated that the Galilean transformation represents only *one possibility* of making the apparent forces vanish, i.e., assuring that $A = D$. Both constants need not to be equal to 1. As it happens, such a generalization is forced by an intriguing experiment performed in 1887.

Michelson and Morley were interested in whether the speed of light differs when measured in two laboratories moving with respect to one another. According to the Galilean transformation, the two velocities of light should be different, in the same way as the speed of train passengers (measured with respect to the platform) differs depending on whether they walk in the same or the opposite

Albert Michelson (1852–1931), American physicist, professor in Cleveland and Chicago, USA. He specialized in precise measurements of the speed of light.
Edward Williams Morley (1838–1923), American physicist and chemist, professor of chemistry at Western Reserve University in Cleveland, USA.

direction with respect to the train motion. Michelson and Morley replaced the train by Earth, which moves along its orbit around the Sun with a speed of about 40 km/s. Fig. 3.1 shows the Michelson–Morley experimental framework schematically. Let us imagine two identical right-angle V-shaped objects with all the arm lengths equal to L.



$$(c\,t_3)^2 = L^2 + (\upsilon t_3)^2$$

Fig. 3.1. The Michelson–Morley experimental framework. We have two identical V-shaped right-angle objects, each associated with a Cartesian coordinate system (with origins O and O'). The first (solid line) is at rest, while the second (dashed line) moves with velocity $v > 0$ with respect to the first (along the x axis). We are going to measure the velocity of light in two laboratories rigidly bound to the two coordinate systems. The mirrors are at the ends of the objects: C, E in O and C', E' in O', while at the origins two semitransparent mirrors Z and Z' are installed. Time $2t_3 \equiv t_\downarrow$ is the time for light to go down and up the vertical arm.

Each of the objects has a semitransparent mirror at its vertex[9] and ordinary mirrors at the ends. We will be interested in how much time it takes the light to travel along the arms of our objects (back and forth). One of the two arms of any object is oriented along the x axis, while the other one must be orthogonal to it. The mirror system enables us to overlap the light beam from the horizontal arm (x axis) with the light beam from the perpendicular arm. If there were any difference in phase between them we would immediately see the interference pattern.[10] The second object moves along x with velocity v (and is associated with coordinate system O') with respect to the first ("at rest," associated with coordinate system O).

---

[9] Such a mirror is made by covering glass with a silver coating.

[10] From my own experience I know that interference measurements are very sensitive. A laser installation was fixed to a steel table 10 cm thick concreted into the foundations of the Chemistry Department building, and the interference pattern was seen on the wall. My son Peter (then five years old) just touched the table with his finger. Everybody could see immediately a large change in the pattern, because the table bent.

3.1.4 The Galilean transformation crashes

In the following we will *suppose that the Galilean transformation is true.* In coordinate system O the time required for light to travel (round-trip) the arm along the x axis ($T_{\rightarrow}$) and that required to go perpendicularly to axis ($T_{\downarrow}$) are the same, i.e.,

$$T_{\rightarrow} = \frac{2L}{c},$$
$$T_{\downarrow} = \frac{2L}{c}.$$

Thus, in the O coordinate system, there will be no phase difference between the two beams (one coming from the parallel, the other from the perpendicular arm) and therefore no interference will be observed. Let us consider now a similar measurement in O'. In the arm colinear with x, when light goes in the direction of $\mathbf{v}$, it has to take more time (t_1) to get to the end of the arm,

$$ct_1 = L + vt_1, \tag{3.4}$$

than the time required to come back (t_2) along the arm,

$$ct_2 = L - vt_2. \tag{3.5}$$

Thus, the total round-trip time $t_{\rightarrow}$ is[11]

$$t_{\rightarrow} = t_1 + t_2 = \frac{L}{c-v} + \frac{L}{c+v} = \frac{L(c+v)+L(c-v)}{(c-v)(c-v)} = \frac{2Lc}{c^2-v^2} = \frac{\frac{2L}{c}}{1-\frac{v^2}{c^2}}. \tag{3.6}$$

What about the perpendicular arm in the coordinate system O'? In this case the time for light to go down (t_3) and up will be the same (let us denote total flight time by $t_{\downarrow} = 2t_3$, Fig. 3.1). Light going down requires time $t_3 = \frac{t_{\downarrow}}{2}$ to travel along the arm. Light goes along the hypotenuse of the rectangular triangle with sides L and $\frac{vt_{\downarrow}}{2}$ (because it goes down, but not only, since after $\frac{t_{\downarrow}}{2}$ it is found at $x = \frac{vt_{\downarrow}}{2}$). We will find, therefore, the time $\frac{t_{\downarrow}}{2}$ from the Pythagorean theorem:

$$(c\frac{t_{\downarrow}}{2})^2 = L^2 + (v\frac{t_{\downarrow}}{2})^2, \tag{3.7}$$

or

$$t_{\downarrow} = \sqrt{\frac{4L^2}{c^2-v^2}} = \frac{2L}{\sqrt{c^2-v^2}} = \frac{\frac{2L}{c}}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3.8}$$

---

[11] Those who have some experience with relativity theory will certainly recognize the characteristic term $1 - \frac{v^2}{c^2}$.

The times $t_\downarrow$ and $t_\rightarrow$ are not equal for the moving system, and there will be the interference we discussed a little earlier.

However, there is no interference!

Lorentz was forced to put the Galilean transformation into doubt (apparently the foundation of the whole science).

3.1.5 The Lorentz transformation

The interference predicted by the Galilean transformation is impossible, because physical phenomena would experience the two systems in a different way, while they differ only by their relative motions ($\mathbf{v}$ has to be replaced by $-\mathbf{v}$).

To have everything back in order, Lorentz assumed that, when a body moves, its length $L_\rightarrow$ (measured by using the unit length at rest in the coordinate system O) along the direction of the motion *contracts* according to equation

$$L_\rightarrow = L\sqrt{1-\frac{v^2}{c^2}}. \tag{3.9}$$

If we insert such a length, instead of L, in the expression for $t_\rightarrow$, then we obtain

$$t_\rightarrow = \frac{\frac{2L_\rightarrow}{c}}{1-\frac{v^2}{c^2}} = \frac{\frac{2L\sqrt{1-\frac{v^2}{c^2}}}{c}}{1-\frac{v^2}{c^2}} = \frac{\frac{2L}{c}}{\sqrt{1-\frac{v^2}{c^2}}} \tag{3.10}$$

and everything is perfect again: $t_\downarrow = t_\rightarrow$. No interference. This means that x' (i.e., the position of a point belonging to a rigid body as measured in O') and x (the position of the same point measured in O) have to be related by the following formula. The coordinate x measured by an observer in his O is composed of the intersystem distance OO', i.e., vt plus the distance O'–point, but measured using the length unit of the observer in O, i.e., the unit that resides in O (thus, *noncontracted* by the motion). Because of the contraction $1:\sqrt{1-\frac{v^2}{c^2}}$ of the rigid body the latter result will be *smaller* than x' (recall, please, that x' is what the observer measuring the position in his O' obtains), hence

$$x = x'\sqrt{1-\frac{v^2}{c^2}} + vt, \tag{3.11}$$

or

$$x' = \frac{x}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{vt}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{3.12}$$

which means that in the linear transformation assuring no apparent forces

$$A = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \tag{3.13}$$

$$B = -\frac{v}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3.14}$$

Of course, Professor O'connor in his laboratory O' would not believe that Professor Oconnor (sitting in his O lab) says that he (O'connor) has a contraction of the rigid body. And indeed, if Professor O'connor measured the rigid body using his standard length unit (he would not know his unit is contracted), then the length measured would be exactly the same as that measured just before separation of the two systems, when both systems were at rest. In a kind of retaliation, Professor O'connor could say that it is certainly not him who has the contraction, but his colleague Oconnor. He would be right, because for him, his system is at rest and his colleague Oconnor flies away from him with velocity $-v$. Indeed, Eq. (3.11) makes that very clear: an exchange $x \leftrightarrow x'$, $t \leftrightarrow t'$, and $v \leftrightarrow -v$ leads to the point of view of Professor O'connor,

$$x' = x\sqrt{1-\frac{v^2}{c^2}} - vt', \tag{3.15}$$

and one can indeed see an evident contraction of the rigid body of Professor Oconnor. This way, neither of these two coordinate systems is privileged. That is very, very good.

As we have already shown, in linear transformation $(x', t') \rightarrow (x, t)$ the diagonal coefficients have to be equal ($A = D$), therefore

$$t' = Cx + Dt, \tag{3.16}$$

$$D = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}. \tag{3.17}$$

To complete determination of the linear transformation we have to calculate the constant C (p. 129). Albert Einstein assumed that if Professors Oconnor and O'connor began (in their own coordinate systems O and O') measurements of the velocity of light, then despite the different distances to go (x and x') and different flight times[12] (t and t'), both scientists would get *the same velocity of light* (denoted by c).

In other words, $x = ct$ and $x' = ct'$.

Using this assumption and by Eqs. (3.12) and (3.17) we obtain

$$ct' = Dct - vDt, \tag{3.18}$$

while multiplying Eq. (3.16) for t' by c we get

$$ct' = cCx + Dct. \tag{3.19}$$

Subtracting both equations we have

$$0 = -vDt - cCx \tag{3.20}$$

or

$$C = -\frac{vtD}{cx} = -\frac{vtD}{cct} = -\frac{vD}{c^2}. \tag{3.21}$$

Thus we obtain the full Lorentz transformation, which on the one hand ensures no apparent forces and on the other hand ensures the same speed of light in *both* systems:

$$\begin{aligned} x' &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}x - \frac{v}{\sqrt{1-\frac{v^2}{c^2}}}t, \\ t' &= -\frac{v}{c^2}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}x + \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}t. \end{aligned}$$

Let us check first of all whether if $v = 0$, then everything is OK. Yes it is. Indeed, the denominator equals 1 and we have $t' = t$ and $x' = x$. Let us see what would happen if the velocity of light were equal to infinity. Then the Lorentz transformation becomes identical with the Galilean one. In general, after expanding t' and x' in a power series of v^2/c^2 we obtain

$$x' = -vt + x + \frac{1}{2}(-vt + x)\frac{v^2}{c^2} + \ldots,$$

---

[12] At the moment of separation $t = t' = 0$.

$$t' = t + (-\frac{x}{v} + \frac{t}{2})\frac{v^2}{c^2} + \dots.$$

This means that only at very high velocity v, we may expect differences between both transformations.

3.1.6 New law of adding velocities

Our intuition worked well for small velocities, much smaller than the velocity of light. The Lorentz transformation teaches us something which seems to contradict our intuition. What does it mean that the velocity of light is constant? Suppose we are flying with the velocity of light and send the light in the direction of our motion. Our intuition tells us: the light will have the velocity equal to 2c. Our intuition has to be wrong. How it will happen?

Let us see. We would like to have the velocity in the coordinate system O, but first let us find the velocity in the coordinate system O', i.e., $\frac{dx'}{dt'}$. From the Lorentz transformation one obtains, step by step,

$$\frac{dx'}{dt'} = \frac{\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}dx - \frac{v}{\sqrt{1-\frac{v^2}{c^2}}}dt}{-\frac{v}{c^2}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}dx + \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}dt} = \frac{\frac{dx}{dt} - v}{1 - \frac{v}{c^2}\frac{dx}{dt}}. \tag{3.22}$$

By extracting $\frac{dx}{dt}$ or using the symmetry relation (when $O' \to O$, then $v \to -v$) we obtain

$$\frac{dx}{dt} = \frac{\frac{dx'}{dt'} + v}{1 + \frac{v}{c^2}\frac{dx'}{dt'}} \tag{3.23}$$

or

VELOCITY ADDITION LAW

$$V = \frac{v' + v}{1 + \frac{vv'}{c^2}}. \tag{3.24}$$

In this way we have obtained a new rule of adding the velocities of the train and its passenger. Everybody naively thought that if the train velocity is v and the passenger velocity with respect to the train corridor is v', then the velocity of the passenger with respect to the platform is

$V = v + v'$. It turned out that this is not true. However, when both velocities are negligibly small with respect to c, then indeed one restores the old rule

$$V = v' + v. \tag{3.25}$$

Now, let us try to fool Mother Nature. Suppose our train is running with the velocity of light, i.e., $v = c$, and we take out a torch and shine the light forward, i.e., $\frac{dx'}{dt'} = v' = c$. What will happen? What will be the velocity V of the light with respect to the platform? $2c$? From (3.24) we have $V = \frac{2c}{2} = c$. This is precisely what is called the universality of the speed of light. Now, let us make a bargain with Nature. We are dashing in the train with the speed of light $v = c$ and walking along the corridor with velocity $v' = 5$ km/h. What will our velocity be with respect to the platform? Let us calculate again:

$$\frac{dx}{dt} = \frac{5 + c}{1 + \frac{c}{c^2}5} = \frac{5 + c}{1 + \frac{5}{c}} = c\frac{5 + c}{5 + c} = c. \tag{3.26}$$

Once more we have been unable to exceed the speed of light c. One last attempt. Let us take the train velocity as $v = 0.95c$, and fire along the corridor a powerful missile with speed $v' = 0.10c$. Will the missile exceed the speed of light or not? We have

$$\frac{dx}{dt} = \frac{0.10c + 0.95c}{1 + \frac{0.95c}{c^2}0.10c} = \frac{1.05c}{1 + 0.095} = \frac{1.05}{1.095}c = 0.9589c, \tag{3.27}$$

so c is not exceeded. Wonderful formula.

3.1.7 The Minkowski space–time continuum

The Lorentz transformation may also be written as

$$\begin{bmatrix} x' \\ ct' \end{bmatrix} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \begin{bmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{bmatrix} \begin{bmatrix} x \\ ct \end{bmatrix}.$$

What would happen if the roles of the two systems were interchanged? To this end let us express x, t by x', t'. By inversion of the transformation matrix we obtain[13]

$$\begin{bmatrix} x \\ ct \end{bmatrix} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \begin{bmatrix} 1 & \frac{v}{c} \\ \frac{v}{c} & 1 \end{bmatrix} \begin{bmatrix} x' \\ ct' \end{bmatrix}. \tag{3.28}$$

---

[13] You may check this by multiplying the matrices of both transformations – we obtain the unit matrix.

We have perfect symmetry, because it is clear that the sign of the velocity has to change. Therefore,

none of the systems is privileged.

Now let us come back to Einstein's morning tram meditation[14] about what he would see on the tram stop clock if the tram had the velocity of light. Now we have the tools to solve the problem. It concerns to two events – two ticks of the clock observed in the coordinate system associated with the tram stop, i.e., $x_1 = x_2 \equiv x$, but happening at two different times t_1 and t_2 (differing by, say, one second, i.e., $t_2 - t_1 = 1$; this is associated with the corresponding movement of the clock hand). What will Einstein see when his tram leaves the stop with velocity v with respect to the stop, or in other words when the tram stop moves with respect to him with velocity $-v$? He will see the same two events, but in his coordinate system they will happen at $t_1' = \frac{t_1}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}}$ and $t_2' = \frac{t_2}{\sqrt{1-\frac{v^2}{c^2}}} - \frac{\frac{v}{c^2}x}{\sqrt{1-\frac{v^2}{c^2}}}$, i.e., *according to the tram passenger* the two ticks at the tram stop will be separated by the time interval

$$t_2' - t_1' = \frac{t_2 - t_1}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Thus, when the tram ran through the streets of Bern with velocity $v = c$, the hands on the tram stop clock when seen from the tram would not move at all, and this second would be equivalent to eternity.

This effect is known as *time dilation*. Of course, for the passengers waiting at the tram stop and watching the clock, its two ticks would be separated by exactly one second. If Einstein took his watch out of his waistcoat pocket and showed it to them through the window *they* would be amazed. The seconds will pass at the tram stop, while Einstein's watch would seem to be stopped. The effect we are describing has been double checked experimentally many times. For example, the meson lives such a short time (in the coordinate system associated with it), that when created by cosmic rays in the stratosphere, it would have no chance of reaching a surface

---

[14] Even today Bern looks quite provincial. In the center Albert Einstein lived at Kramgasse 49. A small house, squeezed by others, next to a small cafe, now with Einstein's achievements on the walls. Einstein's small apartment is on the second floor showing a room facing the backyard, in the middle a child's room (Einstein lived there with his wife Mileva Marič and their son Hans Albert; the personal life of Einstein was complicated), and a large living room facing the street. A museum employee with oriental features says the apartment looks as it did in the "miraculous year 1905," everything is the same (except the wallpaper, she adds), and then: "*maybe this is the most important place for the history of science*."

laboratory before decaying. Nevertheless, as seen from the laboratory coordinate system, the meson's clock ticks very slowly and mesons are observable.

Hermann Minkowski (1864–1909), German mathematician and physicist, professor in Bonn, Königsberg, and Technische Hochschule Zurich, and from 1902 professor at the University of Göttingen. This teacher of Einstein concluded: "*Space of itself and time of itself will sink into mere shadows, and only a kind of union between them shall survive.*"

Hermann Minkowski introduced the seminal concept of the four-dimensional space–time continuum (x, y, z, ct).[15] In our one-dimensional space, the elements of the Minkowski space–time continuum are events, i.e., vectors (x, ct), something happens at space coordinate x at time t, when the event is observed from coordinate system O. When the same event is observed in two coordinate systems, then the corresponding x, t and x', t' satisfy the Lorentz transformation. It turns out that in both coordinate systems *the distance of the event from the origin of the coordinate system is preserved*. The square of the distance is calculated in a strange way, i.e.,

$$(ct)^2 - x^2, \tag{3.29}$$

for the event (x, ct). Indeed, let us check carefully:

$$\begin{aligned}(ct')^2 - (x')^2 &= \frac{1}{1-\frac{v^2}{c^2}}(-\frac{v}{c}x + ct)^2 - \frac{1}{1-\frac{v^2}{c^2}}(x - \frac{v}{c}ct)^2 \\ &= \frac{1}{1-\frac{v^2}{c^2}}[\frac{v^2}{c^2}x^2 + c^2t^2 - 2vxt - x^2 - \frac{v^2}{c^2}c^2t^2 + 2vxt] \\ &= \frac{1}{1-\frac{v^2}{c^2}}[\frac{v^2}{c^2}x^2 + c^2t^2 - x^2 - \frac{v^2}{c^2}c^2t^2] = (ct)^2 - (x)^2.\end{aligned} \tag{3.30}$$

There it is! This equation and Eq. (3.28) enabled Hermann Minkowski to interpret the Lorentz transformation (3.28) as a *rotation* of the event (x, ct) in the Minkowski space about the origin of the coordinate system (since any rotation preserves the distance from the rotation axis).

---

[15] Let me report a telephone conversation between the PhD student Richard Feynman and his supervisor Professor Archibald Wheeler from Princeton Advanced Study Institute (according to Feynman's Nobel Lecture, 1965). Wheeler: "*Feynman, I know why all electrons have the same charge and the same mass!*" F: "*Why?*" W: "*Because they are all the same electron!*" Then, Wheeler explained: "*suppose that the world lines which we were ordinarily considering before in time and space – instead of only going up in time were a tremendous knot, and then, when we cut through the knot by the plane corresponding to a fixed time, we would see many, many world lines and that would represent many electrons.*"

Light cone

All events – the elements of the Minkowski space – can be classified according to what is known as a light cone diagram (Fig. 3.2). We choose an Observer as residing at the origin of the Minkowski space. He is not able to get any message from such events that would require the messenger velocity to exceed the velocity of light (the regions outside the cones). This region is inaccessible for the Observer, and he cannot and will not be able to say anything about this part of Minkowski events.

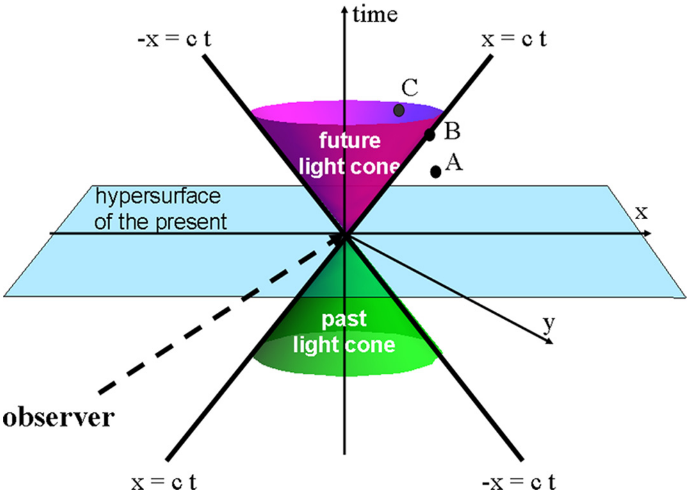


Fig. 3.2. Light cone. An event $(x, ct) = (0, 0)$ for an Observer is chosen as the origin of the Minkowski space–time. Any event A outside the region of the two cones (like, e.g., outside the lines $x = ct$ and $-x = ct$) is inaccessible for the Observer. This is because to get to A from the origin the Observer would necessarily be able to send something with a speed higher than c. Any event of the upper cone is accessible for the Observer (in the future) either by traveling or sending an object with speed $|v| < c$ (inside the upper cone, see point C) or with $|v| = c$ (as shown by B). Similarly, the Observer may get a message from the past, but only from an event belonging to the lower cone, either by receiving a photon ($|v| = c$) or a massive particle (slower than c).

As time goes on, whatever the Observer knows about other events belongs to the past (lower cone), because the message he receives traveled at most at the speed of light. When we look around we do not see how it is, but how it was. Indeed, if we look at stars, we see them not as they are, but as they were (sometimes billions of years ago). Therefore, we are unable to see objects (if any) for which the light has not had enough time to tell us about their existence. Some events in the past cannot affect us.

Similarly, we can affect the future events (upper cone[16]) only if we are able to send a messenger (his speed necessarily is smaller than c).

3.1.8 How do we get $E = mc^2$?

The Schrödinger equation is invariant with respect to the Galilean transformation. Indeed, the Hamiltonian contains the potential energy which depends on interparticle distances, i.e., on the differences of the coordinates, whereas the kinetic energy operator contains the second derivative operators which are invariant with respect to the Galilean transformation. Also, since $t = t'$, the time derivative in the time-dependent Schrödinger equation does not change.

Unfortunately, both Schrödinger equations (time-independent and time-dependent) are not invariant with respect to the Lorentz transformation and therefore are illegal. As a result, one cannot expect the Schrödinger equation to describe accurately objects that move with velocities comparable to the speed of light.

Let us consider a particle moving in the potential V. The Schrödinger equation has been "derived" (see p. 87) from the total energy expression,

$$E = \frac{p^2}{2m} + V, \tag{3.31}$$

where p is the momentum and m the mass.

Einstein was convinced that nothing could be faster than light.[17] Therefore, what would happen if a particle were subject to a constant force? It would eventually attain the velocity of light and what would happen afterwards? There was a problem, and Einstein assumed that in the laboratory coordinate system in which the particle is accelerated, the particle will increase its mass. In the coordinate system fixed on the particle no mass increase will be observed, but in the laboratory system it will. We have to admire Einstein's courage. For millions of people the mass of a body represented an invariant characteristic of the body. How was the mass supposed to increase? Well, the increase law – Einstein reasoned – should be such that the particle was able to absorb *any* amount of the kinetic energy. This means that when $v \to c$, then we have to have $m(v) \to \infty$. One of the possible formulae for $m(v)$ may contain a factor typical of

---

[16] All events on the surface of the future and past cones (see, e.g., $x = \pm ct$) share the same property: their distance $\sqrt{(ct)^2 - x^2}$ from the origin of the Minkowski space is equal to 0.

[17] Maybe this is true, but nothing in the special theory of relativity forces such a statement.

relativity theory (cf. Eq. (3.17)):

$$m(v) = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}, \tag{3.32}$$

where m_0 is the so-called rest mass of the particle (i.e., its mass measured in the coordinate system residing on the particle). It is seen that if v/c were zero (as it is in the nonrelativistic world), then m would be equal to m_0, i.e., to a constant, as it is in nonrelativistic physics.[18]

For the time being, the legitimacy of Eq. (3.32) is questionable as being just one of the possible *ad hoc* suppositions. However, Einstein has shown that this particular formula fits the existing equation of motion. First, after expanding the mass into the Taylor series one obtains something interesting,

$$m(v) = m_0\{1 + \frac{1}{2}\frac{v^2}{c^2} + \frac{3}{8}\frac{v^4}{c^4} + ...\}, \tag{3.33}$$

especially after multiplying the result by c^2:

$$mc^2 - m_0c^2 = \frac{m_0v^2}{2} + \text{smaller terms}. \tag{3.34}$$

It looks as if indeed the kinetic energy was stored directly in the mass m. Einstein deduced that it may be that the total kinetic energy of the body is equal to

$$\boxed{E = mc^2.}$$

---

[18] And therefore no corrections to the Schrödinger equation are needed. At the beginning of this chapter we arrived at the conclusion that electron velocity in an atom is close to its atomic number Z (in a.u.). Hence, for the hydrogen atom ($Z_H = 1$) one may estimate $v/c \simeq 0.7\%$, i.e., v of the electron in the 1s state represents a velocity of the order of 2100 km/s, which is probably very impressive for a car driver, but not for an electron. However, for gold ($Z_{Au} = 79$) we obtain $v/c \simeq 51\%$. This means that in the atom of gold the electron mass is larger by about 15% with respect to its rest mass and therefore the relativistic effects are nonnegligible. For such important elements as C, O, and N (biology) the relativistic corrections may be safely neglected. A young graduate student, Grzegorz Łach, posed an interesting purely academic question (such questions and the freedom to discuss them represent the cornerstone and the beauty of university life): will the human body survive the switching off of relativistic effects? Most of the biomolecules would function practically without significant changes, but the heavy metal atoms in enzyme active sites might direct differently the chemical reactions in which they are involved in. Would they indeed? Would the new direction be destructive for the body? Nobody knows. On the other hand, we have forgotten about the spin concept that follows in a consequent way only from relativistic quantum theory. Without spin no world similar to ours is conceivable.

He convinced himself about this after calculating its time derivative. After assuming that Eq. (3.32) is correct, one obtains

$$\frac{dE}{dt} = c^2\frac{dm}{dt} = c^2\frac{d}{dt}\frac{m_0}{\sqrt{1-\frac{v^2}{c^2}}} = m_0c^2\frac{d}{dt}\frac{1}{\sqrt{1-\frac{v^2}{c^2}}} = m_0c^2\left(-\frac{1}{2}\right)\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}\frac{-2v}{c^2}\frac{dv}{dt} =$$

$$m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}v\frac{dv}{dt} = \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}\frac{1}{1-\frac{v^2}{c^2}}v\frac{dv}{dt} = \frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}\left(1+\frac{\frac{v^2}{c^2}}{1-\frac{v^2}{c^2}}\right)v\frac{dv}{dt} =$$

$$\frac{m_0}{\sqrt{(1-\frac{v^2}{c^2})}}v\frac{dv}{dt} + \frac{v^2}{c^2}m_0\left(1-\frac{v^2}{c^2}\right)^{-\frac{3}{2}}v\frac{dv}{dt} = mv\frac{dv}{dt} + v^2\frac{dm}{dt} = v\frac{d(mv)}{dt}.$$

Precisely the same equation is satisfied in nonrelativistic mechanics, if E denotes the *kinetic energy*:

$$\frac{dE}{dt} = \frac{d(\frac{mv^2}{2})}{dt} = \frac{1}{2}m2v\frac{dv}{dt} = v\frac{d(mv)}{dt}. \tag{3.35}$$

Therefore in relativity theory

$$E_{kin} = mc^2. \tag{3.36}$$

This formula has been verified in laboratories many times. For example, it is possible nowadays to accelerate electrons in cyclotrons up to a velocity that differs from c by $\frac{1}{8\,000\,000}c$. That corresponds to $1-\frac{v^2}{c^2} = \frac{1}{4\,000\,000}$, and the electron's mass m becomes 2000 times larger than its m_0. This means that the electron is pumped up with energy to such an extent that its mass is similar to that of the proton. The energy stored in mass is huge. If, from the mass of a 20 000 TNT atomic bomb, one subtracted the mass of its ashes after explosion,[19] then one would obtain only about 1 g! The energy freed from this 1 g gives an effect similar to the apocalypse.

3.2 Towards relativistic quantum mechanics

The equation of many fathers

We would like to express the kinetic energy E_{kin} through the particle's momentum p, because we would then know how to obtain the corresponding quantum mechanical operators (Chap-

---

[19] R. Feynman, R.B. Leighton, M. Sands, "*Feynman Lectures on Physics*," Addison-Wesley Publishing Company, 1964.

ter 1, p. 22). To this end let us consider the expression

$$E_{kin}^2 - (m_0c^2)^2 = m^2c^4 - m_0^2c^4 = m_0^2c^4\left(\frac{1}{1-v^2/c^2} - 1\right) = \frac{m_0^2c^4}{1-v^2/c^2}\frac{v^2}{c^2}$$
$$= m^2v^2c^2 = p^2c^2. \tag{3.37}$$

Therefore,

$$E_{kin} = c\sqrt{p^2 + m_0^2c^2} \tag{3.38}$$

and the total energy E in the external potential V

$$E = c\sqrt{p^2 + m_0^2c^2} + V. \tag{3.39}$$

What if the particle is subject to an electromagnetic field, given by the electric field $\boldsymbol{\mathcal{E}}$ and the magnetic field $\mathbf{H}$ (or, the magnetic induction $\mathbf{B}$) in every point of the space. Instead of $\boldsymbol{\mathcal{E}}$ and $\mathbf{H}$ (or $\mathbf{B}$) we may introduce two other quantities: the vector field $\mathbf{A}$ and the scalar field ϕ (see Appendix G). As we can show in classical electrodynamics,[20] the kinetic energy of the particle subject to an electromagnetic field is very similar to the same expression without the field (Eq. (3.38)), namely, for a particle of charge q, the momentum $\mathbf{p}$ is to be replaced by $\mathbf{p} - \frac{q}{c}\mathbf{A}$ and the potential V by $q\phi$. Therefore, we obtain the following expression for the total energy of the particle in an electromagnetic field:

$$E = c\sqrt{(\mathbf{p} - \frac{q}{c}\mathbf{A})^2 + m_0^2c^2} + q\phi, \tag{3.40}$$

where $\mathbf{A}$ and ϕ represent functions of the particle's position.

If we wanted to use the last expression to construct the Hamiltonian, then we would find serious difficulty, namely, the momentum operator $\hat{\mathbf{p}} = -i\hbar\nabla$ (replacing $\mathbf{p}$ according to Postulate II, Chapter 1) is under the square root, thus leading to nonlinear operators. Brave scientists noted, however, that if someone made a square, then the danger would disappear. We would obtain

$$(E - q\phi)^2 = c^2\left[(\mathbf{p} - \frac{q}{c}\mathbf{A})^2 + m_0^2c^2\right]. \tag{3.41}$$

All that has been and still is a sort of groping and guessing from some traces or indications.

---

[20] For example, H.F. Hameka, *Advanced Quantum Chemistry*, Addison-Wesley, Reading, 1965, p. 40.

The equations corresponding to physical quantities will be transformed to the corresponding operator equations, and it will be assumed that both sides of them will act on a wave function.

Oskar Klein (1894–1977) was the youngest son of the chief rabbi of Sweden and professor of mathematics and physics at Stockholm Högskola. Walter Gordon (1893-1940) until 1933 was a professor at the University of Hamburg, and afterwards resided in Sweden.

What to insert as the operator $\hat{H}$ of the energy E? This was done by Schrödinger (before Fock, Klein, and Gordon; this is why it is also known as the "equation of many fathers"). Schrödinger inserted what he had on the right-hand side of his time-dependent equation $\hat{H}\Psi = i\hbar\frac{\partial}{\partial t}\Psi$, i.e., $\hat{H} = i\hbar\frac{\partial}{\partial t}$.

This way

$$(i\hbar\frac{\partial}{\partial t} - q\phi)^2 = c^2\left[(-i\hbar\nabla - \frac{q}{c}\mathbf{A})^2 + m_0^2c^2\right], \tag{3.42}$$

or after acting on the wave function we obtain the Fock–Klein–Gordon equation,

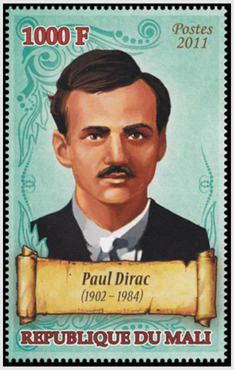
$$(i\hbar\frac{\partial}{\partial t} - q\phi)^2\Psi = c^2\left[(-i\hbar\nabla - \frac{q}{c}\mathbf{A})^2 + m_0^2c^2\right]\Psi. \tag{3.43}$$

This equation has at least one advantage over the Schrödinger equation: ct and x, y, z enter the equation on equal footing, which is required by special relativity. Moreover, the Fock–Klein–Gordon equation is invariant with respect to the Lorentz transformation, whereas the Schrödinger equation is not. This is a prerequisite of any relativity-consistent theory and it is remarkable that such a simple derivation made the theory invariant. The invariance does not mean however that the equation is accurate. The Fock–Klein–Gordon equation describes a boson particle, because Ψ is a usual scalar-type function, in contrast to what we will see in a moment in the Dirac equation.

3.3 The Dirac equation

3.3.1 The electronic sea and the day of glory

Paul Adrien Maurice Dirac (1902–1984), British physicist theoretician, professor at universities in Cambridge, and then Oxford. Dirac was keen on hiking and climbing. He used to practice before expeditions by climbing trees near Cambridge, in the black outfit in which he always gave his lectures.
He spent his last years at the University of Tallahassee (Florida, USA). On being guided through New York City, Dirac remembered old times. The guide remarked that there were visible changes, among others the buses had been painted pink. Dirac quietly agreed, adding that indeed they had, at least from one side...



Paul Dirac used the Fock–Klein–Gordon equation to derive a Lorentz transformation-invariant equation[21] for a single fermion particle. The Dirac equation is solvable only for several very simple cases. One of them is the free particle (Dirac), the other is an electron in the electrostatic field of a nucleus (Darwin, not *that* one).

Charles Galton Darwin (1887–1962), British physicist and mathematician, professor at University of Edinburgh, Scotland, grandson of the evolutionist Sir Charles Robert Darwin. Darwin investigated the scattering of α particles on atoms. Courtesy of Dr. R.C. McGuiness, National Physical Laboratory, UK.

One may add here a few other systems, e.g., the harmonic oscillator, and that's it.

From Eq. (3.41), in the case of a free particle, one obtains two sets of energy eigenvalues, one corresponding to the negative energies

$$E = -\sqrt{p^2c^2 + m_0^2c^4} \tag{3.44}$$

and the other corresponding to the positive energies

$$E = +\sqrt{p^2c^2 + m_0^2c^4}. \tag{3.45}$$

---

[21] See J.D. Bjorken, S.D. Drell, "*Relativistic Quantum Mechanics*," McGraw-Hill, 1964.

Dirac was not worried by the fact that both roots appear after an *ad hoc* decision to square the expression for the energy (Eqs. (3.40) and (3.41)). As we can see, since the momentum may change from 0 to ∞ ($p = mv$, and for $v \to c$ we have $m \to \infty$), we have the *negative energy continuum* and the symmetrically located *positive energy continuum*, both continua separated by the energy gap $2m_0c^2$ (Fig. 3.3).

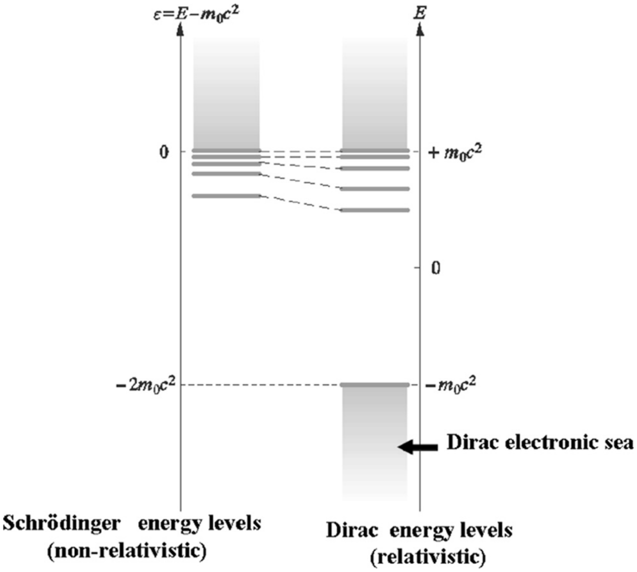


Fig. 3.3. Energy levels for the hydrogen atom according to Schrödinger (left-hand side) and Dirac (right-hand side). Shadowed areas correspond to the positive and negative energy continua.

Dirac (when 26 years old) made an absurd assumption that what people call a vacuum is in reality a sea of electrons occupying the negative energy continuum ("*Dirac electronic sea*"). The sea was supposed to consist of an infinite number of electrons, which had to imply catastrophic consequences concerning, for example, the mass of the Universe (infinite), but Dirac did not doubt and was not discouraged. "*We see only those electrons, that have positive energy*," said Dirac. Why he was so determined? Well, Dirac's concept of the sea was proposed to convince us that due to the Pauli exclusion principle the doubly occupied sea electronic states are simply inaccessible for an extra electron, which is therefore forced to have positive energy. All this looks as no more than an *ad hoc* speculation! Dirac could remove these difficulties by resigning from the negative spectrum, but he did not![22]

---

[22] It seems that no one of many fathers of the Fock–Klein–Gordon equation did dare to take into account also another possibility, the one with the negative square root in Eq. (3.40), a step made by Paul Dirac. In the case of many fathers though the argument about the electron sea and the Pauli exclusion principle would not work, since we have to do with bosons! We would have an abyss of negative energies, a disaster for the theory.

Then, consequently continuing his reasoning, Dirac asked whether he could somehow see those electrons that occupy the sea and answered that yes, it is possible. According to Dirac it is sufficient to excite such an electron by providing the energy of the order of $2m_0c^2$ to cover the energy gap (the energy $2m_0c^2$ is very large, of the order of 1 MeV). Then the sea electron would have the positive energy and could be observed as other electrons with positive energy. However, besides the electron, one would leave a hole in the Dirac sea. Dirac was severely criticized because of what this strange hole would correspond to in experimental physics. Once, when pushed too far, he said desperately that this is a proton. Some seemed to be satisfied, but others began to attack him furiously. However, then Dirac demonstrated that the hole would have the dynamic and electric properties of an electron, except that its sign would be opposite.[23] This was nothing less than a *hypothesis for the existence of antimatter, a state of matter unknown at that time*. Please imagine the shock, when three years later Carl Anderson reported the creation of electron–positron pairs from a vacuum after providing an amount of energy of $2m_0c^2$. This was the day of glory for Dirac and his quantum theory.

Carl David Anderson (1905–1991), American physicist, professor at the Pasadena Institute of Technology. In 1932 Anderson discovered the positron, for which he received the Nobel Prize in 1936. He was also a codiscoverer of the muon. ©The Nobel Foundation.

In a moment we will see the determination with which Dirac attacked the Fock–Klein–Gordon operator equation, which we will write down a little differently:

$$\left[\frac{i\hbar\frac{\partial}{\partial t}-q\phi}{c}\right]^2-\left[(-i\hbar\nabla-\frac{q}{c}\mathbf{A})^2+m_0^2c^2\right]=0. \tag{3.46}$$

Let us first introduce the following abbreviations:

$$\pi_0=\frac{i\hbar\frac{\partial}{\partial t}-q\phi}{c}, \qquad \pi_\mu=-i\hbar\frac{\partial}{\partial\mu}-\frac{q}{c}A_\mu \tag{3.47}$$

for $\mu=x,y,z$ or 1, 2, 3.

---

[23] Paul Dirac, when a pupil in primary school, made his reputation after solving a riddle which very well fits the person who invented the positively charged electron (positron).
Three fishermen went overnight fishing and camping at a lake. After a day of heavy fishing, in the evening they put the fish in a bucket and, tired, fell asleep in the tent. At midnight one of the fishermen woke up and, tired of the whole escapade, decided to take $\frac{1}{3}$ of all the fish, leave the tent quietly, and go home. When he counted the fish splashing in the bucket, it turned out that the number of fish was indivisible by 3. However, when he had thrown one fish back to the lake, the number was divisible by 3, he took his $\frac{1}{3}$ and went away. After a while a second fisherman woke up and did the same, and then the third. The question was, how many fish were in the bucket. Dirac's answer was -2.

Dirac persisted in treating Eq. (3.46) as $a^2 - b^2$ and therefore rewriting it in the form $(a + b)(a - b)$, i.e.,

$$\left(\pi_0 + \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\left(\pi_0 - \sum_{\mu=x,y,z} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right) = 0. \tag{3.48}$$

He was so self-assured that he said Eq. (3.48) has to be satisfied at any price by finding suitable unknowns α_i (independent of coordinates and time). The α's have to satisfy the following relations (anticommutation relations):

$$\alpha_\mu^2 = 1, \tag{3.49}$$

$$\alpha_\mu \alpha_\nu + \alpha_\nu \alpha_\mu = 0 \qquad \text{for} \qquad \mu \neq \nu. \tag{3.50}$$

Indeed, using the anticommutation relations one recovers the Fock–Klein–Gordon equation:

$$\left(\pi_0 + \sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right)\left(\pi_0 - \sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu - \alpha_0 m_0 c\right) =$$

$$= \pi_0^2 - \left[\sum_{\mu=x,y,z}^{3} \alpha_\mu \pi_\mu + \alpha_0 m_0 c\right]^2 = \pi_0^2 - \sum_{\mu,\nu=x,y,z}^{3} \alpha_\mu \alpha_\nu \pi_\mu \pi_\nu -$$

$$- \sum_{\mu=x,y,z}^{3} \left(\alpha_\mu \alpha_0 + \alpha_0 \alpha_\mu\right) \pi_\mu m_0 c - \alpha_0^2 m_0^2 c^2 = \pi_0^2 - \sum_{\mu,\nu=x,y,z}^{3} (\alpha_\mu \alpha_\nu + \alpha_\mu \alpha_\nu) \pi_\mu \pi_\nu -$$

$$- m_0^2 c^2 = \pi_0^2 - \sum_{\mu=x,y,z}^{3} \pi_\mu^2 - m_0^2 c^2.$$

Note that the α's cannot be just numbers, because no numbers can satisfy the anticommutation relation. They have to be matrices. Since we have four variables x, y, z, t, we may expect matrices of order 4, but they could be larger. Here is one of the consistent choices of matrices ($\mathbf{0}$, $\mathbf{1}$ are the zero and unit 2×2 matrices, respectively, while σ_x, σ_y, σ_z are the Pauli matrices, defined on p. 33, that determine electron spin):

$$\alpha_x = \begin{pmatrix} \mathbf{0} & \sigma_x \\ \sigma_x & \mathbf{0} \end{pmatrix}, \qquad \alpha_y = \begin{pmatrix} \mathbf{0} & \sigma_y \\ \sigma_y & \mathbf{0} \end{pmatrix},$$

$$\alpha_z = \begin{pmatrix} \mathbf{0} & \sigma_z \\ \sigma_z & \mathbf{0} \end{pmatrix}, \qquad \alpha_0 \equiv \beta = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix}.$$

The Pauli matrices represent the first sign of what will happen later on: the Dirac equation will automatically describe the spin angular momentum.

3.3.2 The Dirac equations for electron and positron

After the factorization described above Dirac obtained two operator equations. The Dirac equations (for the positron and electron) correspond to these operators acting on the wave function Ψ. Thus, we obtain the equation for the negative electron energies (*positrons*, we use the abbreviation, a kind of the "dot product" of the α matrices and the operator π: $\sum_{\mu=x,y,z} \alpha_\mu \pi_\mu \equiv \boldsymbol{\alpha} \cdot \boldsymbol{\pi}$)

$$(\pi_0 + \boldsymbol{\alpha} \cdot \boldsymbol{\pi} + \alpha_0 m_0 c)\, \Psi = 0 \tag{3.51}$$

and for the positive electron energies (*electrons*)

$$(\pi_0 - \boldsymbol{\alpha} \cdot \boldsymbol{\pi} - \alpha_0 m_0 c)\, \Psi = 0. \tag{3.52}$$

These two equations are coupled together through the same function Ψ which has to satisfy both of them. This coupling caused a lot of trouble in the past. First, people assumed that the equation with the negative electron energies (positron equation) may be ignored, because the energy gap is so large that the Dirac sea is occupied whatever a chemist does with a molecule. This assumption turned out to cause some really vicious or weird performances of numerical procedures (see later on). The electron equation alone reads as

$$i\hbar \frac{\partial \Psi}{\partial t} = (q\phi + c\boldsymbol{\alpha} \cdot \boldsymbol{\pi} + \alpha_0 m_0 c^2)\Psi. \tag{3.53}$$

If one is interested in *stationary states* (cf. p. 25), the wave function has the form $\Psi(x, y, z, t) = \Psi(x, y, z)e^{-i\frac{E}{\hbar}t}$, where we have kept the same symbol for the time-independent factor $\Psi(x, y, z)$. After dividing by $e^{-i\frac{E}{\hbar}t}$ we obtain

THE DIRAC EQUATION FOR STATIONARY ELECTRONIC STATES

$$(E - q\phi - \alpha_0 m_0 c^2 - c\boldsymbol{\alpha} \cdot \boldsymbol{\pi})\Psi = 0. \tag{3.54}$$

The quantity $q\phi = V$ in future applications will denote the Coulomb interaction of the particle under consideration with the external potential.

3.3.3 Spinors and bispinors

The last equation needs a comment. Because the matrices α have dimension 4, Ψ has to be a four-component vector (known as *bispinor*, its connection to the spin concept will be shown later on),

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix},$$

where the first two components (ψ_1 and ψ_2, functions of class Q), for reasons that will become clear in a moment, are called *large components* and are hidden in vector $\boldsymbol{\psi}$, while the two *small components* (ϕ_1 and ϕ_2, functions of class Q)[24] are labeled by vector $\boldsymbol{\phi}$. Vectors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are called the *spinors*.

How do we operate the N-component spinor (for $N = 4$ we have called them bispinors)? Let us construct the proper Hilbert space for the N-component spinors. As usual (p. 595), first, we will define the sum of two spinors in the following way:

$$\begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \cdots \\ \Psi_N \end{pmatrix} + \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \cdots \\ \Phi_N \end{pmatrix} = \begin{pmatrix} \Psi_1 + \Phi_1 \\ \Psi_2 + \Phi_2 \\ \cdots \\ \Psi_N + \Phi_N \end{pmatrix},$$

and then multiply the spinor by a number γ:

$$\gamma \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \cdots \\ \Psi_N \end{pmatrix} = \begin{pmatrix} \gamma\Psi_1 \\ \gamma\Psi_2 \\ \cdots \\ \gamma\Psi_N \end{pmatrix}.$$

Next, we check that the spinors form an Abelian group with respect to the above defined addition (cf. Appendix C, p. 605) and that the conditions for the vector space are fulfilled (Appendix B). Then, we define the scalar product of two spinors,

$$\langle \Phi \mid \Psi \rangle = \sum_{i=1}^{N} \langle \Phi_i \mid \Psi_i \rangle,$$

---

[24] It will be shown that in the nonrelativistic approximation, the large components reduce to the wave function known from the Schrödinger equation, and the small components vanish. The constant E and the function V individually multiply each component of the bispinor Ψ, while $\boldsymbol{\sigma} \cdot \boldsymbol{\pi} \equiv \alpha_x \pi_x + \alpha_y \pi_y + \alpha_z \pi_z$ denotes the "dot product" of the matrices α_μ, $\mu = x, y, z$, by the operators π_μ (in the absence of the electromagnetic field, it is simply the momentum operator component, see p. 673). The matrix β is first multiplied by the constant m_0c^2 and then by the bispinor Ψ.

where the scalar products $\langle \Phi_i | \Psi_i \rangle$ are defined as usual in the Hilbert space of class Q functions. Then, using the scalar product $\langle \Phi | \Psi \rangle$ we define the distance between two spinors, $\|\Phi - \Psi\| \equiv \sqrt{\langle \Phi - \Psi | \Phi - \Psi \rangle}$, and apply the concept of the Cauchy series (the distances between the consecutive terms tend to zero). The Hilbert space of spinors will contain all linear combinations of the spinors together with the limits of all Cauchy series.

An operator acting on a spinor means a spinor with each component resulting from action on the corresponding component,

$$\hat{A} \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \dots \\ \Psi_N \end{pmatrix} = \begin{pmatrix} \hat{A}\Psi_1 \\ \hat{A}\Psi_2 \\ \dots \\ \hat{A}\Psi_N \end{pmatrix}.$$

Sometimes we will use the notation in which a matrix of operators acts on a spinor. In this case the result corresponds to multiplication of the matrix (of operators) and the vector (spinor), i.e.,

$$\begin{pmatrix} \hat{A}_{11} & \hat{A}_{12} & \dots & \hat{A}_{1N} \\ \hat{A}_{21} & \hat{A}_{22} & \dots & \hat{A}_{2N} \\ \dots & \dots & \dots & \dots \\ \hat{A}_{N1} & \hat{A}_{N2} & \dots & \hat{A}_{NN} \end{pmatrix} \begin{pmatrix} \Psi_1 \\ \Psi_2 \\ \dots \\ \Psi_N \end{pmatrix} = \begin{pmatrix} \sum_j \hat{A}_{1j}\Psi_j \\ \sum_j \hat{A}_{2j}\Psi_j \\ \dots \\ \sum_j \hat{A}_{Nj}\Psi_j \end{pmatrix}.$$

3.3.4 What next?

In the following we will show:

a) that the first two components of the bispinor are *much larger* than the last two,
b) that in the limit $c \to \infty$ the Dirac equation gives the *Schrödinger equation*,
c) that the Dirac equation accounts for the spin angular momentum of the electron, and
d) how to obtain, in a simple way, an approximate solution of the Dirac equation to the electron in the field of a nucleus ("*hydrogen-like atom*").

3.3.5 Large and small components of the bispinor

Using matrix multiplication rules, the Dirac equation (3.54) with bispinors can be rewritten in the form of *two equations with spinors* $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$:

$$(E - V - m_0c^2)\boldsymbol{\psi} - c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\boldsymbol{\phi} = \mathbf{0}, \tag{3.55}$$

$$(E - V + m_0c^2)\boldsymbol{\phi} - c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.56}$$

The quantity m_0c^2 represents the energy. Let us use this energy to *shift the energy scale* (we are always free to choose 0 on this scale): $\varepsilon = E - m_0c^2$, in order to make ε comparable in the future to the eigenvalue of the Schrödinger equation (p. 87). We obtain

$$(\epsilon - V)\boldsymbol{\psi} - c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\boldsymbol{\phi} = \mathbf{0}, \tag{3.57}$$

$$(\epsilon - V + 2m_0c^2)\boldsymbol{\phi} - c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.58}$$

This set of equations corresponds to a single matrix equation:

$$\begin{pmatrix} V & c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \\ c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) & V - 2m_0c^2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix} = \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\phi} \end{pmatrix}. \tag{3.59}$$

3.3.6 How to avoid drowning in the Dirac sea

When, in the past, the above equation was solved and the energy ϵ minimized (routine practice in the nonrelativistic case) with respect to the variational parameters in the trial spinors $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ (see Chapter 5), then some serious numerical problems occurred. Either the numerical procedures diverged or the solutions obtained were physically unacceptable. The reason for this was that the existence of the Dirac sea had been totally ignored by neglecting Eq. (3.51) for the positron and taking solely Eq. (3.52) for electron motion. The variational trial functions felt, however, the presence of Dirac sea electronic states (there was nothing in the theory that would prevent the electron from attempting to occupy negative energies) and the corresponding variational energies dived down the energy scale towards the abyss of the sea without bottom.[25] The presence of the Dirac sea makes the Dirac theory, in fact, a theory of an infinite number of particles, whereas formally it is only a theory of a single particle in an external field. This kind of discomfort made people think of the possibility of describing the electron from the Dirac electronic sea by replacing the bispinors by the exact spinor (two components) theory.[26] Such exact separation has been reported by Barysz and Sadlej.[27]

---

[25] How severe the problem might be has been shown by M. Stanke, J. Karwowski, "*Variational Principle in the Dirac Theory: Spurious Solutions, Unexpected Extrema and Other Traps*," in "*New Trends in Quantum Systems in Chemistry and Physics*," vol. I, p. 175–190, eds. J. Maruani et al., Kluwer Academic Publishers. Sometimes an eigenfunction corresponds to a quite different eigenvalue. Nothing of that sort appears in nonrelativistic calculations.

[26] Exact within the Dirac model.

[27] M. Barysz, A.J. Sadlej, J.G. Snijders, *Int. J. Quantum Chem.*, 65(1997)225; M. Barysz, *J. Chem. Phys.*, 114(2001)9315; M. Barysz, A.J. Sadlej, *J. Mol. Struct. (Theochem)*, 573(2001)181; M. Barysz, A.J. Sadlej, *J. Chem. Phys.*, 116(2002)2696. In the latter paper an exact solution to the problem was given. The two-

An approximate (and simple) prescription was also invented to avoid the catastrophic drowning described above. Indeed, Eq. (3.58) can be transformed without any problem to

$$\boldsymbol{\phi} = \left(1 + \frac{(\epsilon - V)}{2m_0c^2}\right)^{-1} \frac{1}{2m_0c}(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\boldsymbol{\psi}.$$

Since $2m_0c^2$ represents a huge energy when compared to the kinetic energy $\varepsilon - V$, the first term on the right-hand side is a good approximation to 1. This means however that

$$\boldsymbol{\phi} \approx \frac{1}{2m_0c}(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\boldsymbol{\psi}, \tag{3.60}$$

which is known as "kinetic balancing." It was shown that the "*kinetically balanced*" trial function achieves the miracle[28] of the energy not tending to $-\infty$. The kinetic balancing indicates some fixed relation between $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$.

Let us focus now on $\boldsymbol{\sigma} \cdot \boldsymbol{\pi}$. This is a 2×2 matrix and in the absence of an electromagnetic field ($\boldsymbol{\pi} = \mathbf{p}$) one has

$$\begin{aligned}\boldsymbol{\sigma} \cdot \boldsymbol{\pi} &= \sigma_x \hat{p}_x + \sigma_y \hat{p}_y + \sigma_z \hat{p}_z \\ &= \begin{pmatrix} 0 & \hat{p}_x \\ \hat{p}_x & 0 \end{pmatrix} + \begin{pmatrix} 0 & -i\hat{p}_y \\ i\hat{p}_y & 0 \end{pmatrix} + \begin{pmatrix} \hat{p}_z & 0 \\ 0 & -\hat{p}_z \end{pmatrix} = \begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix}.\end{aligned}$$

It is seen that $\boldsymbol{\sigma} \cdot \boldsymbol{\pi}$ is of the order of momentum mv, and for the small velocities of the order of m_0v.

Hence, one obtains $\boldsymbol{\phi} \approx \frac{1}{2m_0c}(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\boldsymbol{\psi} \approx \frac{v}{2c}\boldsymbol{\psi}$; therefore the component $\boldsymbol{\phi}$ is for small v much smaller than the component $\boldsymbol{\psi}$,

which justifies the terms "small" and "large" components.[29]

---

component theory, although more appealing, both from the point of view of physics as well as computationally, implies a change in definition of the operators, e.g., the position operator is replaced by a quite complex expression. This fact, ignored in computations using two-component theories, has been analyzed in the following articles: V. Kellő, A.J. Sadlej, B.A. Hess, *J. Chem. Phys.*, 105(1996)1995; M. Barysz, A.J. Sadlej, *Theor. Chem. Acc.*, 97(1997)260; V. Kellő, A.J. Sadlej, *Int. J. Quantum Chem.*, 68(1998)159; V. Kellő, A.J. Sadlej, *J. Mol. Struct. (Theochem)*, 547(2001)35.

[28] This remedy has not only an *ad hoc* character, but moreover does not work for the heaviest atoms, which are otherwise the most important target of relativistic computations.

[29] These terms refer to the positive part of the energy spectrum. For the negative continuum (Dirac sea) the proportion of the components is reversed.

3.3.7 From Dirac to Schrödinger – how to derive the nonrelativistic Hamiltonian?

The *approximate* relation (“kinetic balance”) between the large and small components of the bispinor (that holds for small v/c) may be used to *eliminate* the small components[30] from Eqs. (3.57) and (3.58). We obtain

$$(\epsilon - V)\boldsymbol{\psi} - c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\frac{1}{2m_0c}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \tag{3.61}$$

$$(\epsilon - V)\boldsymbol{\psi} - \frac{1}{2m_0}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \mathbf{0}. \tag{3.62}$$

Let us take a closer look at the meaning of the expression

$$(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) = \begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix}\begin{pmatrix} \hat{p}_z & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & -\hat{p}_z \end{pmatrix} = \begin{pmatrix} \hat{p}^2 & 0 \\ 0 & \hat{p}^2 \end{pmatrix} = \hat{p}^2\mathbf{1}.$$

Now please look carefully. Let us insert this into the last equation. We obtain what is sometimes called the Schrödinger equation with spin (because it is satisfied by a two-component spinor)

$$\left(\frac{\hat{p}^2}{2m_0} + V\right)\boldsymbol{\psi} = \epsilon\boldsymbol{\psi}.$$

Recalling that $\hat{\mathbf{p}}$ represents the momentum operator, we observe each of the large components satisfies the familiar Schrödinger equation

$$\left(-\frac{\hbar^2}{2m_0}\Delta + V\right)\psi = \epsilon\psi.$$

Therefore, the nonrelativistic equation has been obtained from the relativistic one, assuming that the velocity of particle v is negligibly small with respect to the speed of light c. The Dirac equation remains valid even for larger particle velocities.

---

[30] A more elegant solution was reported by Andrzej W. Rutkowski, *J. Phys. B* 19(1986)3431, *ibid.* 19(1986)3431, *ibid.* 19(1986)3443. For the one-electron case, this approach was later popularized by Werner Kutzelnigg as Direct Perturbation Theory (DPT).

3.3.8 How does the spin appear?

It will be shown that the Dirac equation for the free electron in an external electromagnetic field is leading to the spin concept. Thus, in relativistic theory, the spin angular momentum appears in a natural way, whereas in the nonrelativistic formalism it was the subject of a postulate of quantum mechanics (p. 29).

First let us introduce the following identity:

$$(\boldsymbol{\sigma} \cdot \mathbf{a})\,(\boldsymbol{\sigma} \cdot \mathbf{b}) = (\mathbf{a} \cdot \mathbf{b})\,\mathbf{1} + i\boldsymbol{\sigma} \cdot (\mathbf{a} \times \mathbf{b})\,,$$

where we have three times a product of two matrices, each formed by a "scalar product" of matrices $\boldsymbol{\sigma}$ and a vector[31]: $\boldsymbol{\sigma} \cdot \mathbf{a}$, $\boldsymbol{\sigma} \cdot \mathbf{b}$, and $\boldsymbol{\sigma} \cdot (\mathbf{a} \times \mathbf{b})$. Now, taking $\mathbf{a} = \mathbf{b} = \boldsymbol{\pi}$ one obtains the relation

$$(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) = (\boldsymbol{\pi} \cdot \boldsymbol{\pi})\,\mathbf{1} + i\boldsymbol{\sigma}\,(\boldsymbol{\pi} \times \boldsymbol{\pi})\,.$$

If the vector $\boldsymbol{\pi}$ had numbers as its components, the last term would have had to be zero, because the vector product of two parallel vectors would be zero. This, however, need not be true when the vector components are *operators* (as in our case). Since $\boldsymbol{\pi} = \mathbf{p} - \frac{q}{c}\mathbf{A}$, we have $(\boldsymbol{\pi} \cdot \boldsymbol{\pi}) = \boldsymbol{\pi}^2$ and $(\boldsymbol{\pi} \times \boldsymbol{\pi}) = iq\frac{\hbar}{c} curl \mathbf{A}$. To check this, we will obtain the last equality for the x components of both sides (the proof for the other two components looks the same). Let the operator $(\boldsymbol{\pi} \times \boldsymbol{\pi})$ act on an arbitrary function $f(x, y, z)$. As a result we expect the product of f and the vector $iq\frac{\hbar}{c} curl \mathbf{A}$. Let us see:

$$\begin{aligned}(\boldsymbol{\pi} \times \boldsymbol{\pi})_x f &= (\hat{p}_y - q/cA_y)(\hat{p}_z - q/cA_z) f - (\hat{p}_z - q/cA_z)(\hat{p}_y - q/cA_y) f \\ &= \Big[\hat{p}_y\hat{p}_z - q/c\hat{p}_yA_z - q/cA_y\hat{p}_z + (q/c)^2 A_yA_z - \hat{p}_z\hat{p}_y + q/c\hat{p}_zA_y + q/cA_z\hat{p}_y\end{aligned}$$

---

[31] That is, $\boldsymbol{\sigma} \cdot \mathbf{a} = \sigma_x a_x + \sigma_y a_y + \sigma_z a_z$. Let us see, taking an example of $\boldsymbol{\sigma} \cdot \mathbf{a}$, what such a single dot product looks like:

$$\boldsymbol{\sigma} \cdot \mathbf{a} = \begin{pmatrix} 0 & a_x \\ a_x & 0 \end{pmatrix} + \begin{pmatrix} 0 & -ia_y \\ ia_y & 0 \end{pmatrix} + \begin{pmatrix} a_z & 0 \\ 0 & -a_z \end{pmatrix} = \begin{pmatrix} a_z & a_x - ia_y \\ a_x + ia_y & -a_z \end{pmatrix}.$$

Multiplying the matrices $(\boldsymbol{\sigma} \cdot \mathbf{a})$ and $(\boldsymbol{\sigma} \cdot \mathbf{b})$, we therefore get

$$\begin{aligned}(\boldsymbol{\sigma} \cdot \mathbf{a})(\boldsymbol{\sigma} \cdot \mathbf{b}) &= \begin{pmatrix} \mathbf{a} \cdot \mathbf{b} + i(\mathbf{a} \times \mathbf{b})_z & (\mathbf{a} \times \mathbf{b})_y + i(\mathbf{a} \times \mathbf{b})_x \\ -(\mathbf{a} \times \mathbf{b})_y + i(\mathbf{a} \times \mathbf{b})_x & \mathbf{a} \cdot \mathbf{b} - i(\mathbf{a} \times \mathbf{b})_z \end{pmatrix} \\ &= (\mathbf{a} \cdot \mathbf{b})\mathbf{1} + i \begin{pmatrix} (\mathbf{a} \times \mathbf{b})_z & (\mathbf{a} \times \mathbf{b})_x - i(\mathbf{a} \times \mathbf{b})_y \\ (\mathbf{a} \times \mathbf{b})_x + i(\mathbf{a} \times \mathbf{b})_y + & -(\mathbf{a} \times \mathbf{b})_z \end{pmatrix} \\ &= (\mathbf{a} \cdot \mathbf{b})\mathbf{1} + i\boldsymbol{\sigma} \cdot (\mathbf{a} \times \mathbf{b}),\end{aligned}$$

which is the right-hand side of the identity.

$$
\begin{aligned}
&- (q/c)^2 A_z A_y \Big] f \\
&= -q/c(-i\hbar) \left\{ \frac{\partial}{\partial y}(A_z f) - A_z \frac{\partial f}{\partial y} + A_y \frac{\partial f}{\partial z} - \frac{\partial}{\partial z}\left(A_y f\right) \right\} \\
&= i\hbar q/c \left\{ \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right\} f = \frac{iq\hbar}{c} (curl\mathbf{A})_x f.
\end{aligned}
$$

From the Maxwell equations (p. 673), we have $curl\mathbf{A} = \mathbf{H}$, where $\mathbf{H}$ represents the magnetic field intensity. Let us insert this into the Dirac equation (valid for kinetic energy much smaller than $2m_0c^2$, see Eq. (3.61)). We obtain

$$
\begin{aligned}
(\epsilon - V)\boldsymbol{\psi} &= \frac{1}{2m_0}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\psi} = \frac{1}{2m_0}(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\boldsymbol{\psi} + \frac{i}{2m_0}\boldsymbol{\sigma}\cdot(\boldsymbol{\pi}\times\boldsymbol{\pi})\boldsymbol{\psi} \\
&= \frac{1}{2m_0}(\boldsymbol{\pi}\cdot\boldsymbol{\pi})\boldsymbol{\psi} + \frac{i}{2m_0}\frac{iq\hbar}{c}(\boldsymbol{\sigma}\cdot\mathbf{H})\boldsymbol{\psi} \\
&= \left[\frac{\pi^2}{2m_0} - \frac{q\hbar}{2m_0c}\boldsymbol{\sigma}\cdot\mathbf{H}\right]\boldsymbol{\psi} = \left[\frac{\pi^2}{2m_0} + \frac{e\hbar}{2m_0c}\boldsymbol{\sigma}\cdot\mathbf{H}\right]\boldsymbol{\psi}.
\end{aligned}
$$

In the last parentheses, beside the kinetic energy operator (first term), there is a strange second term. The term has the appearance of the interaction energy $-\mathbf{M}\cdot\mathbf{H}$ of a mysterious magnetic dipole moment $\mathbf{M}$ with magnetic field $\mathbf{H}$ (cf. interaction with magnetic field, p. V2-308). The operator of this electronic dipole moment $\mathbf{M} = -\frac{e\hbar}{2m_0c}\boldsymbol{\sigma} = -\mu_B\,\boldsymbol{\sigma}$, where μ_B stands for the Bohr magneton equal to $\frac{e\hbar}{2m_0c}$. The spin angular momentum operator of the electron is denoted by (cf. p. 33) $\mathbf{s}$. Therefore, one has $\mathbf{s} = \frac{1}{2}\hbar\boldsymbol{\sigma}$. Inserting $\boldsymbol{\sigma}$ to the equation for $\mathbf{M}$ we obtain

$$
\mathbf{M} = -2\frac{\mu_B}{\hbar}\mathbf{s} = -\frac{e}{m_0c}\mathbf{s}. \tag{3.63}
$$

It is exactly twice as much as what we get for the *orbital* angular momentum and the corresponding *orbital* magnetic dipole.

When two values differ by an integer factor (as in our case) this should stimulate our mind, because it may mean something fundamental that might depend on, e.g., the number of dimensions of our space or something similar. However, one of the most precise experiments ever performed by humans gave[32] 2.0023193043737 ± 0.0000000000082 instead of 2 (hence the anomalous magnetic spin dipole moment of the electron). Therefore, our excitement must diminish. A more accurate theory (QED, some of the effects of this will be described later) gave

---

[32] R.S. Van Dyck Jr., P.B. Schwinberg, H.G. Dehmelt, *Phys. Rev. Letters*, 59(1990)26.

a result that agreed with the experiment within an experimental error of ±0.0000000008. The extreme accuracy achieved witnessed the exceptional status of QED, because no other theory of mankind has achieved such a level of accuracy.

3.3.9 Simple questions

How to interpret a bispinor wave function? Does the Dirac equation describe a single fermion, an electron, a positron, an electron and a Dirac sea of other electrons (infinite number of particles), or an effective electron or positron (interacting with the Dirac sea)? After 80 years these questions do still not have a clear answer.

Despite the glorious invariance with respect to the Lorentz transformation and despite spectacular successes, the Dirac equation has some serious drawbacks, including a lack of clear physical interpretation. These drawbacks are removed by a more advanced theory – QED.

3.4 The hydrogen-like atom in Dirac theory

After this short escapade we are back with Dirac theory. The hydrogen-like atom may be simplified by immobilizing the nucleus and considering a single particle – the electron[33] moving in the *electrostatic field of the nucleus*,[34] $-Z/r$. This problem has an exact solution first obtained by Charles Galton Darwin (cf. p. 147). The electron state is described by four quantum numbers n, l, m, m_s, where $n = 1, 2, \ldots$ stands for the principal, $0 \leq l \leq n-1$ for the angular, $|m| \leq l$ for the magnetic (both are integers), and $m_s = \frac{1}{2}, -\frac{1}{2}$ for the spin quantum number. Darwin obtained the following formula for the relativistic energy of the hydrogen-like atom (in a.u.):

$$E_{n,j} = -\frac{1}{2n^2}\left[1 + \frac{1}{nc^2}\left(\frac{1}{j+\frac{1}{2}} - \frac{3}{4n}\right)\right],$$

where $j = l + m_s$ and c is the speed of light (in a.u.). For the ground state ($1s$, $n = 1$, $l = 0$, $m = 0$, $m_s = \frac{1}{2}$) we have

$$E_{1,\frac{1}{2}} = -\frac{1}{2}\left[1 + \left(\frac{1}{2c}\right)^2\right].$$

---

[33] In the Dirac equation $\mathbf{A} = \mathbf{0}$ and $-e\phi = V = -\frac{Ze^2}{r}$ were set.

[34] The center-of-mass motion can be easily separated from the Schrödinger equation. Nothing like this has been done for the Dirac equation. The atomic mass depends on its velocity with respect to the laboratory coordinate system, the electron and proton mass also depend on their speeds, and there is also a mass deficit as a result of binding between both particles. All this seems to indicate that center-of-mass separation is not possible. Nevertheless, for an energy expression accurate to a certain power of c^{-1}, such a separation is, at least in some cases, possible.

Thus, instead of the nonrelativistic energy equal to $-\frac{1}{2}$, from the Dirac theory we obtain -0.5000067 a.u., which means a very small stabilizing correction to the nonrelativistic energy. The electron energy levels for the nonrelativistic and relativistic cases are shown schematically in Fig. 3.3.

3.4.1 Step by step: calculation of the hydrogen-like atom ground state within Dirac theory

Matrix form of the Dirac equation

We will use the Dirac equation (3.59). First, the basis set composed of two bispinors will be created, $\boldsymbol{\Psi}_1 = \begin{pmatrix} \boldsymbol{\psi} \\ \mathbf{0} \end{pmatrix}$ and $\boldsymbol{\Psi}_2 = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\phi} \end{pmatrix}$, and the wave function $\boldsymbol{\Psi}$ will be sought as a linear combination $\boldsymbol{\Psi} = c_1\ \boldsymbol{\Psi}_1 + c_2\ \boldsymbol{\Psi}_2$, which represents an approximation. Within this approximation the Dirac equation looks as follows:

$$\begin{pmatrix} V-\epsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\epsilon \end{pmatrix}(c_1\boldsymbol{\Psi}_1 + c_2\boldsymbol{\Psi}_2) = \mathbf{0},$$

which gives

$$c_1\begin{pmatrix} V-\epsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\epsilon \end{pmatrix}\boldsymbol{\Psi}_1 + c_2\begin{pmatrix} V-\epsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\epsilon \end{pmatrix}\boldsymbol{\Psi}_2 = \mathbf{0}.$$

By making a scalar product first with $\boldsymbol{\Psi}_1$ and then with $\boldsymbol{\Psi}_2$ we obtain two equations:

$$c_1\left\langle \boldsymbol{\Psi}_1 \middle| \begin{pmatrix} V-\epsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\epsilon \end{pmatrix} \boldsymbol{\Psi}_1 \right\rangle + c_2\left\langle \boldsymbol{\Psi}_1 \middle| \begin{pmatrix} V-\epsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\epsilon \end{pmatrix} \boldsymbol{\Psi}_2 \right\rangle = 0,$$

$$c_1\left\langle \boldsymbol{\Psi}_2 \middle| \begin{pmatrix} V-\epsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\epsilon \end{pmatrix} \boldsymbol{\Psi}_1 \right\rangle + c_2\left\langle \boldsymbol{\Psi}_2 \middle| \begin{pmatrix} V-\epsilon & c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & V-2m_0c^2-\epsilon \end{pmatrix} \boldsymbol{\Psi}_2 \right\rangle = 0.$$

Taking into account the particular structure of the bispinors $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$, we obtain the same equations expressed in (two component) spinors:

$$c_1 \langle \boldsymbol{\psi}\ |(V-\varepsilon)\boldsymbol{\psi}\rangle + c_2 \langle \boldsymbol{\psi}\ |c(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\boldsymbol{\phi}\rangle = 0,$$

$$c_1 \langle \boldsymbol{\phi} \, | c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \boldsymbol{\psi} \rangle + c_2 \langle \boldsymbol{\phi} \left| (V - 2m_0c^2 - \varepsilon) \boldsymbol{\phi} \right\rangle = 0.$$

This is a set of homogeneous linear equations. To obtain a nontrivial solution,[35] the determinant of the coefficients multiplying the unknowns c_1 and c_2 has to be zero (the secular determinant, cf. variational method in Chapter 5), i.e.,

$$\begin{vmatrix} \langle \boldsymbol{\psi} \, | (V - \varepsilon) \; \boldsymbol{\psi} \rangle & \langle \boldsymbol{\psi} \, | c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \boldsymbol{\phi} \rangle \\ \langle \boldsymbol{\phi} \, | c(\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \boldsymbol{\psi} \rangle & \langle \boldsymbol{\phi} \left| (V - 2m_0c^2 - \varepsilon) \boldsymbol{\phi} \right\rangle \end{vmatrix} = 0. \tag{3.64}$$

The potential V in the above formula will be taken as $-Z/r$, where r is the electron–nucleus distance.

The large component spinor

It is true that we have used an extremely poor basis; however, we will try to compensate for it by allowing a certain flexibility within the large component spinor, $\boldsymbol{\psi} = \begin{pmatrix} 1s \\ 0 \end{pmatrix}$, where the hydrogen-like function $1s = \sqrt{\frac{\zeta^3}{\pi}} exp(-\zeta r)$. The parameter ζ will be optimized in such a way as to minimize the energy ε of the electron. This idea is similar to the variational method in the nonrelativistic theory (Chapter 5 and Appendix H, p. 683), however, it is hardly justified in the relativistic case. Indeed, as proved by numerical experience the variational procedure often fails. As a remedy we will use the kinetic balancing already described of the large and small components of the bispinor (p. 155). The spinor of the small components is therefore obtained automatically from the large components (approximation):

$$\begin{aligned} \boldsymbol{\phi} &= N_\phi (\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \begin{pmatrix} 1s \\ 0 \end{pmatrix} = N_\phi \begin{pmatrix} \hat{p}_z & \hat{p}_x + i\hat{p}_y \\ \hat{p}_x - i\hat{p}_y & \hat{p}_z \end{pmatrix} \begin{pmatrix} 1s \\ 0 \end{pmatrix} \\ &= N_\phi \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\hat{p}_y)(1s) \end{pmatrix}, \end{aligned}$$

where N_ϕ is a normalization constant. In the above formula $\hat{\mathbf{p}}$ represents the momentum operator. The normalization constant N_ϕ will be found from

$$\begin{aligned} \langle \boldsymbol{\phi} \, | \boldsymbol{\phi} \rangle = 1 = \left| N_\phi \right|^2 \left\{ \left\langle \hat{p}_z(1s) \, \middle| \hat{p}_z(1s) \right\rangle + \left\langle (\hat{p}_x + i\hat{p}_y)(1s) \, \middle| (\hat{p}_x + i\hat{p}_y)(1s) \right\rangle \right\} \\ = \left| N_\phi \right|^2 \cdot \left\{ \begin{array}{c} \left\langle \hat{p}_z(1s) \, \middle| \hat{p}_z(1s) \right\rangle + \left\langle \hat{p}_x(1s) \, \middle| \hat{p}_x(1s) \right\rangle + i \left\langle \hat{p}_x(1s) \, \middle| \hat{p}_y(1s) \right\rangle \\ -i \left\langle \hat{p}_y(1s) \, \middle| \hat{p}_x(1s) \right\rangle + \left\langle \hat{p}_y(1s) \, \middle| \hat{p}_y(1s) \right\rangle \end{array} \right\}. \end{aligned}$$

---

[35] It is easy to give a trivial one, but not acceptable (the wave function cannot equal zero everywhere): $c_1 = c_2 = 0$.

In the above formula integrals with the imaginary unit i are equal to zero, because the integrand is an odd function. After using the Hermitian character of the momentum operator we obtain $1 = \left|N_\phi\right|^2 \langle 1s \left|\hat{p}^2 1s\right\rangle = \left|N_\phi\right|^2 \zeta^2$. The last equality follows from Appendix H, p. 683. Thus, one may choose $N_\phi = 1/\zeta$.

Calculating integrals in the Dirac matrix equation

We will calculate one by one all the integrals that appear in the Dirac matrix equation. The integral $\langle \boldsymbol{\psi} \left| -\frac{Z}{r} \boldsymbol{\psi} \right\rangle = -Z\zeta$, because the scalar product leads to the nuclear attraction integral with a hydrogen-like atomic orbital, and this gives the result above (Appendix H, p. 683). The next integral can be computed as follows:

$$\begin{aligned}
\left\langle \boldsymbol{\phi} \left| \tfrac{1}{r} \boldsymbol{\phi} \right.\right\rangle &= \left|N_\phi\right|^2 \left\langle \begin{array}{c} \hat{p}_z(1s) \\ \left(\hat{p}_x + i\hat{p}_y\right)(1s) \end{array} \right| \tfrac{1}{r} \left(\begin{array}{c} \hat{p}_z(1s) \\ \left(\hat{p}_x + i\hat{p}_y\right)(1s) \end{array} \right) \Bigg\rangle \\
&= \left|N_\phi\right|^2 \left\langle \hat{p}_z(1s) \left| \tfrac{1}{r} \hat{p}_z(1s) \right.\right\rangle + \left\langle \left(\hat{p}_x + i\hat{p}_y\right)(1s) \left| \tfrac{1}{r} \left(\hat{p}_x + i\hat{p}_y\right)(1s) \right.\right\rangle \\
&= \left|N_\phi\right|^2 \left\langle (1s) \left| \hat{p}_z \tfrac{1}{r} \hat{p}_z(1s) \right.\right\rangle + \left\langle (1s) \left| \left(\hat{p}_x - i\hat{p}_y\right) \tfrac{1}{r} \left(\hat{p}_x + i\hat{p}_y\right)(1s) \right.\right\rangle \\
&= \left|N_\phi\right|^2 \left\langle (1s) \left| \left(\hat{p}_z \tfrac{1}{r} \right) \hat{p}_z(1s) \right.\right\rangle + \left\langle (1s) \left| \left[\left(\hat{p}_x - i\hat{p}_y\right) \tfrac{1}{r} \right] \left(\hat{p}_x + i\hat{p}_y\right)(1s) \right.\right\rangle \\
&+ \left\langle (1s) \left| \tfrac{1}{r} \hat{p}_z \hat{p}_z(1s) \right.\right\rangle + \left\langle (1s) \left| \tfrac{1}{r} \left(\hat{p}_x - i\hat{p}_y\right) \left(\hat{p}_x + i\hat{p}_y\right)(1s) \right.\right\rangle .
\end{aligned} \tag{3.65}$$

In the second row, the scalar product of spinors is used, in the third row, the Hermitian character of the operator $\hat{p}$. Further,

$$\begin{aligned}
\left\langle \boldsymbol{\phi} \left| \tfrac{1}{r} \boldsymbol{\phi} \right.\right\rangle &= \left|N_\phi\right|^2 \left[\left\langle (1s) \left| \left(\hat{p}_z \tfrac{1}{r} \right) \hat{p}_z(1s) \right.\right\rangle + \left\langle (1s) \left| \tfrac{1}{r} \left(\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2 \right)(1s) \right.\right\rangle \right. \\
&+ \left. \left\langle (1s) \left| \left[\left(\hat{p}_x - i\hat{p}_y\right) \tfrac{1}{r} \right] \left(\hat{p}_x + i\hat{p}_y\right)(1s) \right.\right\rangle \right] \\
&= \left|N_\phi\right|^2 \left[\left\langle (1s) \left| \left(\hat{p}_z \tfrac{1}{r} \right) \hat{p}_z(1s) \right.\right\rangle - \left\langle (1s) \left| \tfrac{1}{r} \Delta(1s) \right.\right\rangle + \left\langle (1s) \left| \left(\hat{p}_x \tfrac{1}{r} \right) \hat{p}_x(1s) \right.\right\rangle \right. \\
&+ \left. \left\langle (1s) \left| \left(\hat{p}_y \tfrac{1}{r} \right) \hat{p}_y(1s) \right.\right\rangle - i \left\langle (1s) \left| \left(\hat{p}_y \tfrac{1}{r} \right) \hat{p}_x(1s) \right.\right\rangle + i \left\langle (1s) \left| \left(\hat{p}_x \tfrac{1}{r} \right) \hat{p}_y(1s) \right.\right\rangle \right] .
\end{aligned} \tag{3.66}$$

We used the atomic units and therefore $\hat{p}^2 = -\Delta$, and the momentum operator is equal to $-i\nabla$. The two integrals at the end cancel each other out, because each of the integrals does not change when the variables are interchanged: $x \leftrightarrow y$.

Finally, we obtain the following formula: $\langle \boldsymbol{\phi} \left| \frac{1}{r}\boldsymbol{\phi} \right\rangle = -N_\phi^2 \left\{ \langle 1s \left| \frac{1}{r}\Delta(1s) \right\rangle + \langle 1s \left| \left(\nabla \frac{1}{r}\right) \nabla(1s) \right\rangle \right\} = -\zeta^{-2}\left\{\left(-3\zeta^3 + 2\zeta^3\right)\right\} = \zeta$, where the equality follows from a direct calculation of the two integrals.[36]

The next matrix element to calculate is equal to $\langle \boldsymbol{\phi} \, | c\,(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\, \boldsymbol{\psi} \rangle$. We proceed as follows (please recall kinetic balancing and we also use Appendix H, p. 683):

$$\begin{aligned}
\langle \boldsymbol{\phi} \, | c\,(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\, \boldsymbol{\psi} \rangle &= N_\phi c \left\langle (\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \begin{pmatrix} 1s \\ 0 \end{pmatrix} \middle| (\boldsymbol{\sigma} \cdot \boldsymbol{\pi}) \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle \\
&= N_\phi c \left\langle \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\,\hat{p}_y)(1s) \end{pmatrix} \middle| \begin{pmatrix} \hat{p}_z(1s) \\ (\hat{p}_x + i\,\hat{p}_y)(1s) \end{pmatrix} \right\rangle \\
&= N_\phi c \left[\left\langle \hat{p}_z(1s) \, \middle| \hat{p}_z(1s) \right\rangle + \left\langle (\hat{p}_x + i\,\hat{p}_y)(1s) \, \middle| (\hat{p}_x + i\,\hat{p}_y)(1s) \right\rangle \right] \\
&= N_\phi c \, \langle 1s \left| \hat{p}^2(1s) \right\rangle = \frac{1}{\zeta} c \zeta^2 = c\zeta .
\end{aligned}$$

The last matrix element reads as

$$\begin{aligned}
\langle \boldsymbol{\psi} \, | c\,(\boldsymbol{\sigma} \cdot \boldsymbol{\pi})\, \boldsymbol{\phi} \rangle &= N_\phi c \left\langle \begin{pmatrix} 1s \\ 0 \end{pmatrix} \middle| (\boldsymbol{\sigma} \cdot \boldsymbol{\pi})^2 \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle \\
&= N_\phi c \left\langle \begin{pmatrix} 1s \\ 0 \end{pmatrix} \middle| \begin{pmatrix} \hat{p}^2 & 0 \\ 0 & \hat{p}^2 \end{pmatrix} \begin{pmatrix} 1s \\ 0 \end{pmatrix} \right\rangle = N_\phi c \, \langle 1s \left| \hat{p}^2 1s \right\rangle = c \frac{1}{\zeta} \zeta^2 = c\zeta .
\end{aligned}$$

Dirac's secular determinant

We have all the integrals needed and may now write the secular determinant corresponding to the matrix form (3.64) of the Dirac equation. After inserting the calculated integrals we get

$$\begin{vmatrix} -Z\zeta - \varepsilon & c\zeta \\ c\zeta & -Z\zeta - 2c^2 - \varepsilon \end{vmatrix} = 0.$$

Expanding the determinant gives the equation for the energy ϵ

$$\varepsilon^2 + \varepsilon(2Z\zeta + 2c^2) + [Z\zeta(Z\zeta + 2c^2) - c^2\zeta^2] = 0.$$

---

[36] In the first integral we have the same situation as a while before. In the second integral we write the nabla operator in Cartesian coordinates, obtain a scalar product of two gradients, and then we get three integrals equal to one another (they contain x, y, z), and it is sufficient to calculate one of them by spherical coordinates by Eq. (H.2) in Appendix H, p. 683.

Hence, we get two solutions:

$$\varepsilon_{\pm} = -(c^2 + Z\zeta) \pm \sqrt{(c^4 + \zeta^2 c^2)}.$$

Note that the square root is of the order of c^2 (in a.u.), and with the (unit) mass of the electron m_0, it is of the order of m_0c^2. Therefore, the minus sign before the square root corresponds to a solution with energy of the order of $-2m_0c^2$, while the plus sign corresponds to energy of the order of zero. Let us recall that we have shifted the energy scale in the Dirac equation and this is the last solution ε_+ (hereafter denoted by ε), which is to be compared to the energy of the nonrelativistic hydrogen-like atom

$$\begin{aligned}\varepsilon &= -\left(c^2 + Z\zeta\right) + \sqrt{c^4 + \zeta^2 c^2} = -\left(c^2 + Z\zeta\right) + c^2\sqrt{1 + \frac{\zeta^2}{c^2}} \\ &= -\left(c^2 + Z\zeta\right) + c^2\left(1 + \frac{\zeta^2}{2c^2} - \frac{\zeta^4}{8c^4} + ...\right) = -Z\zeta + \frac{\zeta^2}{2} + \left(-\frac{\zeta^4}{8c^2} + ...\right).\end{aligned} \tag{3.67}$$

Nonrelativistic solution

If $c \to \infty$, i.e., we approach the nonrelativistic limit, then $\varepsilon = -Z\zeta + \frac{\zeta^2}{2}$. Minimization of this energy with respect to ζ gives its optimum value $\zeta_{opt}^{nonrel} = Z$. In this way one recovers the result known from nonrelativistic quantum mechanics (Appendix H) obtained in the variational approach to the hydrogen atom with the $1s$ orbital as a trial function.

Relativistic contraction of orbitals

Minimizing the relativistic energy given by Eq. (3.67) leads to an equation for optimum $\zeta \equiv \zeta_{opt}^{rel}$:

$$\frac{d\epsilon}{d\zeta} = 0 = -Z + \frac{1}{2}\left(c^4 + \zeta^2 c^2\right)^{-\frac{1}{2}} 2\zeta c^2 = -Z + \left(c^4 + \zeta^2 c^2\right)^{-\frac{1}{2}} \zeta c^2,$$

giving

$$\zeta_{opt}^{rel} = \frac{Z}{\sqrt{1 - \frac{Z}{c^2}}}.$$

The result differs remarkably from the nonrelativistic value $\zeta_{opt}^{nonrel} = Z$, but approaches the nonrelativistic value when $c \to \infty$. Note that the difference between the two values increases with atomic number Z and that the relativistic exponent is always *larger* than its nonrelativistic counterpart. This means that the relativistic orbital *decays faster* with the electron–nucleus distance and therefore

> the relativistic orbital 1*s* is *smaller* (contraction) than the corresponding nonrelativistic one.

Let us see how it is for the hydrogen atom. In that case $\zeta_{opt}^{rel} = 1.0000266$ as compared to $\zeta_{opt}^{nonrel} = Z_H = 1$. And what about the 1*s* orbital of gold? For gold $\zeta_{opt}^{rel} = 96.68$, while $\zeta_{opt}^{nonrel} = Z_{Au} = 79$. This is a large relativistic contraction of the atomic orbitals. Since for a heavy atom, the effective exponent of the atomic orbitals decreases when moving from the low-energy compact 1*s* orbital to higher-energy outer orbitals, this means that the most important relativistic orbital contraction occurs for the inner shells. The chemical properties of an atom depend on what happens to its outer shells (valence shell). Therefore, we may conclude that the relativistic corrections are expected to play a secondary role in chemistry.[37]

If we insert ζ_{opt}^{rel} in Eq. (3.67) we obtain the minimum value of ε, i.e.,

$$\varepsilon_{min} = -\left(c^2 + Z\zeta\right) + \sqrt{c^4 + \zeta^2}. \tag{3.68}$$

Since Z^2/c^2 is small with respect to 1, we may expand the square root in the Taylor series, $\sqrt{1-x} = 1 - \frac{1}{2}x - \frac{1}{8}x^2 - \ldots$. We obtain

$$\begin{aligned}\varepsilon_{min} &= -c^2 + c^2\left\{1 - \left(\frac{1}{2}\right)\left(\frac{Z^2}{c^2}\right) - \frac{1}{8}\left(\frac{Z^2}{c^2}\right)^2 - \ldots\right\} \\ &= -\frac{Z^2}{2}\left(1 + \left(\frac{Z}{2c}\right)^2 + \ldots\right).\end{aligned} \tag{3.69}$$

In the case of the hydrogen atom ($Z = 1$) we have

$$\varepsilon_{min} = -\frac{1}{2}\left(1 + \left(\frac{1}{2c}\right)^2 + \ldots\right), \tag{3.70}$$

---

[37] We have to remember, however, that the relativistic effects also propagate from the inner shells to the valence shell through the orthogonalization condition, that has to be fulfilled after the relativistic contraction. This is why the gold valence orbital 6*s* shrinks, which has an immediate consequence in the relativistic shortening of the bond length in Au_2, which we cited at the beginning of this chapter.

where the first two terms shown give Darwin's exact result[38] (discussed earlier). Inserting $c = 137.036$ a.u. we obtain the hydrogen atom ground-state energy $\varepsilon = -0.5000067$ a.u., which agrees with Darwin's result.

3.5 Larger systems

The Dirac equation is rigorously invariant with respect to the Lorentz transformation, which is certainly the most important requirement for a relativistic theory. Therefore, it seems to be a logically sound approximation for relativistic description of a single quantum particle. Unfortunately, this is not true. Please recall that the Dirac Hamiltonian spectrum contains a continuum of states with negative energies, which is not bound from below (a kind of "energy abyss" with threatening consequences). Dirac desperately postulated that all these negative continuum energy levels are inaccessible for an electron occupying a "normal" energy level with positive energy, like, e.g., the levels of the hydrogen atom.[39] This inaccessibility was supposed to result from postulated occupation of these levels by the electrons from the "Dirac sea" and operation of the Pauli exclusion principle.

The first contradiction within the Dirac theory is the following: the Dirac model ceases to be of one-particle type (by occupying the negative energy continuum by an infinite number of electrons), but despite of this it is unable to describe even two electrons when they are in the positive energy states. A second contradiction is visible from the Fock–Klein–Gordon equation. There also one has a negative energy continuum (one may have also the negative solution when making the square root of the square of Eq. (3.39), p. 145), but since the Fock–Klein–Gordon equation describes bosons,[40] for which the Pauli exclusion principle does not hold (Chapter 1), the concept of Dirac about inaccessibility is useless in this case...[41]

The internal inconsistencies of the Dirac model lead to some absurd results. For example, the Dirac electron velocity, defined as the eigenvalue of the operator being the time derivative of the position operator, is everywhere equal to the velocity of light. These and other inconsistencies are removed only after going to a more advanced theory, which is QED. Unfortunately, QED

---

[38] That is, the exact solution to the Dirac equation for the electron in the external electric field produced by the proton.

[39] Such transitions would cause not only inexistence of atoms, but also a catastrophe on a scale of the Universe, since the infinite amount of energy has to be emitted!

[40] An analysis of the transformation properties of the Fock–Klein–Gordon equation and of the Dirac equation leads to the conclusion that satisfaction of the first of these equations requires a usual (i.e., scalar) wave function, whereas the second equations requires a bispinor character of the wave function. Scalar functions describe spinless particles (because they cannot be associated with the Pauli matrices), while bispinors in the Dirac equation are associated with the Pauli matrices, and describe a particle of spin 1/2.

[41] This means that the Dirac sea is nothing more than a breathtaking fairy tale. Sometimes fairy tales are useful and calming (in a short perspective), for children as well as for scientists.

represents a complicated machinery, which nowadays cannot be applied to large molecules of chemical interest. Therefore, in practice we are forced to use the Dirac equation and to be effective for larger systems we should somehow invent its many-particle generalizations. During nearly a century after the Dirac theory was born, the dangers and traps connected to the Dirac equation have been quite well recognized. Currently the Dirac theory is able to produce results with excellent agreement with experimental data. This however is often associated with a special care during computations.

Noninteracting Dirac electrons

Just to see what kind of problems are associated with application of the Dirac equation let us consider first two noninteracting Dirac electrons. Let us assume that a two-electron Hamiltonian is simply a sum of the two one-electron Hamiltonians. Such two-electron problems can be separated into two one-electron ones. If we have to do with such a situation with the Schrödinger equation nothing special happens – we construct the solution for the two-electron problem from the one-electron solutions. If we solve such a problem in the Dirac model, the negative continuum becomes a source of serious difficulties. Note that the sum of energies of two electrons, one with the positive and one with the negative energy, may give us any energy (Fig. 3.4). We see also that the energy eigenvalues of the two-electron Hamiltonian that correspond to bound states are immersed in the continuum extending from $-\infty$ to $+\infty$. This is strange, but for the time being does not cause any numerical catastrophe, provided we *first separate* the problem into the two one-electron ones (for the noninteracting electrons this is possible) and *then solve* the two problems separately. If we follow this method we get the correct values of the discrete states.

Dirac–Coulomb (DC) model

Things become really very complex when we try to switch the electronic interaction on. First what to use as the interaction energy operator? Nobody knows, but a first idea might be to add to the sum of the Dirac one-electron operators the Coulombic interaction operators of all the particles. This is what is known as the Dirac–Coulomb (DC) model.[42]

The DC approximation is rightly seen by many scholars as methodologically inconsistent. Indeed, as we will show in a minute the Coulomb law needs the infinite velocity of light (Fig. 3.5). This means its nonrelativistic character. Such a nonrelativistic ingredient should not be mixed

---

[42] Some researchers take into account the relativistic effects as a kind of perturbation of the Schrödinger equation. They solve the latter with the corresponding potential energy modified in such a way as to mimic the relativistic effects.

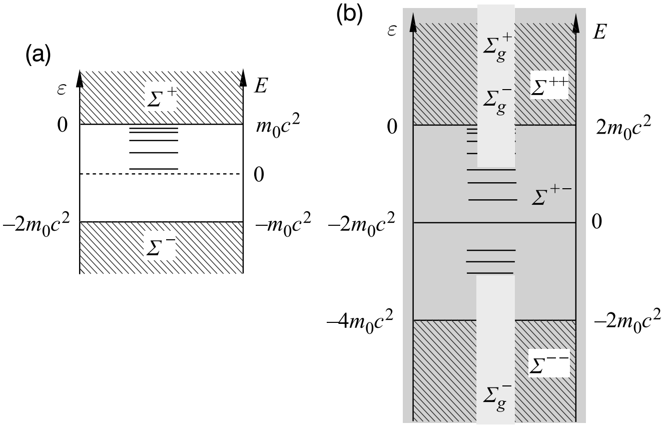


Fig. 3.4. The spectrum of the Dirac Hamiltonian for (a) one electron and (b) two noninteracting electrons. On the right-hand side of each figure the original scale of energy (E) is used, while on the left-hand side the shifted scale (ε) is applied. In the one-electron case one obtains two continua: the first one, Σ^+, describes the states with positive energies, and the second one, Σ^-, pertains to the states with negative energies. In case of two electrons the continua Σ^{++} and Σ^{--} appear, when both electrons belong to the same continuum Σ^+ or Σ^-. However, the total scale of energy is covered by continuum Σ^{+-} that corresponds to one electron in Σ^+ and the other in Σ^-. Continua Σ_g^+ and Σ_g^- appear when one of the electrons occupies a discrete level, while the second one is in a state belonging to a continuum (Σ^+ or Σ^-).

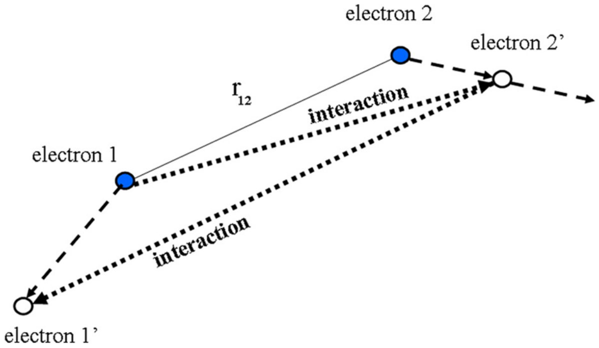


Fig. 3.5. Retardation of the interaction. The distance r_{12} of two particles in the interaction potential (as in Coulomb's law) is bound to represent an approximation, because we assume an instantaneous interaction. However, when the two particles catch sight of each other (which takes time) they are already somewhere else. In Richard Feynman's wording: "The Sun atom shakes, my eye electron shakes eight minutes later."

with any relativistic theory, like, e.g., the Dirac model, because we may get a weird theory which could lead us into scientific swamps with traps and dancing elves, which in reality does not exist. Moreover, Brown and Ravenhall[43] already in the 1950s remarked that due to the interelectronic interaction some states that are of mixed character result: the one of continuum states and the other of bound states of noninteracting particles. This is called in the literature (a bit paramedically) as the Brown–Ravenhall sickness. This means that in the DC model the ground state and all excited states do not represent any bound states, but are in fact resonance states. Resonance states are unstable by definition, i.e., have a finite lifetime and describe inevitable (although after maybe a long life) spontaneous disintegration of the system into pieces.[44] This is obviously in contradiction with experimental facts, because, e.g., the helium atom exists and nothing indicates its internal instability.

Therefore, we have a serious problem, because the modern relativistic computations for atoms and molecules are almost exclusively based on the DC approximation.[45] The problem can be at least partially removed if one applies some methods based on the one-electron model, like for example a mean-field model.[46] A mean field means that we study the motion of a single particle in the field of other particles, but their motion is averaged (their motion and, in a sense, the other particles themselves disappear from the theory through the averaging). In such a model the wave function for all electrons is built from some one-electron functions, which, in this case, represent four-component bispinors. One obtains these bispinors by solving the corresponding Dirac equation in a kinetically balanced bispinor basis (Eq. (3.60), p. 155). The kinetic balance protects the model against the variational collapse, but not against Brown–Ravenhall sickness. One avoids the symptoms of this sickness by projecting the DC equation onto the space of those states that correspond to positive energy.[47] Such a procedure ensures stability of solutions and allows for using the well-known methods developed for bound states. We pay however a price: the variational space used is not complete. Moreover, often construction of a basis set that does not contain admixtures from Σ^- is impossible.[48]

It has been realized only recently that to avoid the abovementioned traps as well as to see what the solutions of the DC equation look like, one has to apply the methods specific to resonance

---

[43] G.E. Brown, D.G. Ravenhall, *Proc. Royal Soc. A*, 208(1951)552.

[44] The resonance states will be discussed in more detail in Chapter 6.

[45] For small systems, like hydrogen-, helium-, and lithium-like atoms, there is a possibility to use more accurate methods of QED.

[46] See Chapter 8.

[47] In practice this looks as follows. The many-electron wave function (let us focus our attention on a two-electron system only) is constructed from those bispinors which correspond to positive energy solutions of the Dirac equation. For example, among two-electron functions built of such bispinors, no function corresponds to Σ^{--}, Σ_g^-, and, most importantly, Σ^{+-}. This means that carrying out computations with such a basis set, we do not use the full DC Hamiltonian, but instead its projection on the space of states with positive energies.

[48] For example, when we are beyond the mean-field model.

states,[49] because they are such states and cannot be legitimately treated otherwise. The energies obtained as solutions of the full DC equation (a correct approach) differ from the solutions of the equation projected onto the space of states with positive energies (approximation) by the terms proportional to $(Z/c)^3$, where Z means the nuclear charge.[50] It turned out that the lifetimes of the resonance states, which correspond to the truly bound states, are proportional to $(Z/c)^{-3}$, and therefore are relatively long (especially for light atoms).[51] It is worth noting that the DC equation, also together with the Breit corrections described in the next section, is exact up to the terms proportional to $(Z/c)^2$. Therefore, the effects connected to the instability of the DC solution, although they were in the past and still are very annoying at the numerical as well as interpretation levels, are smaller than the limits of validity of the DC model itself.[52]

3.6 Beyond the Dirac equation...

How reliable is the relativistic quantum theory presented? The Dirac or Fock–Klein–Gordon equations, as usual in physics, describe only some aspects of reality. The fact that both equations are invariant with respect to the Lorentz transformation indicates *only* that the space–time *symmetry* properties are described correctly. The physical machinery represented by these equations is not so bad, since several predictions have been successfully made (antimatter, electron spin, energy levels of the hydrogen atom). Yet, in the latter case an assumption of the external field $V = -\frac{Ze^2}{r}$ is a positively desperate step, which in fact is unacceptable in a fair relativistic theory for the proton and the electron (and not only of the electron in the external field of the nucleus). Indeed, the proton and the electron move. At a given time their distance is equal to r, but such a distance might be inserted into the Coulombic law if the speed of light were infinite, because the two particles would feel their positions instantaneously. Since, however, any perturbation by a position change of a particle needs time to travel to the other particle, we have to use another distance somehow taking this into account (Fig. 3.5). The same pertains, of course, to any pair of particles in a many-body system (the so-called *retarded potential*).

There is certainly a need for a more accurate theory.

---

[49] G. Pestka, M. Bylicki, J. Karwowski, *J. Phys. B: At. Mol. Opt. Phys.*, 39(2006)2979; *ibid.* 40(2007)2249.

[50] Here we consider an atom.

[51] In principle one should write $Ze^2/\hbar c$ instead of Z/c. The dimensionless constant $e^2/\hbar c \approx 1/137$ is known as *fine structure constant*, in atomic units ($e = 1$, $\hbar = 1$, see Chapter 4) it is equal to $\frac{1}{c}$.

[52] The corrections proportional to $(Z/c)^n$ for $n \geq 3$ that come from QED are nowadays being treated perturbationally (Chapter 5), also for many-electron systems, like, e.g., all atoms. These corrections may become important in interpretation of some particular phenomena, like X-ray spectra.

3.6.1 The Breit equation

Breit constructed a many-electron relativistic theory that takes into account such a retarded potential in an approximate way.[53] Breit explicitly considered only the electrons of an atom, its nucleus (similar to Dirac theory) created only an external field for the electrons. This ambitious project was only partly successful, because the resulting theory turned out to be approximate not only from the point of view of quantum theory (some interactions not taken into account) but also from the point of view of relativity theory (an *approximate* Lorentz transformation invariance).

Gregory Breit (1899–1981), American physicist, professor at the New York, Wisconsin, Yale, Buffalo universities. Breit and Eugene Wigner introduced the resonance states of particles, and with Condon they created the proton–proton scattering theory.

For two electrons the Breit equation has the form (r_{12} stands for the distance between electron 1 and electron 2)

$$\left\{\hat{H}(1)+\hat{H}(2)+\frac{1}{r_{12}}-\frac{1}{2r_{12}}\left[\boldsymbol{\alpha}(1)\boldsymbol{\alpha}(2)+\frac{[\boldsymbol{\alpha}(1)\cdot\mathbf{r}_{12}][\boldsymbol{\alpha}(2)\cdot\mathbf{r}_{12}]}{r_{12}^2}\right]\right\}\Psi=E\Psi, \qquad (3.71)$$

where (cf. Eq. (3.54) with E replaced by the Hamiltonian)

$$\hat{H}(i)=q_i\phi(\mathbf{r}_i)+c\boldsymbol{\alpha}(i)\boldsymbol{\pi}(i)+\alpha_0(i)m_0c^2=-e\phi(\mathbf{r}_i)+c\boldsymbol{\alpha}(i)\boldsymbol{\pi}(i)+\alpha_0(i)m_0c^2$$

is the Dirac Hamiltonian for electron i pointed by vector $\mathbf{r}_i$, whereas the Dirac matrices for electron i, $\boldsymbol{\alpha}(i)=[\alpha_x(i),\alpha_y(i),\alpha_z(i)]$, and the corresponding operators $\pi_\mu(i)$ have been defined on p. 150 and $\phi(\mathbf{r}_i)$ represents the scalar potential calculated at $\mathbf{r}_i$. The wave function Ψ represents a 16-component spinor (here represented by a square matrix of rank 4), because for each electron we would have the usual Dirac bispinor (four-component) and the two-electron wave function depends on the Cartesian product of the components.[54]

The Breit Hamiltonian (in our example, for two electrons in an electromagnetic field) can be approximated by the following useful formula,[55] known as the *Breit–Pauli Hamiltonian*:

$$\hat{H}(1,2)=\hat{H}_0+\hat{H}_1+\ldots\hat{H}_6, \qquad (3.72)$$

---

[53] The Breit equation is invariant with respect to the Lorentz transformation, but only within an accuracy up to some small terms.

[54] In the Breit equation (3.71) the operators in curly brackets "{}" act either by multiplying the 4×4 matrix Ψ by a function (i.e., each element of the matrix) or by a 4×4 matrix resulting from α matrices.

[55] H.A. Bethe, E.E. Salpeter, "*Quantum mechanics of One- and Two-Electron Atoms*," Springer, 1977, p. 181.

where

- $\hat{H}_0 = \frac{\hat{p}_1^2}{2m_0} + \frac{\hat{p}_2^2}{2m_0} + V$ represents the familiar nonrelativistic Hamiltonian;
- $\hat{H}_1 = -\frac{1}{8m_0^3c^2}\left(\hat{p}_1^4 + \hat{p}_2^4\right)$ comes from the velocity dependence of mass, or more precisely from the Taylor expansion of Eq. (3.38) (p. 145) for small velocities;
- $\hat{H}_2 = -\frac{e^2}{2(m_0c)^2}\frac{1}{r_{12}}\left[\hat{\mathbf{p}}_1 \cdot \hat{\mathbf{p}}_2 + \frac{\mathbf{r}_{12}\cdot(\mathbf{r}_{12}\cdot\hat{\mathbf{p}}_1)\hat{\mathbf{p}}_2}{r_{12}^2}\right]$ stands for the correction[56] that accounts in part for the abovementioned retardation; alternatively, the term may be viewed as the interaction energy of two magnetic dipoles, each resulting from the orbital motion of an electron (*orbit–orbit term*);
- $\hat{H}_3 = \frac{\mu_B}{m_0c}\{\left[\boldsymbol{\mathcal{E}}(\mathbf{r}_1) \times \hat{\mathbf{p}}_1 + \frac{2e}{r_{12}^3}\mathbf{r}_{12} \times \hat{\mathbf{p}}_2\right] \cdot \mathbf{s}_1 + \left[\boldsymbol{\mathcal{E}}(\mathbf{r}_2) \times \hat{\mathbf{p}}_2 + \frac{2e}{r_{12}^3}\mathbf{r}_{21} \times \hat{\mathbf{p}}_1\right] \cdot \mathbf{s}_2\}$ is the interaction energy of the electronic magnetic moments (resulting from the abovementioned orbital motion) with the spin magnetic dipole moments (*spin–orbit coupling*), μ_B stands for the Bohr magneton, and $\boldsymbol{\mathcal{E}}$ denotes the electric field vector; since we have two orbital magnetic dipole moments and two spin orbital dipole moments, there are four spin–orbit interactions; the first term in square brackets stands for the spin–orbit coupling of the same electron, while the second term represents the coupling of the spin of one particle with the orbit of the second; altogether we have four interactions;
- $\hat{H}_4 = \frac{ie\hbar}{(2m_0c)^2}\left[\hat{\mathbf{p}}_1 \cdot \boldsymbol{\mathcal{E}}(\mathbf{r}_1) + \hat{\mathbf{p}}_2 \cdot \boldsymbol{\mathcal{E}}(\mathbf{r}_2)\right]$ is a nonclassical term particular to Dirac theory (also present in the one-electron Dirac Hamiltonian) called the *Darwin term*;
- $\hat{H}_5 = 4\mu_B^2\{-\frac{8\pi}{3}(\mathbf{s}_1 \cdot \mathbf{s}_2)\,\delta(\mathbf{r}_{12}) + \frac{1}{r_{12}^3}\left[\mathbf{s}_1 \cdot \mathbf{s}_2 - \frac{(\mathbf{s}_1\cdot\mathbf{r}_{12})(\mathbf{s}_2\cdot\mathbf{r}_{12})}{r_{12}^2}\right]\}$ corresponds to the spin dipole moment interactions of the two electrons (*spin–spin term*). The first term is called the *Fermi contact term,* since it is nonzero only when the two electrons touch one another (see Appendix E, p. 659), whereas the second term represents the classical *dipole-dipole interaction* of the two electronic spins (cf. the multipole expansion in Appendix V2-G, p. V2-613 and Chapter V2-5), i.e., the interaction of the two spin magnetic moments of the electrons (with the factor 2, according to Eq. (3.63), p. 158);
- $\hat{H}_6 = 2\mu_B\left[\mathbf{H}(\mathbf{r}_1) \cdot \mathbf{s}_1 + \mathbf{H}(\mathbf{r}_2) \cdot \mathbf{s}_2\right] + \frac{e}{m_0c}\left[\mathbf{A}(\mathbf{r}_1) \cdot \hat{\mathbf{p}}_1 + \mathbf{A}(\mathbf{r}_2) \cdot \hat{\mathbf{p}}_2\right]$ is known as the *Zeeman interaction*, i.e., the interaction of the spin (the first two terms) and the orbital electronic magnetic dipole moments (the second two terms) with the external magnetic field **H** (cf. Eq. (3.63)).

The terms listed above are of prime importance in the theory of the interaction of matter with the electromagnetic field (e.g., in nuclear magnetic resonance).

---

[56] For noncommuting operators $\hat{\mathbf{a}}\left(\hat{\mathbf{a}} \cdot \hat{\mathbf{b}}\right)\hat{\mathbf{c}} = \sum_{i,j=1}^{3} \hat{a}_i\hat{a}_j\hat{b}_j\hat{c}_i$.

3.6.2 A few words about quantum electrodynamics

The Dirac and Breit equations do not account for several subtle effects. They are predicted by QED, a many-particle theory.

The QED energy may be conveniently developed in a series of $\frac{1}{c}$:

- in zero order we have the nonrelativistic approximation (solution to the Schrödinger equation);
- there are no first-order terms;
- the second order contains the Breit corrections;
- the third and further orders are called the *radiative corrections*.

Radiative corrections

The radiative corrections include:

- *Interaction with the vacuum* (Fig. 3.6a). According to contemporary physics the perfect vacuum does not just represent nothing. The electric field of the vacuum itself fluctuates about zero and these instantaneous fluctuations influence the motion of any charged particle. When a strong electric field operates in a vacuum, the latter undergoes a polarization (*vacuum polarization*), which means a *spontaneous creation of matter*, more specifically, of particle–antiparticle pairs.
- *Interaction with virtual photons*. The electric field influences the motion of electrons. What about its own electric field? Does it influence its motion as well? The latter effect is usually modeled by allowing the electron to emit photons and then to absorb them ("*virtual photons*")[57] (Fig. 3.6d).

The QED calculations performed to date have been focused on the energy. The first calculations of atomic susceptibilities (helium) within an accuracy including the c^{-2} terms were carried out independently by Pachucki and Sapirstein[58] and by Cencek and coworkers,[59] and with accuracy up to c^{-3} (with estimation of the c^{-4} term) by Łach and coworkers (see Table 3.1). To get a flavor what subtle effects may be computed nowadays, Table 3.1 shows the components of the first ionization energy and of the dipole polarizability (see Chapter V2-4) of the helium atom.

---

[57] As remarked by Richard Feynman (see Additional literature in the present chapter, p. 193), for unknown reasons physics is based on the interaction of objects of spin $\frac{1}{2}$ (like electrons or quarks) mediated by objects of spin 1 (like photons, gluons, or W particles).

[58] K. Pachucki, J. Sapirstein, *Phys. Rev. A*, 63(2001)12504.

[59] W. Cencek, K. Szalewicz, B. Jeziorski, *Phys. Rev. Letters*, 86(2001)5675.

Table 3.1. Contributions of various physical effects (nonrelativistic, Breit, QED, and beyond QED) to the ionization energy (third column) and the dipole polarizability α (fourth column) of the helium atom as well as comparison with the experimental values (all quantities in atomic units, i.e., $e = 1, \hbar = 1, m_0 = 1$, where m_0 denotes the rest mass of the electron). The first column gives the symbol of the term in the Breit–Pauli Hamiltonian (3.72) as well as of the QED corrections given order by order (first corresponding to the electron–positron vacuum polarization (QED), then, beyond QED, to other particle–antiparticle pairs (non-QED): $\mu, \pi, \ldots$) split into several separate effects. The second column contains a short description of the effect. The estimated error is given in parentheses in the units of the last figure reported.

Term	Physical interpretation	Ionization energy (MHz)	α (a.u. x 10^{-6})[1]
$\hat{H}_0$	**Schrödinger equation**	**5 945 262 288.62(4)**	**1 383 809.986(1)**
δ	**nonzero size of the nucleus**	**−29.55(4)**	**0.022(1)**
$\hat{H}_1$	**p^4 term**	**1 233 305.45(1)**	**−987.88(1)**
$\hat{H}_2$ (el-el)	electron–electron retardation (Breit interaction)	48 684.88(1)	−23.219(1)
$\hat{H}_2$ (el-n)	electron–nucleus retardation (Breit interaction)	319.16(1)	−0.257(3)
$\hat{H}_2$	**Breit (total)**	**49 004.04(1)**	**−23.476(3)**
$\hat{H}_3$	**spin–orbit**	0	0
$\hat{H}_4$ (el-el)	electron–electron Darwin term	117 008.83(1)	−66.083(1)
$\hat{H}_4$ (el-n)	electron–nucleus Darwin term	−1 182 100.99(1)	864.85(2)
$\hat{H}_4$	**Darwin (total)**	**−1 065 092.16(1)**	**798.77(2)**
$\hat{H}_5$	**spin–spin (total)**	**−234 017.66(1)**	**132.166(1)**
$\hat{H}_6$	**spin–field**	0	0
QED (c^{-3})	vacuum polarization correction to electron–electron interaction	−72.48(1)	0.41(1)
QED (c^{-3})	vacuum polarization correction to electron–nucleus interaction	1 463.00(1)	−1.071(1)
QED (c^{-3})	total vacuum polarization in c^{-3} order	1 390.52(1)	−1.030(1)
QED (c^{-3})	**vac. pol. + other c^{-3} QED correction**	**−40 483.98(5)**	**30.66(1)**
QED (c^{-4})	vacuum polarization	12.26(1)	0.009(1)
QED (c^{-4})	**total c^{-4} QED correction**	**−834.9(2)**	**0.56(22)**
QED-h.o.	**estimation of higher-order QED correction**	**84(42)**	**−0.06(6)**
non-QED	**contribution of virtual muons, pions, etc.**	**0.05(1)**	**−0.004(1)**
$\sum$	**Theory (total)**	**5 945 204 223(42)**[2]	**1 383 760.79(23)**
	Experiment	**5 945 204 238(45)**[3]	**1 383 791(67)**[4]

[1] G. Łach, B. Jeziorski, K. Szalewicz, *Phys. Rev. Letters*, 92(2004)233001

[2] G.W.F. Drake, W.C. Martin, *Can. J. Phys.*, 76(1998)679; V. Korobov, A. Yelkhovsky, *Phys. Rev. Letters*, 87(2001)193003

[3] K.S.E. Eikema, W. Ubachs, W. Vassen, W. Hogervorst, *Phys. Rev. A*, 55(1997)1866

[4] F. Weinhold, *J. Phys. Chem.*, 86(1982)1111

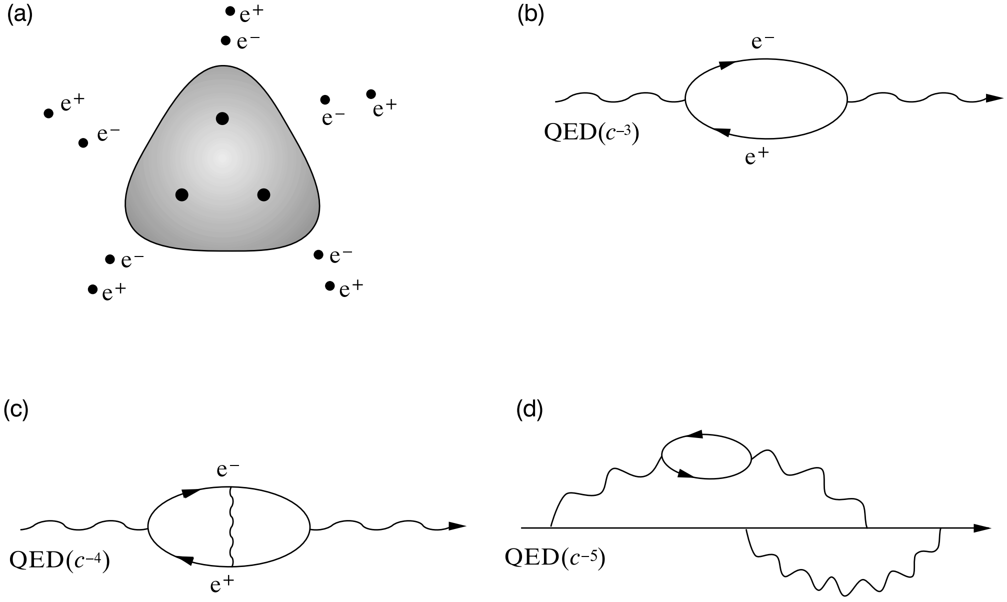


Fig. 3.6. Radiative effects. (a) The electric field close to the proton (composed of three quarks) is so strong that it *creates matter and antimatter* (shown as electron–positron pairs). The three quarks visible in scattering experiments represent the valence quarks. (b) One of the radiative effects in the QED correction of the c^{-3} order (see Table 3.1). The pictures show the sequence of the events from left to right. A photon (wavy line on the left) polarizes the vacuum, an electron–positron pair (solid lines) is created, and the photon vanishes. Then the created particles annihilate each other and a photon is created. (c) A similar event (of the c^{-4} order in QED), but during the existence of the electron–positron pair the two particles interact by exchange of a photon. (d) An electron (horizontal solid line) emits a photon, which creates an electron–positron pair, which annihilates producing another photon. Meanwhile the first electron emits a photon, then first absorbs the photon from the annihilation, and then absorbs the photon emitted by itself earlier. This effect is of the order c^{-5} in QED.

Comments to the table

- $\hat{H}_0$ denotes the result obtained from accurate solution of the Schrödinger equation (i.e., the nonrelativistic and finite nuclear mass theory). Today the solution of the equation could be obtained with greater accuracy than reported here. Imagine that here the theory is limited by the *error by which we know the helium atom mass*, which is "only" 12 significant figures.
- The effect of the nonzero size of the nucleus is small; it is practically never taken into account in computations. If we enlarged the nucleus to the size of an apple, the first Bohr orbit would be 10 km from the nucleus. And still (sticking to our analogy) the electron is able to distinguish a point from an apple? Not quite. It sees the (tiny) difference because the electron knows the region close to the nucleus: it is there that it resides most often. Anyway the theory is able to compute such a tiny effect.

- *The term p^4 and the total Darwin effect nearly cancel each other out for unclear reasons*. This cancellation is persistently confirmed also in other systems. Mysteriously enough, this pertains not only to the ionization energy, but also to the polarizability.
- After the abovementioned cancellation (of p^4 and Darwin terms), the retardation becomes one of the most important relativistic effects. As seen from the table, the effect is about 100 times larger (both for the ionization energy and the polarizability) for the electron–electron retardation than for that of the nucleus–electron. This is quite understandable, because the nucleus represents a "massive rock" (it is about 7000 times heavier) in comparison to an electron, it moves slowly and in the nucleus–electron interaction only the electron contributes to the retardation effect. Two electrons make the retardation much more serious.
- Term $\hat{H}_3$ (spin–orbit coupling) is equal to zero for symmetry reasons (for the ground state).
- In the Darwin term, the nucleus–electron versus electron–electron contribution have reversed magnitudes (about 1 : 10 as compared to 100 : 1 in retardation). Again this time it seems intuitively correct. We have the sum of the particle–particle terms in the Hamiltonian $\hat{H}_4 = \frac{ie\hbar}{(2m_0c)^2}\left[\hat{\mathbf{p}}_1 \cdot \boldsymbol{\mathcal{E}}(\mathbf{r}_1) + \hat{\mathbf{p}}_2 \cdot \boldsymbol{\mathcal{E}}(\mathbf{r}_2)\right]$, where $\boldsymbol{\mathcal{E}}$ means an electric field created by two other particles on the particle under consideration. Each of the terms is proportional to $\nabla_i \nabla_i V = \Delta_i V = 4\pi q_i \delta(\mathbf{r}_i)$, where δ is the Dirac delta function (Appendix E, p. 659) and q_i denotes the charge of the particle i. The absolute value of the nuclear charge is twice the electron charge.
- In term $\hat{H}_5$ the spin–spin relates to the electron–electron interaction, because the helium nucleus has spin angular momentum 0.
- The Coulombic interactions are modified by polarization of vacuum (similarly as two charges in a dielectric medium interact weaker). The table reports such corrections[60] to the electron–electron and the electron–nucleus interactions (QED (c^{-3})) taking into account that electron–positron pairs jump out from the vacuum. One of these effects is shown in Fig. 3.6a. As seen from the table, the nucleus polarizes the vacuum much more easily (about ten times more that the polarization by electrons). Once again the larger charge of the nucleus makes the electric field larger and qualitatively explains the effect. Note that the QED corrections (corresponding to e-p creation) decrease quickly with their order. One of such higher-order corrections is shown in Fig. 3.6d.
- What about the creation of other (than e-p) particle–antiparticle pairs from the vacuum? The larger the rest mass the more difficult it is to squeeze out the corresponding particle–antiparticle pair. And yet, we have some tiny effect (see non-QED entry) corresponding

---

[60] However, these effects represent a minor fraction of the total QED (c^{-3}) correction.

to the creation of such pairs as muon–antimuon (μ), pion–antipion[61] (π), etc. This means that the helium atom is composed of the nucleus and the two electrons only, when we look at it within a certain approximation. To tell the truth, the atom contains also photons, electrons, positrons, muons, pions, and whatever you wish, but with smaller and smaller probability of appearance. All that has only a minor effect of the order of something like the seventh significant figure (both for the ionization potential and for the polarizability).

Summary

The beginning of the 20th century has seen the birth and development of two revolutionary theories: relativity and quantum mechanics. These two theories turned out to be incompatible, and attempts were made to make them consistent. This chapter consists of two interrelated parts:

- introduction of the elements of relativity theory, and
- attempts to make quantum theory consistent with relativity (relativistic quantum mechanics).

Special theory of relativity

- If experiments are to be described in the same way in two laboratories that move with respect to the partner laboratory with constant velocities v and $-v$, respectively, then the *apparent forces have to vanish*. The same event is described in the two laboratories (by two observers) in the corresponding coordinate system (in one the event happens at coordinate x and time t, in the second at x' and t'). A sufficient condition that makes the forces vanish is based on linear dependence: $x' = Ax + Bt$ and $t' = Cx + Dt$, where A, B, C, D denote some constants.
- In order to put both observers on the same footing, we have to have $A = D$.
- The Michelson–Morley experiment has shown that each of the observers will note that in the partner's laboratory there is a *contraction* of the dimension pointing to the partner. As a consequence there is a time *dilation*, i.e., each of the observers will note that time flows slower in the partner's laboratory.
- Einstein assumed that in spite of this, any of the observers will measure the same speed of light (c) in his coordinate system.
- This leads to the *Lorentz transformation* that says where and when the two observers see the same event. The Lorentz transformation is especially simple after introducing the Minkowski space (x, ct): $\begin{bmatrix} x' \\ ct' \end{bmatrix} = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \begin{Bmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{Bmatrix} \begin{bmatrix} x \\ ct \end{bmatrix}$. None of the two coordinate systems is privileged (relativity principle).
- Finally, we derived Einstein's formula $E_{kin} = mc^2$ for the kinetic energy of a body with mass m (this depends on its speed with respect to the coordinate system where the mass is measured).

---

[61] Pions are π mesons, the subnuclear particles with mass comparable to that of the muon, a particle about 200 times more massive than an electron. Pions were discovered in 1947 by C.G. Lattes, G.S.P. Occhialini, and C.F. Powell.

Relativistic quantum dynamics

- Fock, Klein, and Gordon found the total energy for a particle using the Einstein formula for kinetic energy $E_{kin} = mc^2$, adding the potential energy and introducing the momentum[62] $p = mv$. After introducing an external electromagnetic field (characterized by the vector potential $\mathbf{A}$ and the scalar potential ϕ) they obtained the following relation among operators: $\left[\frac{i\hbar\frac{\partial}{\partial t} - q\phi}{c}\right]^2 - \left[(-i\hbar\nabla - \frac{q}{c}\mathbf{A})^2 + m_0^2c^2\right] = 0$, where m_0 denotes the rest mass of the particle.
- Paul Dirac factorized the left-hand side of this equation by treating it as the difference of squares. This gave two continua of energy separated by a gap of width $2m_0c^2$. Dirac assumed that the lower (negative energy) continuum is fully occupied by electrons ("vacuum"), while the upper continuum is occupied by the single electron (our particle). If we managed to excite an electron from the lower continuum to the upper one, then in the upper continuum we would see an electron, while the hole in the lower continuum would have the properties of a positive electron (positron). This corresponds to the creation of the electron–positron pair from the vacuum.
- The Dirac equation for the electron has the form $\left(i\hbar\frac{\partial}{\partial t}\right)\Psi = (q\phi + c\sum_{\mu=x,y,z}\alpha_\mu\pi_\mu + \alpha_0 m_0c^2)\Psi$, where π_μ in the absence of a magnetic field is equal to the momentum operator $\hat{p}_\mu$, $\mu = x, y, z$, while α_μ stand for the square *matrices* of rank 4, which are related to the Pauli matrices (cf. introduction of spin, Chapter 1). In consequence, the wave function Ψ has to be a four-component vector composed of square integrable functions (*bispinor*).
- The Dirac equation demonstrated "pathological" behavior when a numerical solution was sought. The very reason for this was the decoupling of the electron and positron equations. The exact separation of the negative and positive energy continua has been demonstrated by Barysz and Sadlej, but it leads to a more complex theory. Numerical troubles are often removed by an *ad hoc* assumption known as *kinetic balancing*, i.e., fixing a certain relation among the bispinor components. By using this relation we prove that there are two *large* and two *small* (smaller by a factor of about $2\frac{c}{v}$) components of the bispinor.[63]
- The kinetic balance can be used to eliminate the small components from the Dirac equation. Then, the assumption $c = \infty$ (nonrelativistic approximation) leads to the *Schrödinger equation* for a single particle.
- The Dirac equation for a particle in the electromagnetic field contains the interaction of the spin magnetic moment with the magnetic field. In this way spin angular momentum appears in the Dirac theory in a natural way (as opposed to the nonrelativistic case, where it had to be postulated).
- The problem of an electron in the external electric field produced by the nucleus (i.e., the hydrogen-like atom) has been solved exactly within the Dirac model. It turned out that the relativistic corrections are important only for systems with heavy atoms.
- It has been demonstrated in a step-by-step calculation how to obtain an approximate solution of the Dirac equation for the hydrogen-like atom. One of the results is that the relativistic orbitals are contracted compared to the nonrelativistic ones.

---

[62] They wanted to involve the momentum in the formula to be able to change the energy expression to an operator ($\mathbf{p} \rightarrow \hat{\mathbf{p}}$) according to the postulates of quantum mechanics.

[63] For solutions with negative energies this relation is reversed.

- Finally, the Breit equation has been given. The equation goes beyond the Dirac model, by taking into account the *retardation* effects. The Pauli–Breit expression for the Breit Hamiltonian contains several easily interpretable physical effects.
- QED provides an even better description of the system by adding *radiative* effects that take into account the interaction of the particles with the vacuum.

Main concepts, new terms

anticommutation relation (p. 150)
apparent forces (p. 127)
bispinors (p. 152)
Breit equation (p. 171)
Breit–Pauli Hamiltonian (p. 171)
contact term (p. 171)
contraction of orbitals (p. 164)
Darwin solution (p. 159)
Darwin term (p. 176)
Dirac–Coulomb model (p. 167)
Dirac electronic sea (p. 148)
Dirac equation (p. 149)
Dirac vacuum (p. 148)
Einstein equation (p. 143)
electron spin (p. 158)
Fermi contact term (p. 171)
Fock–Klein–Gordon equation (p. 144)
inertial system (p. 129)
Galilean transformation (p. 130)
kinetic balance (p. 154)
Klein–Gordon equation (p. 144)
length contraction (p. 134)
Lorentz transformation (p. 134)
Michelson–Morley experiment (p. 131)
Minkowski space–time (p. 138)
negative energy continuum (p. 148)
positive energy continuum (p. 148)
positron (p. 148)
particle–antiparticle creation (p. 173)
radiative corrections (p. 173)
relativistic mass (p. 143)
relativistic mass effect (p. 125)
relativity principle (p. 138)
retarded potential (p. 171)
spinors (p. 152)
spin–orbit coupling (p. 171)
spin–spin coupling (p. 171)
time dilation (p. 139)
vacuum polarization (p. 173)
velocity addition law (p. 137)
virtual photons (p. 173)

From the research front

DC theory within the mean-field approximation (Chapter 8) is routinely applied to molecules and allows us to estimate the relativistic effects even for large molecules. In the computer era, this means that there are commercial programs available that allow anybody to perform relativistic calculations.

Much worse is the situation with the calculations going beyond the Dirac approach. The first estimation for molecules of relativistic effects beyond the Dirac approximation has been carried out by Ladik[64] and

---

[64] J. Ladik, *Acta. Phys. Hung.*, 10(1959)271.

then by Jeziorski and Kołos.[65] Besides the computation of the Lamb shift for the water molecule,[66] not much has been computed in this area for years.

Then it turned out that a promising approach (the test was for the hydrogen molecule) is to start from an accurate solution to the Schrödinger equation[67] and go directly towards expectation value of the Breit–Pauli Hamiltonian with this wave function (i.e., to *abandon the Dirac equation*), and then to the QED corrections.[68] This Breit–Pauli contribution represents the first complete nonzero relativistic effect proportional to[69] $\left(\frac{1}{c}\right)^2$. Then the complete QED contribution proportional to $\left(\frac{1}{c}\right)^3$ is computed directly, as well as a major part of the $\left(\frac{1}{c}\right)^4$ QED term. The dissociation energy calculated in this way for the hydrogen molecule amounts to $36\,118.0695 \pm 0.0010\ \text{cm}^{-1}$ as compared to the experimental result[70] of $36\,118.0696 \pm 0.0004\ \text{cm}^{-1}$. The estimated size of the neglected QED effects is $\pm 0.0004\ \text{cm}^{-1}$. An even more stringent and equally successful test of this approach of my colleagues will be reported in Chapter 6. It seems therefore that, as for now, when taking the best-known (and the simplest) molecule we may say that the present theory describes Nature with extreme accuracy.

Ad futurum

In comparison with typical chemical phenomena, the relativistic effects in almost all instances remain of marginal significance for biomolecules or molecules typical of traditional organic chemistry. In inorganic chemistry, these effects could however be much more important. Probably the DC theory combined with the mean-field approach will for a few decades remain a satisfactory standard for the vast majority of researchers. At the same time there will be theoretical and computational progress for small molecules (and for atoms), where Dirac theory will be progressively replaced by QED.

In most applications we do not need an accuracy like that reported in the section *From the research front*, but the reason why such results are important is general, pertaining to all those who perform any kind of quantum mechanical calculations. In this way science tests the reliability of its best tools, both experimental and theoretical. The agreement achieved makes all of us confident that we know quite well what we are doing and that the world works in a way very similar to that we think it should.

---

[65] The calculations were performed for the hydrogen molecular ion H_2^+ (B. Jeziorski, W. Kołos, *Chem. Phys. Letters*, 3(1969)677).

[66] P. Pyykkö, K.G. Dyall, A.G. Császár, G. Tarczay, O.L. Polyansky, J. Tennyson, *Phys. Rev. A*, 63(2001)24502.

[67] With the center of mass separated out.

[68] K. Piszczatowski, G. Łach, M. Przybytek, J. Komasa, K. Pachucki, B. Jeziorski, *J. Chem. Theory Comput.*, 5(2009)3039.

[69] The $\frac{1}{c}$ contribution vanishes.

[70] J. Liu, E.J. Salumbides, U. Hollenstein, J.C.J. Koelemeji, K.S.E. Eikema, W. Ubachs, F. Merkt, *J. Chem. Phys.*, 130(2009)174306.

Sooner or later, however, as so many times in the past, we will meet an irreducible discrepancy. If this happens a better theory will have to be constructed – this is how science operates.

Additional literature

H. Bethe, E. Salpeter, "Quantum Mechanics of One- and Two-Electron Atoms," Springer, Berlin, 1957.

This book is exceptional. It is written by excellent specialists in such a competent way and with such care (no misprints) that despite the lapse of many decades it remains the fundamental and best source.

I.M. Grant, H.M. Quiney, "Foundations of the Relativistic Theory of Atomic and Molecular Structure," *Adv. At. Mol. Phys.*, 23(1988)37.

Very good review.

L. Pisani, J.M. André, M.C. André, E. Clementi, *J. Chem. Educ.*, 70(1993)894–901; also **J.M. André, D.H. Mosley, M.C. André, B. Champagne, E. Clementi, J.G. Fripiat, L. Leherte, L. Pisani, D. Vercauteren, M. Vracko, Exploring Aspects of Computational Chemistry: vol. I, Concepts**, Presses Universitaires de Namur, pp. 150–166(1997), **vol. II, Exercises**, Presses Universitaires de Namur, p. 249–272(1997).

Fine article, fine book, written clearly; its strength is also in very simple examples of the application of the theory.

R.P. Feynman, "QED – The Strange Theory of Light and Matter," Princeton University Press, Princeton, 1988.

Excellent book written by one of the celebrities of our times in the style "*quantum electrodynamics not only for poets*."

Questions

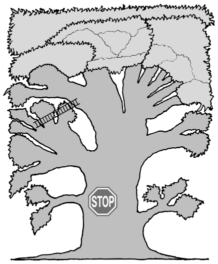
1. In two inertial systems the same forces operate if the two coordinate systems:
    a. are related by Galilean transformation.
    b. move with opposite velocities.
    c. x' and t' are some functions of x and t.
    d. x' and t' depend linearly on x and t.
2. The Michelson–Morley experiment has shown that when an observer in the coordinate system O measures a length in O' (both coordinate systems fly apart; $v' = -v$), then he obtains:
    a. a contraction of the length unit of the observer in O'.
    b. a contraction of all dimensions of the moving body.
    c. the same result that is obtained by an observer in O'.
    d. a contraction of any lengths along the direction of the motion.

3. An observer in O measures the times that a phenomenon takes in the laboratory with the coordinate system O and in the laboratory with the coordinate system O' (both coordinate systems fly apart; $v' = -v$):
   a. time goes with the same speed in O and in O' only if $v = 0$.
   b. the phenomenon in O' goes with the same speed.
   c. if one exchanged the roles of the scientists in O and in O' their conclusions would be the same.
   d. the time of any phenomenon going on in O will be shorter than the same phenomenon going on in O'.
4. In the Minkowski space, the distance of any event from the origin (both coordinate systems fly apart; $v' = -v$) is:
   a. the same for observers in O and in O'
   b. equal to $ct - x$.
   c. the Minkowski distance between the supernova explosion a billion years ago and the event of its observation in our laboratory is equal to 0.
   d. is calculated according to a non-Euclidean metric.
5. A bispinor always represents:
   a. a two-component vector with each component being a complex number.
   b. a four-component vector with functions as components.
   c. a solution of the Dirac equation.
   d. a square integrable vector function.
6. The Dirac electronic sea:
   a. represents the only states accessible for the relativistic free electron.
   b. turned out to be a consequence of the Pauli exclusion principle.
   c. results from forcing the Lorentz invariance for a single free electron.
   d. is infinitely deep.
7. The Schrödinger equation can be deduced from the Dirac equation under the assumption that:
   a. v/c is small.
   b. the speed of the electron is close to c.
   c. all components of the bispinor have equal length.
   d. the magnetic field is zero.
8. The electron spin:
   a. is postulated by the nonrelativistic quantum mechanics.
   b. results from the Darwin term in the Dirac equation
   c. appears in the Dirac equation for the free electron.
   d. appears in the Dirac equation for a single electron in an electromagnetic field.
9. The Dirac hydrogen atom orbitals:
   a. when compared to Schrödinger ones are of the same size, because the nuclear charge has not changed.
   b. when compared to Schrödinger's are more diffuse.
   c. show the largest relativistic effects close to the nucleus.
   d. describe the relativistic electron in an external Coulomb field of the nonrelativistic proton.
10. The Breit equation:
    a. is invariant with respect to the Lorentz transformation.
    b. takes into account the interaction of the magnetic moments of electrons resulting from their orbital motion.

c. takes into account the interaction of the spin magnetic moments.
d. describes many electrons.

Answers

1abd, 2ad, 3acd, 4acd, 5bd, 6cd, 7a, 8ad, 9cd, 10bcd



CHAPTER 4

Exact Solutions – Our Beacons

Simplicity is the extreme degree of sophistication.
Leonardo da Vinci

Where are we?

We are in the middle of the TREE trunk.

An example

When in the 19th century Joseph von Fraunhofer examined sunlight using a glass prism, he found the familiar spectrum is incomplete. In some fixed positions definite colors (characterized by the light wave lengths) were lacking! It is presumed that the interior of the Sun emits the *full* spectrum, but a relatively cold Sun's surface *absorbs some definite colors (absorption lines).* As it turned out later, the reason were atoms present in the Sun's surface. In this way the atoms left in the spectrum their fingerprints. The simplest is the hydrogen atom, responsible for the black absorption lines shown in the upper part of Fig. 4.1 (the fingerprints of other types of atoms have been intentionally removed from the spectrum). The atoms, when heated in a laboratory, emit light with the same frequencies (this emission spectrum is shown in the lower part of Fig. 4.1). Thus, strangely enough, the atoms behave like musical instruments of the Universe! This analogy fits to such an extent, that what determines the position (frequency) of the mysterious spectral lines is the Schrödinger (or *wave*) equation, *essentially the same as that describing functioning of musical instruments*[1]. The exact solution of this equation for the hydrogen atom is given in the present chapter.

What is it all about?

---

[1] At the beginning of the quantum revolution, when no textbooks of quantum physics existed, young enthusiasts used to learn the basics of the new quantum wave mechanics from a music textbook ("*Theory of Sound*" by John William Strutt, baron Rayleigh, London, Macmillan and Co., 1877).

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00012-4

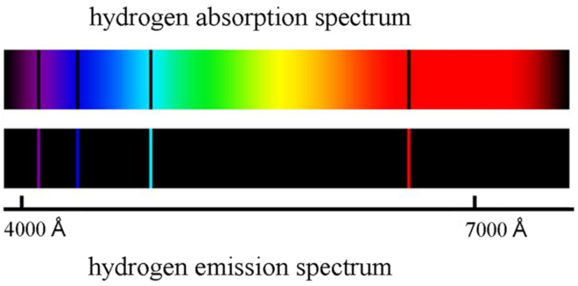


Fig. 4.1. The UV-VIS spectrum of the hydrogen atom (in the wave length scale).

- Four particles
- N particles

Charming SUSY or new solutions (♦) **p. 249**

- SUSY partners
- Relation between SUSY partners
- Examples

Beacons and pearls of physics (△) **p. 255**

Why is this important?

Thanks to computers the Schrödinger equation is nowadays quite easy to solve with *a desired accuracy* for many systems. However, there are only a few systems for which the *exact* solutions are possible. These problems and solutions play an extremely important role in physics, since they represent a kind of beacons for our navigation in science, where as a rule we deal with complex systems. Real systems may most often be approximated by those for which exact solutions exist. For example, a real diatomic molecule is an extremely complex system, difficult to describe in detail, and certainly does not represent a harmonic oscillator. Nevertheless, the main properties of diatomics follow from the simple harmonic oscillator model. When a chemist or physicist has to describe a complex system, he/she always first tries to simplify the problem,[2] to make it similar to one of the simple problems described in the present chapter. Thus, from the beginning *we know the (idealized) solution*. This is of prime importance when discussing the (usually complex) solution to a higher level of accuracy. If this higher-level description differs dramatically from that of the idealized one, most often this indicates that there is an error in our calculations and nothing is more urgent than to find and correct it.

What is needed?

- The postulates of quantum mechanics (Chapter 1, necessary),
- separation of the center-of-mass motion (Appendix J on p. 691, necessary),
- operator algebra (Appendix B on p. 595, necessary).

In the present textbook we assume that the reader knows most of the problems described in the present chapter from a basic course in quantum chemistry. This is why the problems are given in short – only the most important results, without derivation, are reported. On the other hand, such a presentation, in most cases, will be sufficient for our goals.

---

[2] One of the cardinal strategies of science, when we have to explain a strange phenomenon, is to first simplify the system and create a model or series of models (more and more simplified descriptions) that still exhibit the phenomenon. The first model to study should be as simple as possible, because it will shed light on the main machinery.

Classical works

The hydrogen atom problem was solved by Werner Heisenberg in "*Über quantentheoretischen Umdeutung kinematischer und mechanischer Beziehungen*," published in *Zeitschrift für Physik*, 33(1925)879. ★ Erwin Schrödinger arrived at an equivalent picture within his wave mechanics in "*Quantisierung als Eigenwertproblem. I.*," published in *Annalen der Physik*, 79(1926)361. Schrödinger also gave the solution for the harmonic oscillator in a paper (under almost the same title) which appeared in *Annalen der Physik,* 79(1926)489. ★ The Morse oscillator problem was solved by Philip McCord Morse in "*Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels*," in *Physical Review*, 34(1929)57.[3] ★ The tunneling effect was first considered by Friedrich Hund in "*Zur Deutung der Molekelspektren*," published in *Zeitschrift für Physik*, 40(1927)742. ★ The idea of supersymmetry was introduced to quantum mechanics by Candadi V. Sukumar in a work "*Supersymmetry, factorisation of the Schrödinger equation and a Hamiltonian hierarchy*," published in the *Journal of Physics A*, 18(1985)*L*57. ★ The Schrödinger equation for the harmonium[4] was first solved by Sabre Kais, Dudley R. Herschbach, and Raphael David Levine in "*Dimensional Scaling as a Symmetry Operation*," which appeared in the *Journal of Chemical Physics*, 91(1989)7791.

* * *

4.1 Free particle

The potential energy for a free particle is a constant (taken arbitrarily as zero), $V = 0$, and, therefore, energy E represents the kinetic energy only. The Schrödinger equation takes the form

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} = E\Psi$$

or in other words

$$\frac{d^2\Psi}{dx^2} + \kappa^2\Psi = 0,$$

with $\kappa^2 = \frac{2mE}{\hbar^2}$. The constant κ in *this situation*[5] is a real number.

The special solutions to this equation are $\exp(i\kappa x)$ and $\exp(-i\kappa x)$, $\kappa \geq 0$. Their linear combination with arbitrary complex coefficients A' and B' represents the general solution:

$$\Psi = A'\exp(i\kappa x) + B'\exp(-i\kappa x)\,. \tag{4.1}$$

---

[3] Note the spectacular speed at which the scholars worked (Heisenberg's first paper appeared in 1925).

[4] A harmonic model of the helium atom.

[5] The kinetic energy is always positive.

This is a wave of wave length $\lambda = \frac{2\pi}{\kappa}$. Function $\exp(i\kappa x)$ represents the eigenfunction of the momentum operator: $\hat{p}_x \exp(i\kappa x) = -i\hbar \frac{d}{dx} \exp(i\kappa x) = -i\hbar i\kappa \exp(i\kappa x) = \kappa\hbar \exp(i\kappa x)$. For eigenvalue $\hbar\kappa > 0$ the eigenfunction $\exp(i\kappa x)$ describes, therefore, a particle moving towards $+\infty$. Similarly, $\exp(-i\kappa x)$ corresponds to a particle of the same energy, but moving in the opposite direction. The function $\Psi = A' \exp(i\kappa x) + B' \exp(-i\kappa x)$ is a superposition of these two states. A measurement of the momentum can give only two values: $\kappa\hbar$ with probability proportional to $|A'|^2$ or $-\kappa\hbar$ with probability proportional to $|B'|^2$.

4.2 Box with ends (and the music)

The problem pertains to a single particle in a potential (Fig. 4.2a),

$$\begin{aligned} V(x) &= 0 \qquad \text{for } 0 \le x \le L, \\ V(x) &= \infty \qquad \text{for other } x. \end{aligned}$$

Just because the particle will never go outside the section $0 \le x \le L$ (where it would find $V(x) = \infty$), the value of the wave function outside the section is equal to 0. It remains to find the function in $0 \le x \le L$.

Let us write down the Schrödinger equation for $0 \le x \le L$ with the Hamiltonian containing the kinetic energy only (since $V = 0$, one has $E \ge 0$)

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} = E\Psi. \tag{4.2}$$

The solution is given by Eq. (4.1), which may also be written as (A and B are complex numbers)

$$\Psi = A \sin\kappa x + B \cos\kappa x, \tag{4.3}$$

with

$$\kappa^2 = \frac{2mE}{\hbar^2}. \tag{4.4}$$

Now, the key is to recall (Chapter 2) that the wave function has to be continuous and, therefore, two conditions have to be fulfilled: (1) $\Psi = 0$ for $x = 0$ and (2) $\Psi = 0$ for $x = L$. The first condition immediately gives $B = 0$, the second in this situation[6] is equivalent to $\kappa L = n\pi$, for $n = 0, 1, \ldots$. *From this follows energy quantization*, because κ contains energy E. One obtains, therefore, the following solution (a standing wave[7]):

---

[6] A has to be nonzero; otherwise $\Psi = 0$, which is forbidden.

[7] Recall that any stationary state has a trivial time dependence through the factor $\exp(-i\frac{E}{\hbar}t)$. A standing wave at any time t has a standing-still pattern of the nodes, i.e., the points x with $\Psi = 0$.

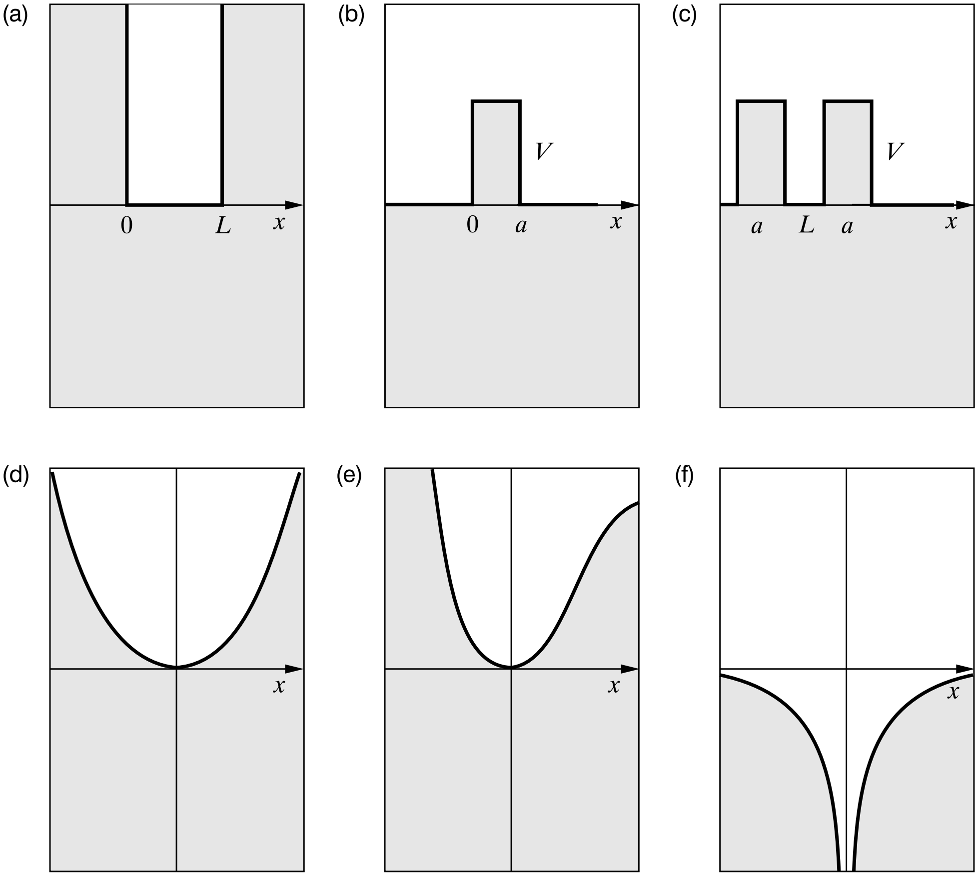


Fig. 4.2. The potential energy functions for (a) particle in a box, (b) single barrier, (c) double barrier, (d) harmonic oscillator, (e) Morse oscillator, and (f) hydrogen atom.

$$E_n = \frac{n^2 h^2}{8mL^2}, \tag{4.5}$$

$$\Psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi}{L} x, \tag{4.6}$$

$n = 1, 2, 3, \ldots,$

because $n = 0$ has to be excluded as it leads to the wave function being equal to zero everywhere, while $n < 0$ may be safely excluded as they lead to the same wave functions as[8] $n > 0$.

---

[8] With the opposite sign, but it does not matter.

We have *chosen* the normalization constant as a positive real number.[9] Fig. 4.3 shows the wave functions for $n = 1, 2, 3$.

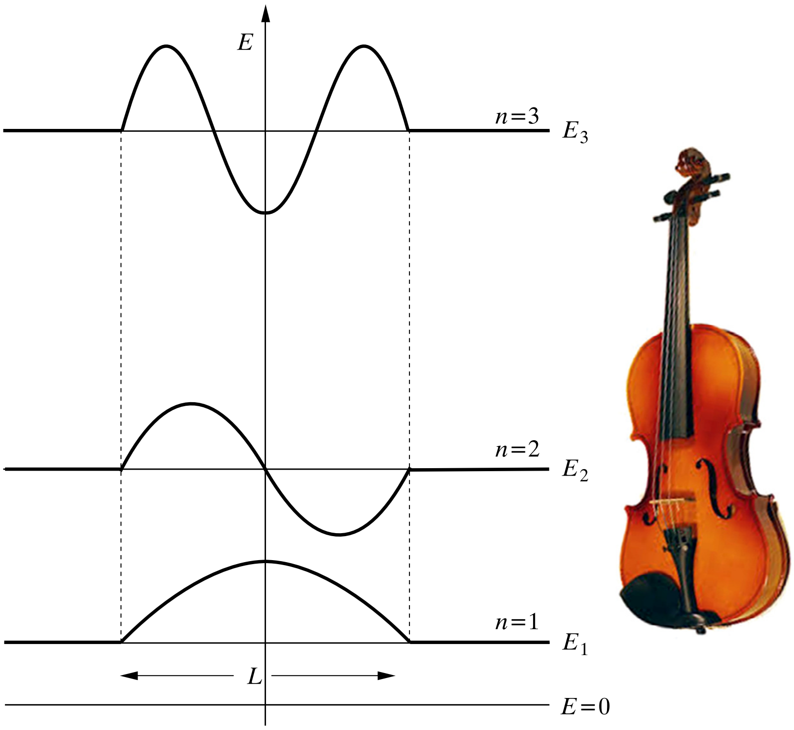


Fig. 4.3. The wave functions for the particle in a box corresponding to $n = 1, 2, 3$. Note the increasing number of nodes when the energy E_i of the stationary state increases. These quantum mechanical solutions are identical with those which the theory of music gives for the vibrations of a (mathematical) string.

Note that the wave functions (plotted as functions of $x, 0 \le x \le L$) are similar to what one may expect for a vibrating violin string of length L (its ends are *immobilized*). In fact, they *are* solutions of the mathematical string equation, identical to Eq. (4.2). If someone had a generator of vibrations with variable frequency and tried to transfer the vibrational energy to the string, he would succeed only for frequencies close to some particular resonance frequencies. The

---

[9] Here is a small exercise showing how to calculate the normalization factor N_n. We have to assure that $\int_0^L |\Psi_n|^2 \, dx = 1$. Inserting the wave function into the left hand side one gets $\int_0^L |\Psi_n|^2 \, dx = |N_n|^2 \int_0^L \left|\sin \frac{n\pi}{L} x\right|^2 dx = |N_n|^2 \frac{L}{n\pi} \int_0^{n\pi} \sin^2 y dy = |N_n|^2 \frac{L}{\pi} \int_0^{\pi} \sin^2 y dy$. The integral $\int_0^{\pi} \sin^2 y dy$ is equal to $\int_0^{\pi} \cos^2 y dy$ and their sum is equal π. Therefore, we continue $|N_n|^2 \frac{L}{\pi} \int_0^{\pi} \sin^2 y dy = |N_n|^2 \frac{L}{\pi} \frac{\pi}{2} = |N_n|^2 \frac{L}{2} = 1$. From this one has $|N_n| = \sqrt{\frac{2}{L}}$ and one *may* choose $N_n = \sqrt{\frac{2}{L}}$ as the normalization constant.

lowest resonance frequency corresponds to such motion of the string that at any given time all its parts have the same sign of the amplitude (this corresponds to Ψ_1). Increasing frequency will not result in energy transfer until the next resonance is reached, with the amplitude of the string such as shown by Ψ_2 (single node, where $\Psi_2 = 0$), similarly for $\Psi_n, n = 3, 4, \ldots$, with the number of nodes equal to $n - 1$.

Example 1 (Butadiene naively). The particle-in-box problem has more to do with chemistry than would appear at first glance.

In organic chemistry, we consider some molecules with conjugated double and single carbon–carbon bonds. One of the simplest is butadiene: $H_2C{=}CH{-}CH{=}CH_2$.

What does this molecule have to do with the particle in a box? Nothing, it seems. First, we have not a single particle but 40 particles (10 nuclei and 30 electrons). Second, where is this constant potential for the motion of the particle? Nowhere. Third, a molecule does not represent a one-dimensional but a three-dimensional object, and in addition, a curved one instead of a section of the x axis. It would seem that any attempt to apply such a primitive theory to our molecule is ridiculous and yet in such a difficult situation we will see the power of the exact solutions reported in the present chapter. All above objections are perfectly justified, but let us try to simplify our system a little.

In the molecule under study the CC bonds are "averaged," which facilitates the motion of the π electrons (this notion will become clear in Chapter 8) along the system. The four π electrons are loosely bound to the molecule, and we may assume that other electrons are always rigidly bound and they will therefore be ignored.

If

- we removed the π electrons from the molecule (and put them temporarily into a safe),
- "ground up" the remaining (positively charged) molecular core and distributed the ground mass uniformly along the x axis within a section of length L equal to the length of the molecule (averaging the potential energy for a charged particle) to construct a kind of highway for the π electrons, and
- added the first π electron from the safe, then

this single electron would represent something similar to a particle in a box.[10] Assuming this simplified model we know all the details of the electron distribution, including the ground state ψ_1 and excited wave functions $\psi_2, \psi_3, \ldots$ (in the one-particle case called the orbitals). If we now took the π electrons from the safe, added them one by one to the system, assuming that

---

[10] Almost, because the potential is not quite constant. Also one might remove the particle from the box at the expense of a large but finite energy (ionization), which is not feasible for the particle in a box.

they would not see one another,[11] taking into account the Pauli exclusion principle (the double occupancy by electrons of the individual orbitals described in more detail in Chapter 8), then we would obtain information about the electron density distribution in the molecule. The idea we are describing is called the Free Electron Molecular Orbitals (FEMO) method.

In our example the orbitals are real and the total electron density distribution (normalized to four π electrons, i.e., giving 4 (electrons) after integration over x) is given as[12]

$$\rho(x) = 2\psi_1^2 + 2\psi_2^2 = 2\frac{2}{L}\sin^2\frac{\pi}{L}x + 2\frac{2}{L}\sin^2\frac{2\pi}{L}x = \frac{4}{L}\left(\sin^2\frac{\pi}{L}x + \sin^2\frac{2\pi}{L}x\right).$$

The function $\rho(x)$ is shown in Fig. 4.4a.

It is seen that

1. $\rho(x)$ is largest on the outermost bonds in the molecule, exactly where chemists put their two little lines to symbolize a double bond;
2. π electron density, i.e., $\rho(x)$, is nonzero in the center. This means that the bond over there is not strictly a single bond.

This key information about the butadiene molecule has been obtained at practically no cost from the primitive FEMO model.

Of course, we cannot expect the description to reflect all the details of the charge distribution in the butadiene molecule, but one may expect this approach to be able to reflect at least some rough features of the π electron distribution. *If the results of more advanced calculations contradicted the rough particle-in-box results, then we should take a closer look at them and search for an error.* This is the strength of the simple exact model systems. They play the role of the beacons – points of reference.

4.3 Cyclic box

The one-dimensional box described above is similar to a stick in which the particle can move. The butadiene molecule is rather similar to such a stick and, therefore, the one-dimensional box models it quite well.

---

[11] As we will see in Chapter 8, this approximation is more realistic than it sounds.

[12] The student "i" is characterized by a probability density distribution $\rho_i(x)$ of finding him at coordinate x (we limit ourselves to a single variable, measuring his position, say, on his way from the dormitory to the university). If all students moved independently, the sum of their individual probability densities at point x_0, i.e., $\rho(x_0) = \sum_i \rho_i(x_0)$ would be proportional to the probability density of finding any student at x_0. The same pertains to electrons, when assumed to be independent.

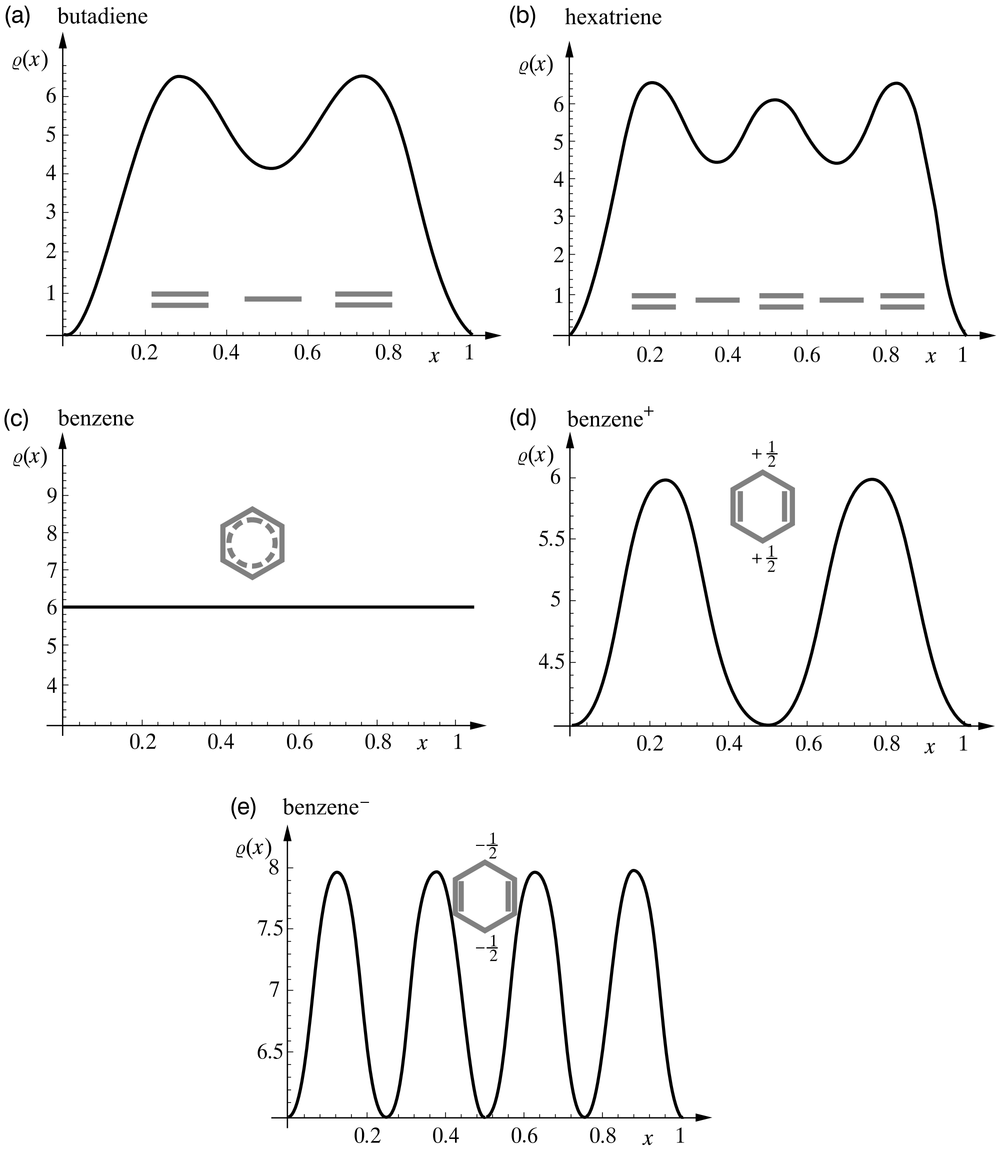


Fig. 4.4. π electron density charge distributions for several molecules computed by the FEMO method. The length L of each molecule has been assumed to be equal to 1. For other lengths the charge distributions are similar. The electron density for four electrons (a) in butadiene and (b) of six electrons in hexatriene. The electron density maxima coincide with the positions chemists write as double bonds. (c) *The six-electron density distribution in the benzene molecule is peculiar, because it is constant along the perimeter of the molecule*. If we (d) subtract an electron from benzene or (e) add an electron to it, then maxima and minima of the π electron density appear. If an electron is subtracted (d) there are two maxima (double bonds) and two π electron-deficient regions denoted as having charge $+\frac{1}{2}$. After one π electron is added (e) we obtain four maxima (two double bonds and two electron-rich regions denoted by charge $-\frac{1}{2}$).

And what can model the benzene molecule? In a crude approximation we may think of benzene as a stick with the two ends *joined* in such a way as to be unable to recognize where the union has taken place. Limiting ourselves to this effect,[13] we may use the solution given by Eq. (4.3) and impose appropriate boundary conditions. What could these boundary conditions be? The wave function at the two ends of the box has to be stitched together without leaving any trace of the seam. This is achieved by two boundary conditions: $\Psi(0) = \Psi(L)$, forcing the two wave function values to match, and $\Psi'(0) = \Psi'(L)$, making the seam "invisible".[14] The two conditions mean:

$$\begin{gathered} A \sin \kappa 0 + B \cos \kappa 0 = A \sin \kappa L + B \cos \kappa L, \\ A\kappa \cos \kappa 0 - B\kappa \sin \kappa 0 = A\kappa \cos \kappa L - B\kappa \sin \kappa L \end{gathered}$$

or

$$\begin{aligned} B &= A \sin \kappa L + B \cos \kappa L, \\ A &= A \cos \kappa L - B \sin \kappa L. \end{aligned}$$

To find a nontrivial solution the determinant of the coefficients at the unknown quantities A and B has to vanish[15]: $\begin{vmatrix} \sin \kappa L & \cos \kappa L - 1 \\ \cos \kappa L - 1 & -\sin \kappa L \end{vmatrix} = 0$, which is equivalent to

$$\cos \kappa L = 1.$$

The last condition gives $\kappa L = 2\pi n, n = 0, \pm 1, \pm 2, \ldots$. This immediately produces a formula for the energy very similar to that for the box with ends, but with the replacement $n \to 2n$:

---

[13] And neglecting such effects as the particular shape of the benzene (curvature, etc.).

[14] There is no such a thing in Nature as infinitely steep potential energy walls or infinite values of the potential energy (like in the particle-in-a-box problem). This means we should treat such idealized cases as limit cases of possible continuous potential energy functions. From the Schrödinger equation $-\frac{\hbar^2}{2m}\frac{\partial^2 \psi}{\partial x^2} = E\psi - V\psi$ we see that in such a case the continuity of V implies (ψ must be continuous) the continuity of $\frac{\partial^2 \psi}{\partial x^2}$. The continuity of the second derivative results in the continuity of the first derivative as well. What about a nonphysical case of a discontinuous V like in the rectangular barrier case? Well, then we lose the continuity of the second derivative by definition, but still we may have and indeed we have the continuity of the first derivative (this we force successfully in the case just being described). In a more drastic (and nonphysical) case of the discontinuity like in the particle in a box (infinitely steep V and, on top of that, $V = \infty$), we lose even continuity of the first derivative (at $x = 0, L$).

[15] This is a set of homogeneous linear equations.

$$E_n = \frac{(2n)^2 h^2}{8mL^2}, \tag{4.7}$$

where this time $n = 0, \pm 1, \pm 2, \ldots$.

The corresponding wave functions are

$$\begin{aligned}
\psi_0 &= \sqrt{\frac{1}{L}} \text{ for } n = 0,\\
\psi_{n>0} &= A \sin \frac{2\pi n}{L} x + B \cos \frac{2\pi n}{L} x,\\
\psi_{n<0} &= -A \sin \frac{2\pi\, |n|}{L} x + B \cos \frac{2\pi\, |n|}{L} x.
\end{aligned}$$

Since $\psi_{n>0}$ and $\psi_{n<0}$ correspond to the same energy, any combination of them also represents an eigenfunction of the Schrödinger equation corresponding to the same energy (Appendix B on p. 595). Taking therefore as the new wave functions (for $n \neq 0$) the normalized sum and difference of the above wave functions, we finally obtain the solutions to the Schrödinger equation

$$\begin{aligned}
\Psi_0 \equiv \psi_0 &= \sqrt{\frac{1}{L}} \text{ for } n = 0,\\
\Psi_{n>0} &= \sqrt{\frac{2}{L}} \sin \frac{2\pi n}{L} x \;\text{ for } n > 0,\\
\Psi_{n<0} &= \sqrt{\frac{2}{L}} \cos \frac{2\pi n}{L} x \;\text{ for } n < 0.
\end{aligned}$$

4.3.1 Comparison of two boxes: hexatriene and benzene

Let us take an example of two molecules: hexatriene and benzene (i.e., cyclohexatriene) (Fig. 4.5). Let us assume for simplicity that the length of hexatriene L is equal to the perimeter of benzene.[16] Both molecules have six π electrons (any of them). The electrons doubly occupy (the Pauli exclusion principle) three one-electron wave functions corresponding to the lowest

---

[16] This is to some extent an arbitrary assumption, which simplifies the final formulae nicely. In such cases we have to be careful that the conclusions are valid.

energies. Let us compute the sum of the electron energies as the "total electron energy"[17] (in the units $\frac{h^2}{8mL^2}$, to have the formulae as compact as possible):

- HEXATRIENE: $E_{heks} = 2 \times 1 + 2 \times 2^2 + 2 \times 3^2 = 28$,
- BENZENE: $E_{benz} = 2 \times 0 + 2 \times 4 + 2 \times 4 = 16$.

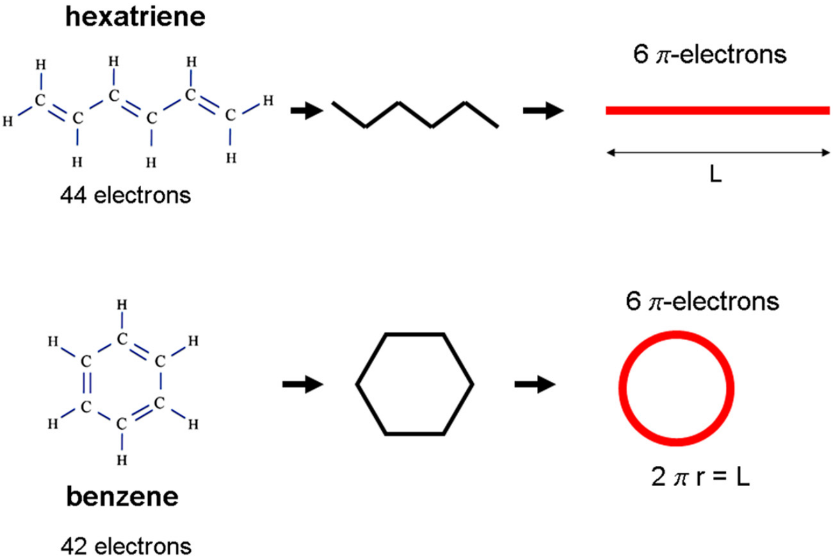


Fig. 4.5. Hexatriene (C_6H_8) and benzene (C_6H_6) contain 44 and 42 electrons, respectively. Both molecules are planar and contain six π electrons (very mobile ones) each. In the Free Electron Molecular Orbitals approach one considers explicitly only these six electrons per molecule. Other electrons as well as the atomic details are ignored. One ends up with each molecule representing a highway for the six π electrons, which are treated as being effectively mutually independent. The hexatriene highway represents a section $(0, L)$, while that for the benzene molecule is a ring of circumference L.

We conclude that six π electrons in the benzene molecule correspond to lower energy (i.e., are more stable) than the six π electrons in the hexatriene molecule. Chemists find this experimentally: the benzene ring with its π electrons survives in many chemical reactions, whereas this rarely happens to the π electron system of hexatriene.

> Our simple theory predicts the benzene π electron system is more stable than that of the hexatriene molecule.

And what about the electronic density in both cases? We obtain (Fig. 4.4b,c)

---

[17] As will be shown in Chapter 8, this method represents an approximation.

- HEXATRIENE: $\rho(x) = 2 \times \frac{2}{L}\left[\sin^2 \frac{\pi}{L}x + \sin^2 \frac{2\pi}{L}x + \sin^2 \frac{3\pi}{L}x\right]$,
- BENZENE: $\rho(x) = 2 \times \frac{1}{L} + 2 \times \frac{2}{L}[\sin^2 \frac{2\pi}{L}x + \cos^2 \frac{2\pi}{L}x] = \frac{6}{L}$.

This is an extremely interesting result.

> The π electron density is constant along the perimeter of the benzene molecule.

No single and double bonds – *all CC bonds are equivalent* (Fig. 4.4c). Chemical experience led chemists already a long time ago to the conclusion that all the CC bonds in benzene are equivalent. This is why they decided to write down the benzene formula in the form of a regular hexagon with a circle in the middle (i.e., not to give the single and double bonds). The FEMO method reflected that feature in a naive way. Do the π electrons not see where the carbon nuclei are? Of course they do. We will meet some more exact methods in further chapters of this textbook, which give a more detailed picture, but it will turn out that *all CC bonds would have the same density distribution,* similar to the solution given by the primitive FEMO method. From Eq. (4.7) it follows that this will happen not only for benzene, but also for all the systems with $(4n+2)$ electrons, $n = 1, 2, \ldots$ because of a very simple (and, therefore, very beautiful) reason that $\sin^2 x + \cos^2 x = 1$ for any x.

The addition or subtraction of an electron makes the distribution nonuniform (Fig. 4.4d,e). Also in the six-π electron hexatriene molecule, uniform electron density is out of the question (Fig. 4.4b). Note that the maxima of the density coincide with the double bonds chemists like to write down. However, even in this molecule, there is still a certain equalization of bonds, since the π electrons are also where the chemists write a single bond (although the π electron density is smaller over there[18]).

Again important information has been obtained at almost no cost.

Two-dimensional rectangular box

Let us consider a rectangular box (Fig. 4.6) with sides L_1 and L_2 and $V = 0$ inside and $V = \infty$ outside. We first separate the variables x and y (the total Hamiltonian represents a sum of two independent one-dimensional Hamiltonians), which leads to the two one-dimensional Schrödinger equations (solved as shown above).

The energy eigenvalue is, therefore, equal to the sum of the energies for the one-dimensional problems,

---

[18] Where, according to the classical chemical picture, no π electron should be.

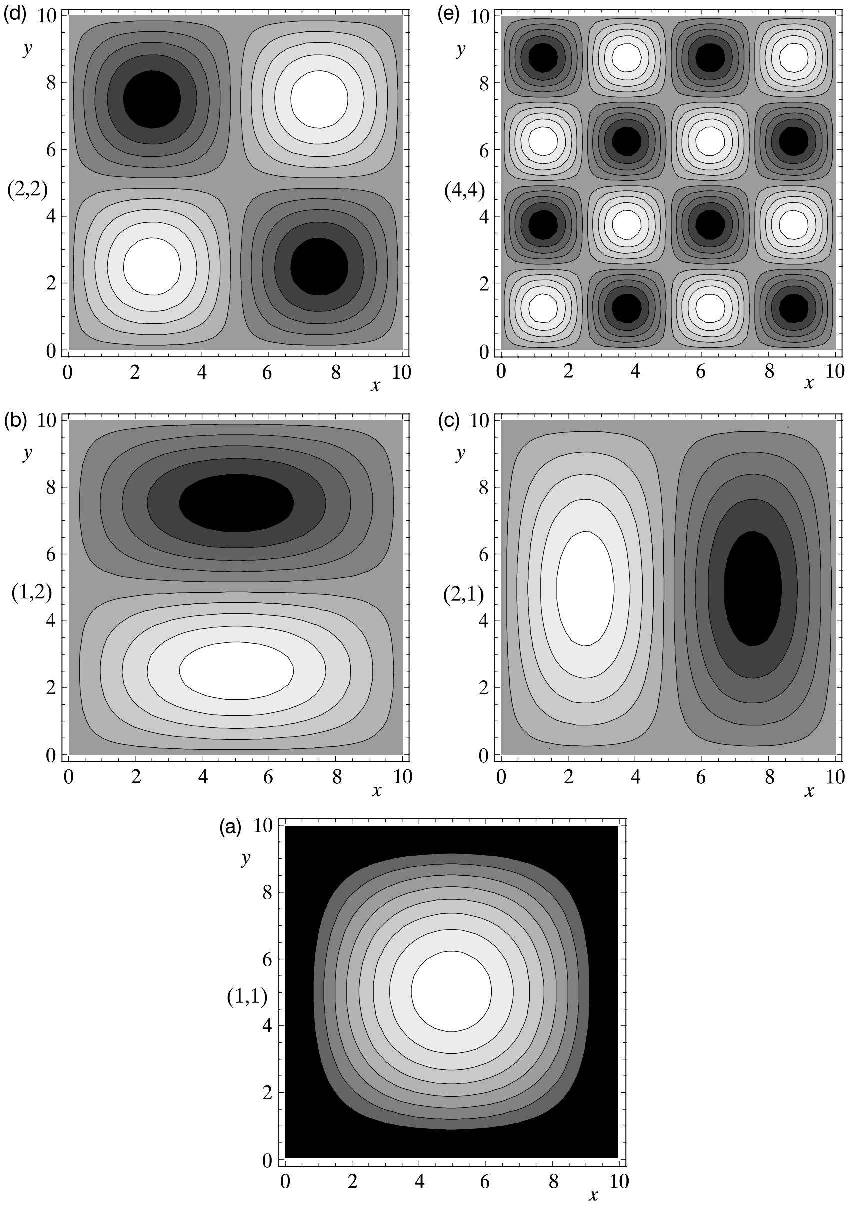


Fig. 4.6. Examples of the wave functions for a particle in a square box, the quantum numbers (n_1, n_2) correspond to (a) $(1, 1)$, (b) $(1, 2)$, (c) $(2, 1)$, (d) $(2, 2)$, and (e) $(4, 4)$. In the cases shown, the higher the energy, the more nodes in the wave function. This rule is not generally true. For example, in a rectangular box with $L_1 \gg L_2$ even a large increase of n_1 does not raise the energy too much, while introducing a lot of nodes. On the other hand, increasing n_2 by 1 raises the energy much more, while introducing only one extra node. A reader acquainted with hydrogen atom orbitals will easily recognize the resemblance of the above figures to some of them (cf. pp. 235–239), because of the rule mentioned above.

$$E_n = \frac{h^2}{8m}\left(\frac{n_1^2}{L_1^2} + \frac{n_2^2}{L_2^2}\right), \tag{4.8}$$

while the wave function has the form of the product of both one-dimensional solutions

$$\Psi_{n_1 n_2} = 2\sqrt{\frac{1}{L_1 L_2}} \sin\frac{n_1\pi}{L_1}x \cdot \sin\frac{n_2\pi}{L_2}y, \tag{4.9}$$

where $n_1, n_2 = 1, 2, \ldots$.

If someone cut out a plywood square, immobilized the square sides, and tried to transfer vibrations using a vibration generator, he/she would be most effective using some resonance frequencies. The distribution of amplitudes of vibrations of the plywood (as functions of x and y) would be very similar to what we see in Fig. 4.6. Note that for the plywood square any frequency except the lowest one corresponds to *two* vibrational modes, exactly as in the two-dimensional square box ($L_1 = L_2 = L$) one has two states $\Psi_{n_1 n_2}$ and $\Psi_{n_2 n_1}$ (double degeneracy). If one made the sides different (of the plywood or of the two-dimensional box) the degeneracy would be lifted. Thus, the degeneracy is a consequence of symmetry.

4.4 Carbon nanotube

We know the graphite from our pencils. It is useful because of a remarkable feature: it is able to leave traces on paper, which are in fact layered carbon structures. Why? This is related to the unique structure of graphite. The graphite represents a stack of honeycomb resembling identical carbon monolayers (graphenes), each atom in the graphene bound by three *strong* identical carbon–carbon chemical bonds with its nearest neighbors (Fig. 4.7a). What we profit from however in our pencils is that the graphene sheets themselves are loosely bound together by the relatively *weak* intermolecular forces only. This is why the graphite leaves traces on paper. A single graphene sheet, similarly as the benzene ring, has the most mobile π electrons (one per carbon atom). These electrons may be treated in the spirit of the FEMO method as a set of independent electrons. Thus, similarly to what we have seen for hexatriene and benzene (p. 192) a free graphene may be approximated by a rectangular box for an electron, as described on p. 200.

One may roll the graphene sheet and seam the sides by the carbon–carbon bonds to make a cylinder – the famous carbon nanotube[19] (Fig. 4.7c). Soon we will see how to apply the FEMO method to a nanotube.

Note that there is however a possibility to introduce in the graphene sheet some number of carbon pentagons instead of hexagones. This causes some off-plane deformation of the graphene

---

[19] The rolling up may be accomplished in several different ways: making the seam directly "head-to-head", i.e., without any shift when seaming, with one carbon shift, two carbon shifts, etc. The resulting nanotubes preserve the locally hexagonal structure, but their physical and chemical properties strongly depend not only on the radius of the nanotube, but first of all on the abovementioned shift.

sheet. Such deformations, when combined, may lead even to closing a space forming fullerenes (the most important of them is C_{60}, of a football shape, Fig. 4.7b). Thus all the carbon allotropes shown are structurally related to each other.

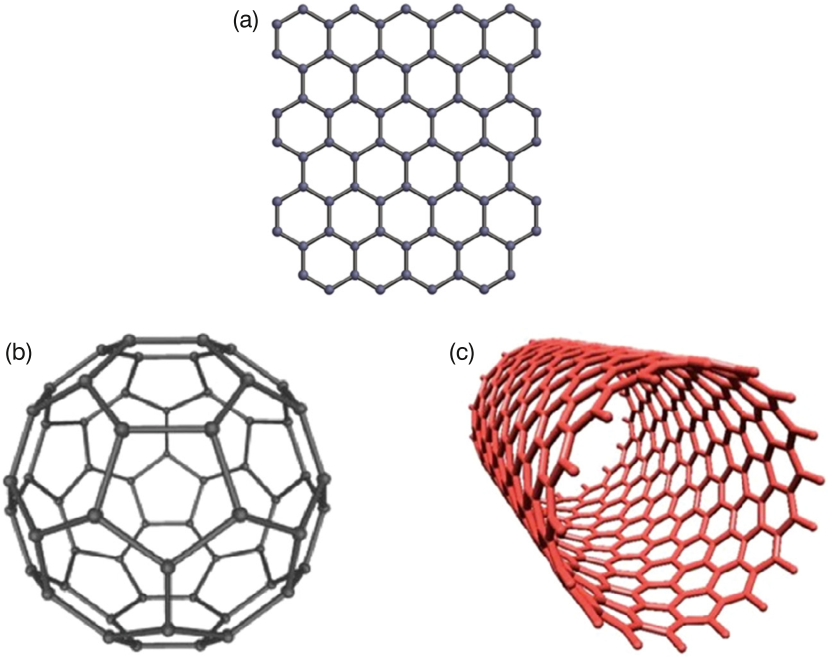


Fig. 4.7. Some important carbon allotropic forms. (a) graphene, (b) fullerene, and (c) nanotube.

Massimo Fusaro[20] treated the nanotube within the FEMO method as a surface for free motion of independent electrons. The calculated energy levels are then occupied by a corresponding number of electrons (one from each carbon atom). Radius $R = \frac{3ja}{2\pi}$ of the cylinder is defined through the number $(2j)$ of the carbon atoms terminating the nanotube, where $a = 1.42$ Å is the CC nearest neighbor distance in typical systems with the conjugated single and double CC bonds. The length of the cylinder can be calculated as $L = (\frac{N_C}{2j} + 1)\frac{\sqrt{3}a}{2}$, where N_C stands for the total number of carbon atoms in the nanotube. The total number of electrons (also N_C) considered in the FEMO model can be calculated from the radius and the length of the nanotube as $N_C = \frac{4}{3}\left(\frac{2\pi R \cdot L}{a^2\sqrt{3}} - \frac{\pi R}{a}\right)$.

A nanotube is a complicated many-electron molecule and, it seems, its electronic structure can be revealed only after long computations giving an approximation to the solution of the

---

[20] M. Fusaro, *J. Comput. Theor. Nanoscience*, 6(2009)1175. A general approach to infinite nanotubes using their translational symmetry (see Chapter V2-1) is described in D.J. Klein, W.A. Seitz, T.G. Schmalz, *J. Phys. Chem.*, 97(1993)1231.

Schrödinger equation. And again we may admire the power of simple models. After assuming the continuous cylinder model one may expect for $R = 0$ that we have to recover the solution for the box with ends, while for $L = 0$ and $R \neq 0$ one should get the solution for the cyclic box. It turns out that in principle we get indeed something like that, except that the shortest cylinder built of carbon atoms cannot have zero length.

The Schrödinger equation for a particle on a side surface of a cylinder (assuming the cylinder axis as x, position of the particle is given by x, ϕ, where ϕ is an angle measuring rotation about x) can be solved exactly. After applying the boundary conditions (the wave function has to vanish at the ends of the cylinder) one gets the following solution (A_{nl} is a normalization constant):

$$\psi_{n,l>0}(x,\phi) = A_{nl}\sin(l\phi)\sin\left(\frac{n\pi}{L}x\right) \qquad \text{for} \quad l > 0,$$
$$\psi_{n,l<0}(x,\phi) = A_{nl}\cos(l\phi)\sin\left(\frac{n\pi}{L}x\right) \qquad \text{for} \quad l < 0,$$
$$\psi_{n,l=0}(x,\phi) = A_{n0}\sin\left(\frac{n\pi}{L}x\right) \qquad \text{for} \quad l = 0,$$

where $n = 1, 2, ..., \frac{N_C}{2j}$ (due to the atomic resolution of our model a larger n would lead to repeating of the already obtained solutions), while $l = 0, \pm 1, \pm 2, ...$ The wave functions given above represent nothing else but products of the wave function for the box with ends and for the cyclic box.

The calculated energy eigenvalues are

$$E_{nl} = \frac{n^2h^2}{8mL^2} + \frac{(2l)^2\,h^2}{8m\,(2\pi R)^2},$$

which means that we have a sum of energy for the box with ends and of the cyclic box (for the length of the box equal to $2\pi R$). For $l \neq 0$ we get double degeneracy, for $l = 0$ the levels are not degenerated.

Thus, a particle confined to the side surface of a cylinder belongs also to those problems of quantum mechanics that can be solved exactly.

Nanotube in more detail

Does the FEMO orbital model describe a real nanotube? Maybe it is just a nice exercise unrelated to reality. In contemporary quantum chemistry we have a toolbox with many methods that are much more accurate than FEMO. One of these methods is known as Density Functional

Theory (DFT, Chapter V2-3) and proved to be a reliable tool. Fusaro carried out DFT computations of what is known as reactivity indices[21] that measure which positions on the nanotube surface are more reactive than others. These particular reactivity indices have been associated with what is known as the Fukui function,[22] which describes the rate at which the electric charge *at a particular point* changes when the total electric charge of the nanotube varies. Therefore, this is an indication where an attacking ion would be directed when it approaches the nanotube (it will choose the place with the largest charge opposite to its own charge). It turned out that the Fukui function for the nanotube as a function of position changes quasiperiodically (not counting the end effects) along the nanotube as well as around the nanotube axis. The values of the Fukui function calculated within the FEMO model and those computed by the much more advanced DFT approach agreed semiquantitatively, both revealing the abovementioned quasiperiodicity.

Such results are encouraging. We see that at least in this case we can understand the machinery by which the world is operating by seeing how its main features appear from something very simple – a kind of primitive theory in which we control and understand every detail. A more sophisticated theory may add some new (and sometimes useful and important) features, but in many cases does not have the power of showing convincingly why something happens.

4.5 Single barrier

Is it possible to pass through a barrier with less energy than the barrier height? Yes.

4.5.1 Tunneling effect below the barrier height

Let us imagine a rectangular potential energy barrier (Fig. 4.2): $V(x) = V_0$ for $0 \leq x \leq a$, with $V(x) = 0$ for other values of x ($V_0 > 0$ is a number). Let us assume a particle of mass m going from the negative values of x (i.e., from the left-hand side); its kinetic energy is equal to E.

The x axis will be divided into three regions:

1. $-\infty < x < 0$, then $V(x) = 0$,
2. $0 \leq x \leq a$, then $V(x) = V_0$,
3. $a < x < \infty$, then $V(x) = 0$.

So, the barrier region has length a. In each of these regions the Schrödinger equation will be solved, and then the solutions will be stitched together in such a way as to make it smooth at

---

[21] M. Fusaro, *J. Comp. Theor. Nanoscience*, 11(2010)2393.

[22] K. Fukui, T. Yonezawa, H. Shingu, *J. Chem. Phys.*, 20(1952)722; K. Fukui, T. Yonezawa, C. Nagata, *J. Chem. Phys.*, 21(1953)174; T. Yonezawa, C. Nagata, *Bull. Chem. Soc. Japan*, 27(1954)423.

any boundary between the regions. The solutions for each region separately can be written very easily as[23]

$$\Psi_1(x) = A_1 e^{i\kappa_0 x} + B_1 e^{-i\kappa_0 x}, \tag{4.10}$$

$$\Psi_2(x) = A_2 e^{i\kappa x} + B_2 e^{-i\kappa x}, \tag{4.11}$$

$$\Psi_3(x) = A_3 e^{i\kappa_0 x} + B_3 e^{-i\kappa_0 x}, \tag{4.12}$$

where A and B represent amplitudes of the de Broglie waves running right and left, respectively, where $\kappa_0^2 = \frac{2mE}{\hbar^2}$ in regions 1 and 3, while in region[24] 2 we have $\kappa^2 = \frac{2m(E-V_0)}{\hbar^2}$. The quantities $|A_i|^2$ and $|B_i|^2$ are proportional to the probability of finding the particle going right ($|A_i|^2$) and left ($|B_i|^2$) in the region i.

In regions 1 and 2 one may have the particle going right, but also left because of reflections from the boundaries, hence in regions 1 and 2 A and B will have nonzero values. However, in region 3 we will have $B_3 = 0$, because there will be no wave going back (it has nothing to reflect on). We have to work on choosing such A and B, so that wave functions of connected regions match at the boundaries (the functions should "meet") and, at each boundary, have the same value of the first derivative (they should meet smoothly). Satisfaction of these requirements is sufficient to determine the ratios of the coefficients; this is what we are looking for when aiming at describing what will happen more often than something else.

Therefore, such a perfect stitching means that for $x = 0$ and $x = a$ one has

$$\begin{aligned}
\Psi_1(x=0) &= \Psi_2(x=0),\\
\Psi_2(x=a) &= \Psi_3(x=a),\\
\Psi_1'(x=0) &= \Psi_2'(x=0),\\
\Psi_2'(x=a) &= \Psi_3'(x=a).
\end{aligned}$$

Using Eqs. (4.10)–(4.12) we may rewrite the same as

$$\begin{aligned}
A_1 + B_1 &= A_2 + B_2,\\
A_2 e^{i\kappa a} + B_2 e^{-i\kappa a} &= A_3 e^{i\kappa_0 a},\\
\kappa_0 A_1 - \kappa_0 B_1 &= \kappa A_2 - \kappa B_2,\\
A_2 \kappa e^{i\kappa a} - B_2 \kappa e^{-i\kappa a} &= A_3 \kappa_0 e^{i\kappa_0 a}.
\end{aligned}$$

---

[23] This is the wave function for free particles. The particle has the possibility (and therefore also the probability) to move left or right. The formulae are natural, because in a particular region considered the potential energy represents a constant (in regions 1 and 3 $V(x) = 0 = const$, while in region 2 $V(x) = V_0 = const'$).

[24] We will obtain these formulae after solving in each region the Schrödinger equation $\frac{\partial^2 \Psi}{\partial x^2} + \kappa^2 \Psi = 0$.

We have five unknowns: A_1, B_1, A_2, B_2, and A_3, but only four equations to determine them. Please note that we are interested only in $\frac{|A_3|^2}{|A_1|^2} \equiv D$, because this is the probability to pass the barrier assuming the particle was sent from the left for sure. Therefore, after dividing all equations by A_1 we have only four unknowns, $\frac{B_1}{A_1} \equiv b_1$, $\frac{A_2}{A_1} \equiv a_2$, $\frac{B_2}{A_1} \equiv b_2$, and $\frac{A_3}{A_1} \equiv a_3$, and four equations to determine them:

$$\begin{gathered}
1 + b_1 = a_2 + b_2, \\
a_2 e^{i\kappa a} + b_2 e^{-i\kappa a} = a_3 e^{i\kappa_0 a}, \\
\kappa_0 - \kappa_0 b_1 = \kappa a_2 - \kappa b_2, \\
a_2 \kappa e^{i\kappa a} - b_2 \kappa e^{-i\kappa a} = a_3 \kappa_0 e^{i\kappa_0 a}.
\end{gathered}$$

Solution of this set of equations (for a given $E > 0$) gives the *transmission coefficient* $D(E) = |a_3|^2$ as function of energy

$$D = \frac{1}{1 + \frac{1}{4\beta(\beta-1)} \sin^2 \kappa a}, \tag{4.13}$$

where $\beta = \frac{E}{V_0}$ is a ratio of the impact energy and the barrier height. The quantity $\kappa = \sqrt{\frac{2m(E-V_0)}{\hbar^2}}$ is positive for $E > V_0$ and we use directly Eq. (4.13). For $E < V_0$ there is a problem, because κ is imaginary, $\kappa = \sqrt{\frac{2m(E-V_0)}{\hbar^2}} = ik$, where $k = \sqrt{\frac{2m(V_0-E)}{\hbar^2}} > 0$. However, calculation of D is simple again; for $E < V_0$ we obtain[25]

$$D = \frac{1}{1 + \frac{1}{4\beta(1-\beta)} sh^2 ka}. \tag{4.14}$$

When $E = V_0$, we are in trouble, because $\beta = 1$ and $\frac{1}{(\beta-1)} \sin^2 \kappa a$ is an expression of the type $\frac{0}{0}$. Using, however, the de l'Hospital rule we get[26]

$$D = \frac{1}{1 + \frac{ma^2}{2\hbar^2} V_0}.$$

---

[25] Because $\sin \kappa a = \frac{\exp(i\kappa a) - \exp(-i\kappa a)}{2i} = \frac{\exp(-ka) - \exp(+ka)}{2i} = -\frac{1}{i} sh(ka) = i\ sh(ka)$ and therefore $\sin^2 \kappa a = -sh^2(ka)$.

[26] The de l'Hospital rule gives a series of transformations $\frac{1}{(\beta-1)} \sin^2 \kappa a = \lim_{E \to V_0} \frac{\sin^2\left(\sqrt{\frac{2m(E-V_0)}{\hbar^2}}a\right)}{\frac{E}{V_0} - 1} =$
$\lim_{E \to V_0} \frac{2 \sin\left(\sqrt{\frac{2m(E-V_0)}{\hbar^2}}a\right) \frac{\frac{2m}{\hbar^2}}{2\sqrt{\frac{2m(E-V_0)}{\hbar^2}}a} a^2}{\frac{1}{V_0}} = \frac{2ma^2}{\hbar^2} V_0.$

What is called common sense tells us that if $E > V_0$, the particle will pass the barrier for sure: we say it will go over the barrier. We will postpone a bit checking out the value of common sense. For the time being let us consider the cases for which the energy is smaller than the barrier, i.e., $E < V_0$. If our point-like particle behaved according to classical mechanics it would not pass the barrier. It is like a car having velocity[27] v and mass m (therefore its kinetic energy equals $E = \frac{mv^2}{2}$) that cannot pass a hill of height h if $E < V_0$ with $V_0 = mgh$ (g is the gravitational acceleration). Let us see how it will look in quantum mechanics. From Eq. (4.14) it can be seen that for any $E < V_0$ we have $D > 0$ except if[28] $E = 0$. This means that the particle can pass through a wall as if the car, in our analogy, passed through a tunnel made in the hill. This intriguing phenomenon is known as *tunneling*.[29]

Let us see what the effectiveness of the tunneling (D) depends on. Fig. 4.8 and Eq. (4.14) show that

for tunneling:

- the larger the impact energy of the particle, the larger D;
- a higher barrier (V_0) or a wider barrier (larger a) decrease D;
- tunneling of a lighter particle is easier.

All these changes are monotonic. This means there are no discrete “magic” values of the impact energy or the barrier heights or widths, for which the tunneling would be easier.

What about the wave function for a tunneling particle ($E < V_0$)? The answer can be found in Fig. 4.9.

As one can see:

- The real and imaginary parts of the wave function are not equal to zero along the barrier, i.e., the particle will appear there!
- The real and imaginary parts of the wave function vanish exponentially along the barrier.

---

[27] Here we assume the engine is switched off.

[28] In such a case in Eq. (4.14) $\beta = 0$, with as a consequence $D = 0$.

[29] The wonder I am trying to convince you about will however fade a bit. What kinetic energy is really necessary for an object to be transported over the hill? Should it always be equal to or greater than mgh? Well, imagine our car divided into a series of pieces, say a 1000 of them, tied one to the next one by a thin thread of a certain length. What about transporting the car now? Well, we will pay an energy to transport the first piece to the top of the barrier, say $\frac{mgh}{1000}$, but the second piece would be then pulled by the first one descending behind the hill! The same happens for other pieces. We see therefore that, when the object to transport is not point-like, but instead has some dimension in space, the energy cost of passing the barrier may be smaller! A quantum particle is described by a wave function extended in space and therefore its passing the barrier is possible even if the particle itself has insufficient kinetic energy. This is what we observe in the tunneling effect.

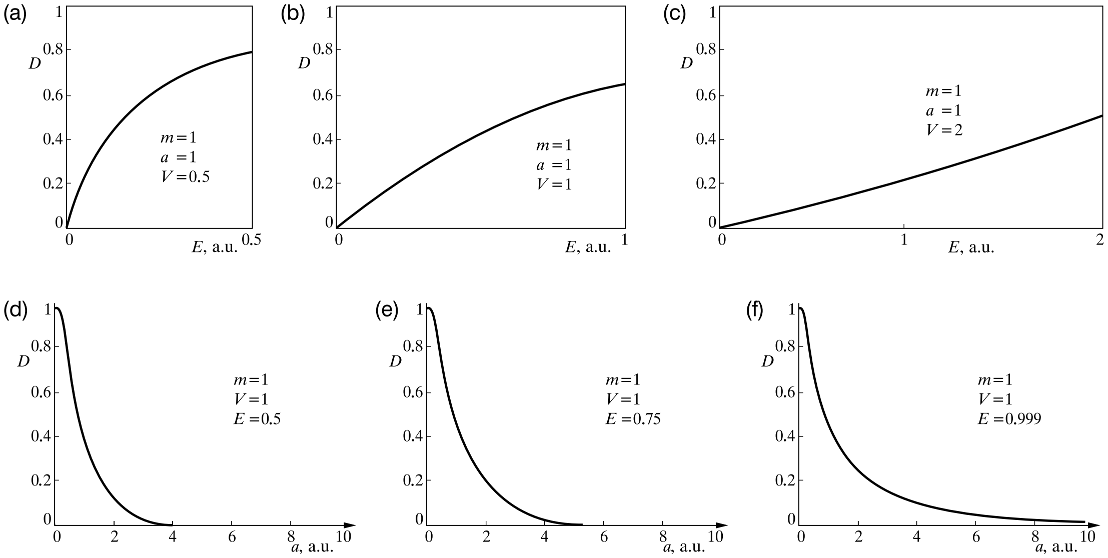


Fig. 4.8. A classical particle cannot tunnel through a barrier, while a quantum particle *can*. The figures show the transmission coefficient (tunneling) of the electron having various energies (always lower than the barrier) and passing through a barrier of various heights and widths. (a)–(c) The larger the energy, the easier to tunnel, and the higher the barrier, the harder to pass the barrier (at the same energy of the particle). (d)–(f) The dependence of the transmission coefficient on the barrier width: the wider the barrier, the harder it is to go through.

- The latter means that since the barrier has a finite width, the wave function will not vanish at the exit from the barrier region. Then, after going out of the barrier region, the wave function again begins to oscillate having *the same wave length, but a smaller amplitude*! This means that the particle passes the barrier with a certain probability D and reflects from the barrier with probability $1 - D$.

4.5.2 Surprises for energies larger than the barrier

Now let us see what will happen if the impact energy E is larger than the barrier height V_0. The transmission coefficient is given by the general Eq. (4.13).

Fig. 4.10 shows the transmission coefficient D as a function of $\frac{E}{V_0}$. We have a surprise, clearly visible from Eq. (4.13). *We expected that for* $\frac{E}{V_0} > 1$ *the particle would always pass the barrier*, while because of $\sin \kappa a$ this will happen only rarely, for impact energies E with $\kappa = \sqrt{\frac{2m(E-V_0)}{\hbar^2}}$ satisfying $\kappa a = n\pi$, where $n = 1, 2, \ldots$. For other values of κ the particle has only a certain chance to pass (Fig. 4.10). This result is a bit puzzling. What is special in these magic values

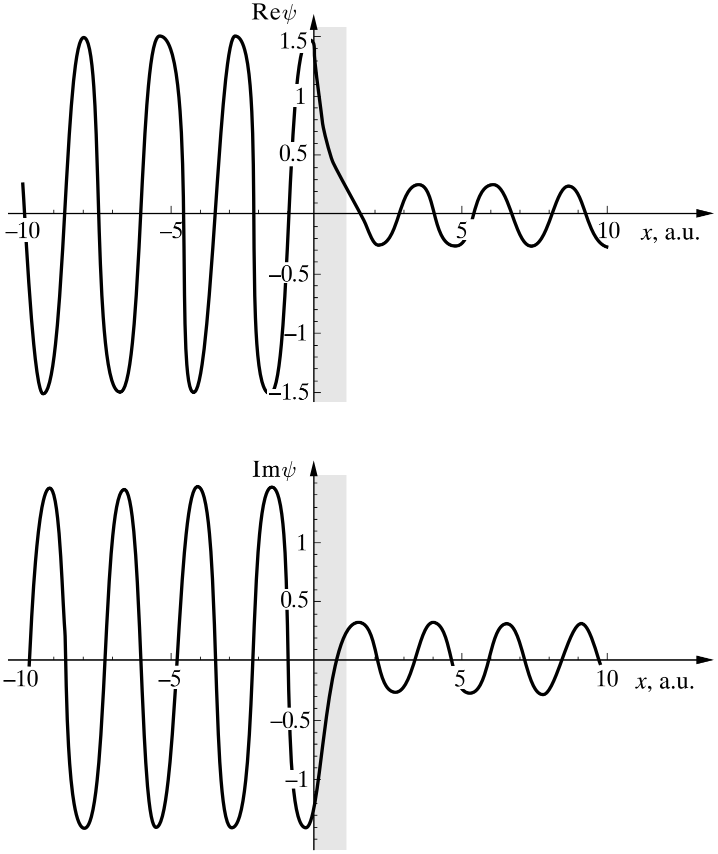


Fig. 4.9. Tunneling of an electron ($m = 1$) with energy $E = 2.979$ a.u. through a single barrier of height 5 a.u. and width 1 a.u. The wave function plot (real and imaginary parts) corresponds to the following values of the coefficients: $A_1 = 1$ (as a reference), $B_1 = 0.179 - 0.949i$, $A_2 = 1.166 - 0.973i$, $B_2 = 0.013 + 0.024i$, $A_3 = -0.163 - 0.200i$, and it represents a wave.

of κ, for which the particle does not see the barrier? For the time being we do not know, but, wait a minute, we see something interesting. For these magic impact energies we have $\frac{2m(E-V_0)}{\hbar^2}a^2 = \left(n^2\pi^2\right)$ or

$$E \equiv E_n = V_0 + \frac{n^2h^2}{8ma^2}, \tag{4.15}$$

but this means energy levels of the particle in a box of the length equal to the barrier width and placed exactly on top of the barrier! What!? The barrier and the box of the same length represent extremes in their character... But, in spite of that we have the miracle coming true! It looks as if the de Broglie wave describing the particle should fit a half-integer number of its

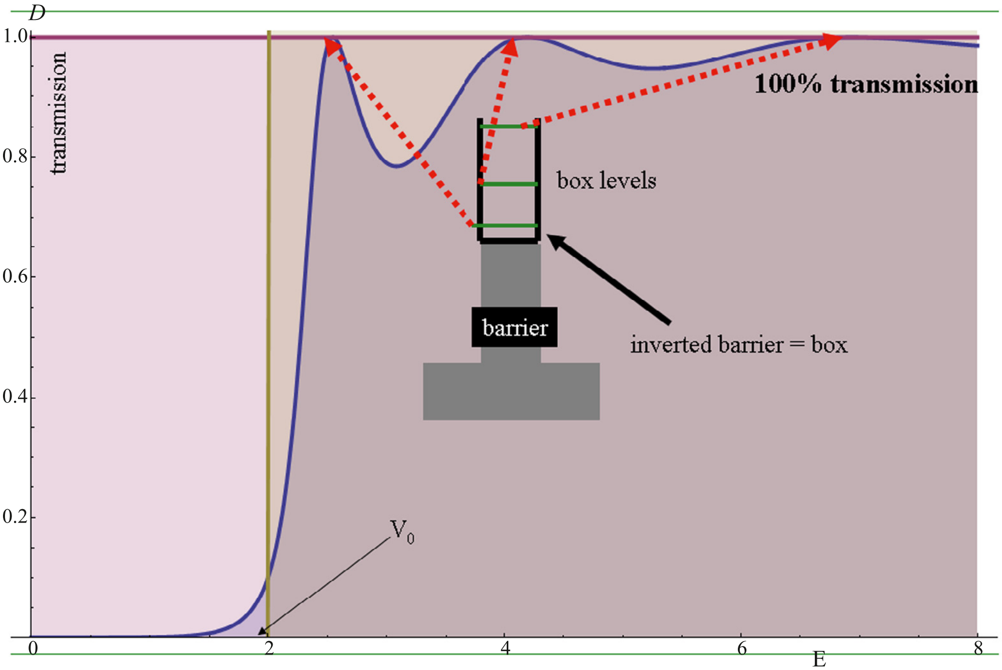


Fig. 4.10. Probability (D) of passing the barrier by a particle (its mass is 1 a.u. and is equal to the mass of an electron; the barrier height $V_0 = 2$ a.u. and width $a = 3$ a.u.) as a function of the impact energy (E). For $E < V_0$ one has the tunneling probability D increasing with E. The vertical line corresponds to $E = V_0$; for a particle with this energy the probability of passing the barrier is only about 0.1. Then, for $E > V_0$, i.e., the impact energy being larger than the barrier height, the probability of reflection is in general large. However, if the impact energy matches a particle-in-a-box eigenvalue of energy the transmission coefficient increases to 100%!.

wave lengths within the region of perturbation (either the energy barrier or the energy well). Only then the wave does not lose its rhythm, a fact important for constructive interference, and only then the particle passes over the barrier as if without seeing it. It is interesting that to get this effect one has to have $E > V_0$. There was no such effect for tunneling (for $E < V_0$), the interference was impossible due to the exponential decay of the wave function instead of its oscillatory behavior.

One more intriguing problem. Eq. (4.15) tells us that if the impact energy increases, the magic-like 100% passages will be rarer and rarer. We might expect a reverse behavior, because for very large E the particle should not pay attention to the barrier and pass without seeing it. Fig. 4.10 makes the situation clear: the 100% passages indeed will be rarer and rarer with increasing E, but it will be easier and easier to pass for impact energies which *do not* satisfy $\kappa a = n\pi$. For large values of the impact energies the particle practically will not see the barrier (in agreement with common sense).

4.6 The magic of two barriers

If we take *two* rectangular barriers of height V_0 with a well between them (Fig. 4.2c), then we will also witness magic. We allow for any energy of the particle ($E > 0$).

How will the problem be solved?

We have five nonoverlapping sections of the x axis. In each section the wave function will be assumed in the form $\Psi(x) = Ae^{i\kappa x} + Be^{-i\kappa x}$ with the same A and B coefficients and with the corresponding values of $\kappa^2 = \frac{2m(E-V)}{\hbar^2}$. In section 4, however, the particle goes right and never left, hence $B_5 = 0$. Now, the other coefficients A and B will be determined by stitching the wave function nicely at each of the four boundaries in order to have it going smoothly through the boundary (the wave function values and the first derivative values have to be equal for the left and right section to meet at this boundary). In this way we obtain a set of eight linear equations with eight unknown ratios: $\frac{A_i}{A_1}, i = 2, 3, 4, 5$, and $\frac{B_i}{A_1}, i = 1, 2, 3, 4$. The most interesting ratio is A_5/A_1, because this coefficient determines the transmission coefficient through the two barriers. Using the program Mathematica,[30] we obtain an amazing result.

Transmission coefficient

Let us check how the transmission coefficient (in our case identical to the transmission probability) changes through two identical barriers of height $V_0 = 5$ (all quantities in a.u.), each of width $a = 1$, when increasing the impact energy E from 0 to $V_0 = 5$. In general the transmission coefficient is very small. For example, for $E = 2$ the transmission coefficient through the *single* barrier (D_{single}) amounts to 0.028, that is, the chance of transmission is about 3%, while the transmission coefficient through the double barrier (D_{double}) is equal to 0.00021, i.e., about 100 times smaller. It stands to reason, it is encouraging. It is fine that it is harder to cross two barriers than a single barrier.[31] And the story will certainly be repeated for other values of E. To be sure, let us scan the whole range $0 \leq E < V_0$. The result is shown in Fig. 4.11 and represents a big surprise!

---

[30] See the web annex, the file Mathematica\Dwiebar.ma.

[31] This is even more encouraging for a prison governor; of course, a double wall is better than a single one!

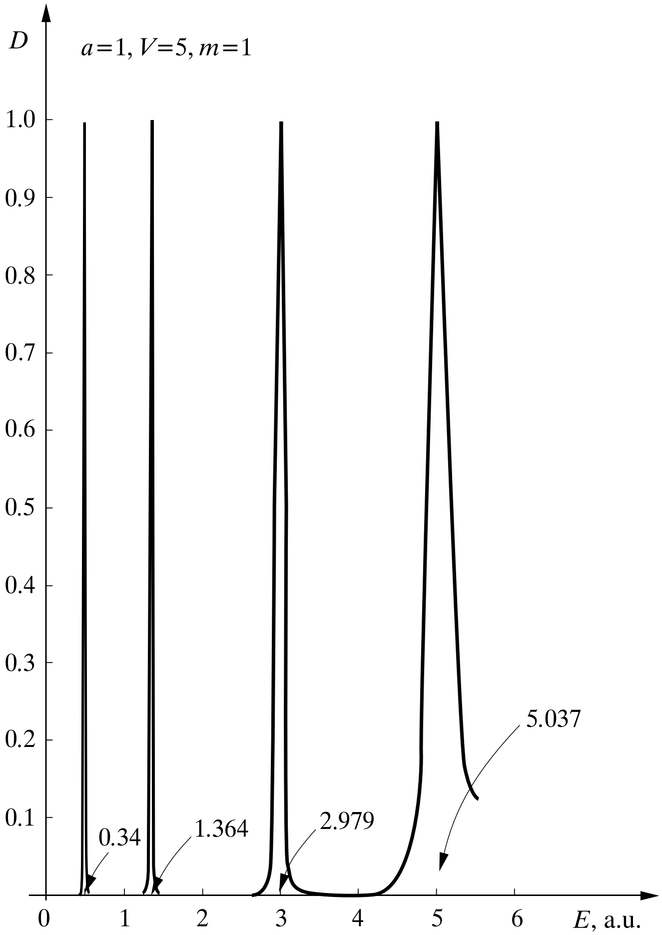


Fig. 4.11. The transmission coefficient (D) for a particle going through a potential double barrier (of height 5 a.u.) as a function of the particle impact energy E. We see some sudden increases of the transmission coefficient (resonance states).

4.6.1 Magic energetic gates (resonance states)

There is something really exciting going on. In our case we have three energies $E \leq V_0$, at which the transmission coefficient D_{double} increases dramatically in a narrow energy range. These energies are 0.34, 1.364, and 2.979. Thus, there are three secret energetic gates for going through the double barrier. It is sufficient just to hit the right energy (resonance energy). Is the chance of passing two barriers large? Let us take a look. For all three resonances the transmission coefficient is equal to $D_{\text{double}} = 1$, it only drops down differently when going off-resonance! Thus, there are three particle energies, for which the particle goes through the two barriers like a ghost through walls, completely undisturbed, as if the barriers did not exist[32]! Moreover, as we can see for the third resonance, the transmission coefficient through the single barrier amounts to $D_{\text{single}} = 0.0669$ (i.e., only 7%), but through two barriers 100%! It looks as if it would be hard for a prisoner to pass through a single armored prison door, but when the anxious prison governor made a second armored door behind the first, the prisoner (trained in quantum mechanics) disappeared through the two doors like a ghost.[33] Visibly at the resonances the back-going waves reflected from the four borders just added up to nothing, thus producing their perfectly destructive interference. At the same particular resonance energy the forward-going waves produced a perfectly constructive interference.

But why? What happens over there? Let us stress once more that the phenomenon is 100 % of a quantum nature, because a classical particle would tunnel neither through the double nor through the single barrier. Why do we observe such dramatic changes in the transmission coefficient for the two barriers? *We may have some suspicions.* From the time the second barrier is created, a new situation appears: a well *between* the two barriers, something similar to the box discussed earlier.[34] A particle in a box has some peculiar energy values: the energies of the stationary states (cf. p. 190). In our situation all these states correspond to a continuum, but something magic might happen if the particle had just one of these energies. Let us calculate the stationary state energies *assuming* that $V_0 = \infty$. Using atomic units in the energy formula, we have $E_n = \frac{h^2}{8m}\frac{n^2}{L^2} = \frac{\pi^2}{L^2}\frac{n^2}{2}$. To simplify the formula even more let us take $L = \pi$. Finally, we have simply $E_n = \frac{n^2}{2}$. Hence, we might expect something strange for the energy E equal to $E_1 = \frac{1}{2}$, $E_2 = 2$, $E_3 = \frac{9}{2}$, $E_4 = 8$ a.u., etc. (the last energy level, $E_4 = 8$, is already higher than the barrier height). Note, however, that the resonance states obtained appear at quite different energies: 0.34, 1.364, 2.979. Our intuitive hypothesis was wrong...

Maybe this intuition nevertheless contains a grain of truth? Let us concentrate on E_1, E_2, E_3, E_4. One may expect that the wave functions corresponding to these energies are similar to the ground state (nodeless), the first (single node) excited state, and the second (two nodes) excited state of the particle in a box. What does then happen to the nodes of the wave function for the particle going through two barriers? Here are the plots for the off-resonance (Fig. 4.12) and resonance (of the highest energy, Fig. 4.13) cases.

These figures and similar figures for lower-energy resonances seem to support our hypothesis: if an integer number of the half-waves of the wave function fit the region of the "box" between the

---

[32] This news should be strictly confidential in penitentiary departments. This phenomenon has been described first by David Bohm, *Quantum Theory*, Prentice-Hall, New York, 1951.

[33] There is experimental evidence for such resonance tunneling through two energy barriers in semiconductors. One of the first reports on this topic was a paper by T.C.L.G. Soliner, W.D. Goodhue, P.E. Tannenwald, C.D. Parker, D.D. Peck, *Appl. Phys. Letters*, 43(1983)588.

[34] Note, however, that the box has finite well depth and final width of the walls.

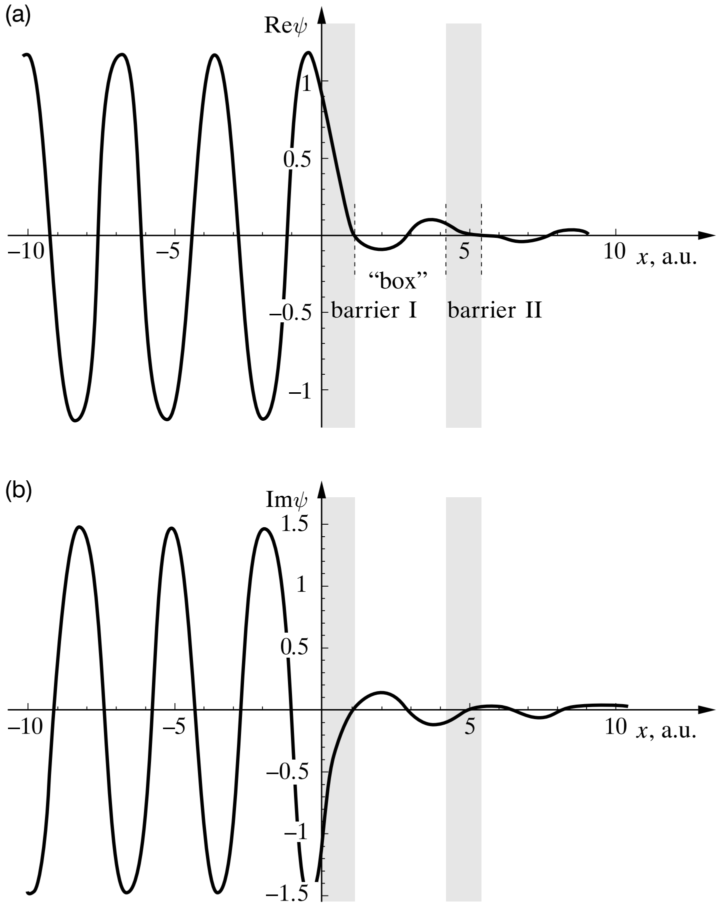


Fig. 4.12. Off-resonance case below the barriers' height. Tunneling of an electron with energy $E = 2$ a.u. through two barriers of height $V_0 = 5$ and width $a = 1$, the barrier separation is $L = \pi$ (all quantities in a.u.). (a) The real part of the wave function oscillates before the first barrier, is reduced by an order of magnitude in the first barrier, oscillates for ca. one period between the barriers, decays in the second barrier, and leaves the barrier region with an amplitude representing about 5% of the starting amplitude. (b) A similar picture follows from the imaginary part of the wave function.

barriers ("barrier-box-barrier"[35]), then we may expect resonance – a secret gate to go through

---

[35] We see one more time that a quantum particle is a kind of magic object. Even its passing over a single barrier revealed already some unexpected difficulties. Now we see some strange things happen for two barriers. It looks as if the particle's passing (through a single or a double barrier) goes smoothly, under the condition that the particle does not "lose its rhythm" in the barrier region. This "not losing the rhythm" reflects its wave nature and means fitting an integer number of its de Broglie half-waves within the obstacle region.

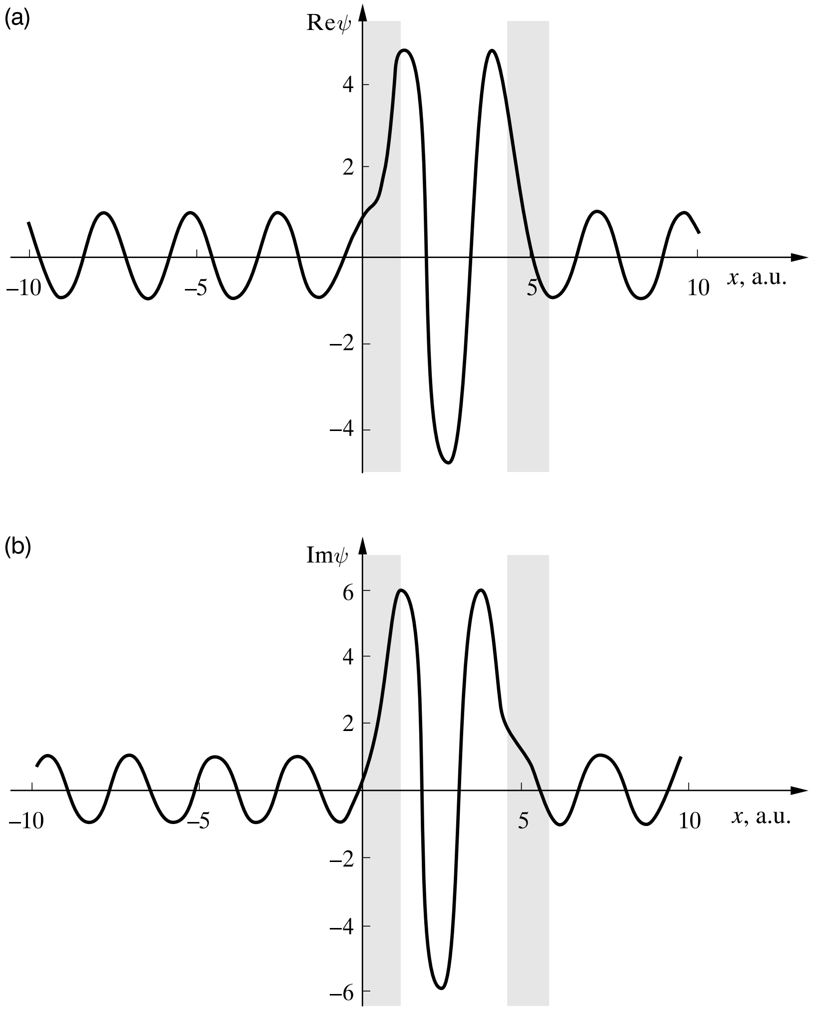


Fig. 4.13. The resonance case below the barriers' height. Tunneling of an electron with energy $E = 2.979$ a.u. through two barriers of height $V_0 = 5$ and width $a = 1$, the barrier separation is $L = \pi$ (all quantities in a.u.). (a) The real part of the wave function oscillates before the first barrier with amplitude 1, increases by a factor of about 3.5 within the first barrier, makes slightly more than about one period between the barriers, decays in the second barrier, and leaves the barrier region with an amplitude representing about 100% of the starting amplitude. (b) A similar picture follows from the imaginary part of the wave function.

the barriers.[36] As we can see, indeed we have been quite close to guessing the reason for the

---

[36] As one can see in this case, contrary to what happened with a single barrier, the wave function does not vanish exponentially within the barriers.

resonances. However, it turned out that the box length should include not only the box itself but also the barrier widths. Maybe to obtain the right resonance energies we simply have to adjust the box length? Since, instead of resonance at $E_1 = \frac{1}{2}$, we have resonance at energy 0.34, we may guess that it is sufficient to change the box width L to $L' = \sqrt{\frac{0.5}{0.34}}L = 1.21L$, to make the first resonance energies match. Then, instead of $E_1 = \frac{1}{2}$, we have exactly the first resonance energy equal to $E_1' = 0.34$, an agreement forced by us, but later, instead of $E_2 = 2$ we obtain $E_2' = 1.36$, and this agrees very well with the second resonance energy 1.364! Then, instead of $E_3 = 4.5$, we obtain $E_3' = 3.06$, a good approximation to 2.979, but evidently the closer the barrier energy the harder it is to obtain agreement.[37] The next resonance state is expected to occur at $E_4' = 8 \times 0.68 = 5.44$, but we have forgotten that this energy already exceeds the barrier height ($V_0 = 5$ a.u.). We will come back to this state in a moment.

4.6.2 Strange flight over two barriers

Let us consider the two barriers and an electron with *higher energy than the barrier* height V_0. What will happen? Well, we might say naively that this means the particle energy is sufficient to pass the barrier. Note, however, that we had strange phenomena to happen even for a single barrier.

Let us assume the barrier height $V_0 = 5$ and the particle energy is equal to 5.5 a.u. We solve our equations and we obtain a transmission coefficient equal to 0.138, hence the electron will bounce back with a probability of about 86%. How it did bounce off? Difficult to say.

Fig. 4.11 shows the transmission coefficient also for energies higher than the barrier height. It turns out that at energy $E = 5.037$ a.u. (i.e., higher than the barrier height) another resonance state is hidden, which ensures almost 100% certainty of transmission (whereas the particle energies in the energetic neighborhood lead to a considerable reflection rate as described above). We expected naively such behavior for all $E > V_0$, but it turned out to be true for the resonance state. Let us recall that we have already predicted "by mistake" a box stationary state with energy $E_4' = 5.44$, higher than the barrier height V_0. This, together with the number of the nodes within the barrier range seen in Fig. 4.14, tells us that indeed this is the state.[38]

---

[37] Note, please, that there is such a thing as resonance width, and that this width is different for each resonance. The most narrow resonance corresponds to the lowest energy, the widest to the highest energy. The width of resonances is related to the notion of the resonance lifetime τ (τ is proportional to the inverse of the resonance width).

[38] It corresponds to a lower energy than we predicted (similar to the case of E_3). No wonder that due to the finite well depth, the states corresponding to the upper part of the well "feel" the box is longer.

What makes the difference between the resonance and off-resonance states for $E > V_0$? The corresponding wave functions (real and imaginary parts) are given in Figs. 4.14 and 4.15.

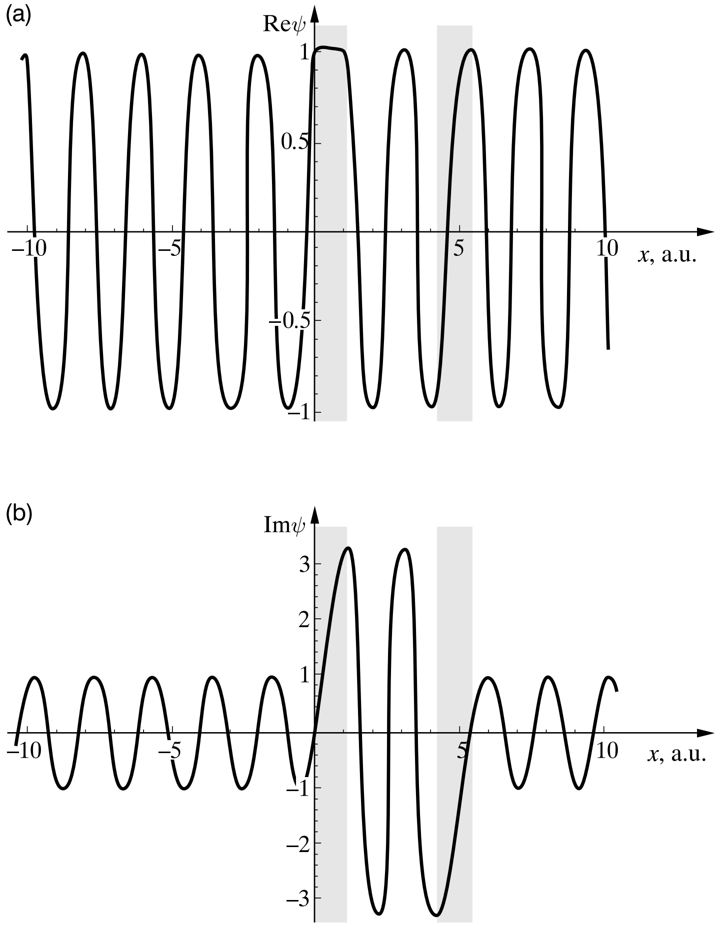


Fig. 4.14. Resonance case over the barriers. The wave function for an electron with energy $E = 5.037$ a.u., i.e., *over the barrier height* $V_0 = 5$, is shown. As we can see the amplitude is nearly the same for the wave function before and after the barriers (this means the transmission coefficient of the order of 100%). The real part and especially the imaginary part both wobble within the domain of the barriers, i.e., within section $(0, 5.14)$ (the imaginary part has a large amplitude). We may guess that the state is related to the three-node stationary state.

Thus, resonance states may also hide in that part of the continuum which has energy higher than the barriers (with a short lifetime, because such resonances are wide, cf. Fig. 4.11). They are also a reminder of the stationary states of the particle in a box longer than the separation of the barriers and infinite well depth.

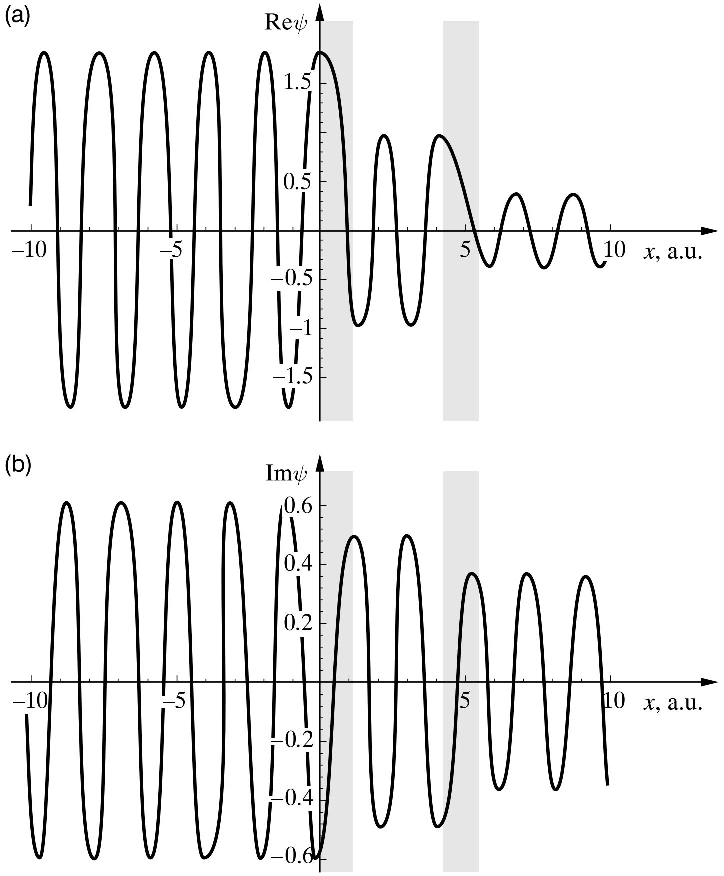


Fig. 4.15. The off-resonance case over the barriers. The wave function for an electron with energy $E = 5.5$ a.u., i.e., *over the barrier height* $V_0 = 5$, is shown. Despite the fact that $E > V_0$, the amplitude of the outgoing wave is considerably reduced after passing the range of the barriers $(0, 5.14)$. This means that the particle flying over the barriers may reflect from them.

4.7 Harmonic oscillator

A one-dimensional harmonic oscillator is a particle of mass m, subject to the force $-kx$, where the force constant $k > 0$, and x is the deviation of the particle from its equilibrium position[39]

---

[39] A harmonic oscillator represents a single particle. Sometimes, however, we mean by it *two* point-like particles bound together by a spring, having a certain equilibrium length, and when the length deviates from the equilibrium the potential energy is proportional to the square of the deviation. This two-particle model looks much more attractive to us, since diatomic molecules, like H_2, HCl, etc., resemble it. As a consequence one might apply to diatomics conclusions from solution of the Schrödinger equation for the harmonic oscillator. The diatomics will be elaborated in detail in Chapter 6. It will turn out there that a relative motion of two atoms can indeed

($x = 0$). The potential energy is given as a parabola $V = \frac{1}{2}kx^2$ (this property is known as harmonicity).

> Despite the fact that there is no such a thing as a perfect harmonicity of oscillations, this notion plays in physics an outstanding role, probably the most important and most fruitful among all models considered! The reason is that our Universe is full of vibrations of all kinds.

This is why the present section is of prime importance.

New variable and the transformed Hamiltonian

First, let us write down the harmonic oscillator Hamiltonian:

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2.$$

Now, let us introduce a useful scaled coordinate: $q = \sqrt[4]{\frac{km}{\hbar^2}}x$, with $-\infty < q < \infty$. The Hamiltonian written using the new variable reads very interestingly as

$$\hat{H} = -\frac{\hbar^2}{2m}\left(\frac{dq}{dx}\frac{d}{dq}\right)^2 + \frac{1}{2}k\frac{1}{\sqrt{\frac{km}{\hbar^2}}}q^2 = -\frac{1}{2}\hbar\sqrt{\frac{k}{m}}\frac{d^2}{dq^2} + \frac{1}{2}\hbar\sqrt{\frac{k}{m}}q^2 = \hbar\omega\left[-\frac{1}{2}\frac{d^2}{dq^2} + \frac{1}{2}q^2\right]. \tag{4.16}$$

Creation and annihilation operators

Now let us prepare two tools, which are the operators: $\hat{B} = \frac{1}{\sqrt{2}}(\frac{d}{dq} + q)$, known as a creation operator, and $\hat{b} = \frac{1}{\sqrt{2}}(-\frac{d}{dq} + q)$, an annihilation operator (their names will become clear in a minute). It turns out[40] that $\hat{b} = \hat{B}^\dagger$. As we will see in a moment these operators will create

---

be reduced (in a certain approximation) to a harmonic motion of a single particle, however under the condition that its deviation from the equilibrium position cannot be lower than $-R_0$, where R_0 stands for the equilibrium length of the diatomic. The reason is very simple: even if the spring connecting the two atoms were an ideal one, the harmonicity can pertain to extending the spring, but certainly not to squeezing it. When squeezing by $-R_0$ the two atoms seat one on the other and this is the end of squeezing at all. This means one cannot have a harmonic molecule for some basic reasons.

[40] Indeed, for f and g being arbitrary functions of class Q (please recall that $i\frac{d}{dq}$ represents a Hermitian operator), $\left\langle \hat{b}f|g\right\rangle = \frac{1}{\sqrt{2}}\left\langle (-\frac{d}{dq} + q)f|g\right\rangle = \frac{1}{i}\frac{1}{\sqrt{2}}\left\langle i\frac{d}{dq}f|g\right\rangle + \frac{1}{\sqrt{2}}\langle qf|g\rangle = \frac{1}{i}\frac{1}{\sqrt{2}}\left\langle f|i\frac{d}{dq}g\right\rangle + \frac{1}{\sqrt{2}}\langle qf|g\rangle = \frac{1}{\sqrt{2}}\left\langle f|\frac{d}{dq}g\right\rangle + \frac{1}{\sqrt{2}}\langle qf|g\rangle = \frac{1}{\sqrt{2}}\left\langle f|(\frac{d}{dq} + q)g\right\rangle = \left\langle f|\hat{B}g\right\rangle$, which means $\hat{b} = \hat{B}^\dagger$.

or annihilate the energy quanta of our system, which will be called *phonons*, each of energy $h\nu = \hbar\omega$.

Let us see what the operator $\hat{B}^\dagger \hat{B}$ does. For an arbitrary function[41] f we have $\hat{B}^\dagger \hat{B} f = \left(\frac{1}{\sqrt{2}}(-\frac{d}{dq} + q)\right)\left(\frac{1}{\sqrt{2}}(\frac{d}{dq} + q)\right) f = -\frac{1}{2}\frac{d^2 f}{dq^2} - \frac{1}{2}\frac{d}{dq}(qf) + \frac{1}{2}q\frac{df}{dq} + \frac{1}{2}q^2 f = -\frac{1}{2}\frac{d^2 f}{dq^2} - \frac{1}{2}(f + q\frac{df}{dq}) + \frac{1}{2}q\frac{df}{dq} + \frac{1}{2}q^2 f = \left(-\frac{1}{2}\frac{d^2}{dq^2} + \frac{1}{2}q^2 - \frac{1}{2}\right) f$. Hence, we recognize immediately that the Hamiltonian has the form

$$\hat{H} = h\nu(\hat{B}^\dagger \hat{B} + \frac{1}{2}). \tag{4.17}$$

Eigenfunctions and eigenvalues

Now we will find the eigenfunctions (4.22) and eigenvalues (4.21) of the Hamiltonian $\hat{H}$. To this end we note first that[42]

$$[\hat{B}, \hat{B}^\dagger] = 1 \quad \text{or} \quad \hat{B}\hat{B}^\dagger = \hat{B}^\dagger \hat{B} + 1. \tag{4.18}$$

Let us consider the following function:

$$\psi_v = (v!)^{-\frac{1}{2}} (\hat{B}^\dagger)^v \psi_0, \tag{4.19}$$

where ψ_0 stands for the normalized Gaussian function[43]: $\psi_0(q) = \frac{1}{\sqrt[4]{\pi}} \exp(-\frac{1}{2}q^2)$. Let us note first that $\hat{B}\psi_0 = \left(\frac{1}{\sqrt{2}}(\frac{d}{dq} + q)\right) \frac{1}{\sqrt[4]{\pi}} \exp(-\frac{1}{2}q^2) = \frac{1}{\sqrt[4]{\pi}} \left(\frac{1}{\sqrt{2}}(-q + q)\right) \exp(-\frac{1}{2}q^2) = 0$. This re-

---

[41] When transforming operators one has to remember that they act on a function (this is why we consider such a function explicitly), otherwise it is easy to make a mistake.

[42] We have

$$\hat{B}\hat{B}^\dagger f = \left(\frac{1}{\sqrt{2}}(\frac{d}{dq} + q)\right)\left(\frac{1}{\sqrt{2}}(-\frac{d}{dq} + q)\right) f =$$
$$-\frac{1}{2}\frac{d^2 f}{dq^2} + \frac{1}{2}\frac{d}{dq}(qf) - \frac{1}{2}q\frac{df}{dq} + \frac{1}{2}q^2 f =$$
$$-\frac{1}{2}\frac{d^2 f}{dq^2} + \frac{1}{2}(f + q\frac{df}{dq}) - \frac{1}{2}q\frac{df}{dq} + \frac{1}{2}q^2 f = \left(-\frac{1}{2}\frac{d^2}{dq^2} + \frac{1}{2}q^2 + \frac{1}{2}\right) f,$$

and therefore using Eq. (4.17) we get the commutation relation needed.

[43] We briefly check the normalization: $\int_{-\infty}^{\infty} |\psi_0(q)|^2 dq = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-q^2) dq = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-q^2) dq = \frac{1}{\sqrt{\pi}} \sqrt{\pi} = 1$.

sult will be used in a minute. We have also[44]

$$\hat{B}^{\dagger}\hat{B}\psi_v = v\psi_v. \tag{4.20}$$

Now, please recall that what we have in the Hamiltonian (4.17) is just $\hat{B}^{\dagger}\hat{B}$. Therefore, function ψ_v given by Eq. (4.19) for $v = 0, 1, 2, \ldots$ represents an eigenfunction of the Hamiltonian: $\hat{H}\psi_v = h\nu(\hat{B}^{\dagger}\hat{B} + \frac{1}{2})\psi_v = E_v\psi_v$, with the energy

$$E_v = h\nu(v + 1/2), \tag{4.21}$$

with $v = 0, 1, 2, \ldots$. Note that the oscillator energy is never equal to zero and is additive, i.e., the phonons, each of energy $\hbar\omega$, do not interact. The smaller the oscillating mass or the larger the force constant, the larger the energy of the phonon (Fig. 4.16).

The harmonic oscillator has an infinite number of energy levels (all of them nondegenerate; their separation is constant and is equal to $h\nu$).

The state ψ_0 describes the absence of phonons (phonon vacuum) and therefore an attempt of annihilation of the vacuum gives zero: $\hat{B}\psi_0 = 0$. The function $\psi_v = (v!)^{-\frac{1}{2}}(\hat{B}^{\dagger})^v\psi_0$, which is a general solution of the Schrödinger equation, *represents therefore a state with v noninteracting phonons, each of energy hν, created from the phonon vacuum* (by using v creation operators $\hat{B}^{\dagger}$).

---

[44] Let us see what we will get when applying the operator $\hat{B}^{\dagger}\hat{B}$ (and using the commutation relation (4.18)) to the wave function ψ_v:

$$\hat{B}^{\dagger}\hat{B}\psi_v = (v!)^{-\frac{1}{2}}\hat{B}^{\dagger}\hat{B}(\hat{B}^{\dagger})^v\psi_0 = (v!)^{-\frac{1}{2}}\hat{B}^{\dagger}\hat{B}\hat{B}^{\dagger}(\hat{B}^{\dagger})^{v-1}\psi_0 = (v!)^{-\frac{1}{2}}\hat{B}^{\dagger}\left(\hat{B}^{\dagger}\hat{B}+1\right)(\hat{B}^{\dagger})^{v-1}\psi_0 =$$
$$(v!)^{-\frac{1}{2}}(\hat{B}^{\dagger})^2\hat{B}(\hat{B}^{\dagger})^{v-1}\psi_0 + (v!)^{-\frac{1}{2}}\hat{B}^{\dagger}(\hat{B}^{\dagger})^{v-1}\psi_0 = (v!)^{-\frac{1}{2}}(\hat{B}^{\dagger})^2\hat{B}(\hat{B}^{\dagger})^{v-1}\psi_0 + \psi_v.$$

The commutation relation replaces $\hat{B}\hat{B}^{\dagger}$ by $\hat{B}^{\dagger}\hat{B} + 1$. The $\hat{B}^{\dagger}\hat{B}$ part made that in the sequence of v operators $\hat{B}^{\dagger}$ in ψ_v the operator $\hat{B}$ has been shifted right by one position (see the first term) and on its right-hand side it has only $v - 1$ of $\hat{B}^{\dagger}$ operators. *On the other hand, the presence of the unit operator made that we got the second term in the form* $1 \cdot \psi_v = \psi_v$. Now we will use once more the commutation relation. Also this time it will cause a one-position right shift of $\hat{B}$ resulting in $(v!)^{-\frac{1}{2}}(\hat{B}^{\dagger})^3\hat{B}(\hat{B}^{\dagger})^{v-2}\psi_0$ together with a *new additive term* ψ_v (altogether we have already $2\psi_v$). This procedure is repeated v times, up to the point when we get (note the first term turns out to be zero) $(v!)^{-\frac{1}{2}}(\hat{B}^{\dagger})^{v+1}\hat{B}\psi_0 + v\psi_v = 0 + v\psi_v = v\psi_v$, where we have used the relation $\hat{B}\psi_0 = 0$. Hence, $\hat{B}^{\dagger}\hat{B}\psi_v = v\psi_v$.

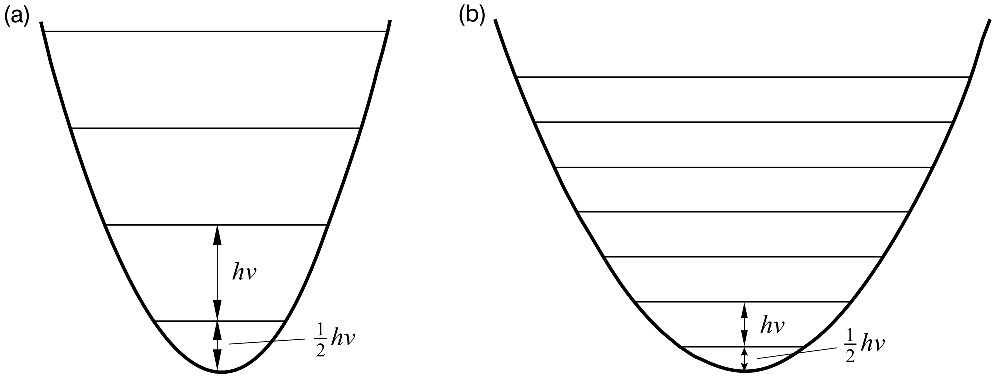


Fig. 4.16. A "finger print" of a harmonic oscillator – an infinite number of equidistant energy levels. Such figures represent in fact several independent figures one on top of the other: the parabola means the potential energy as a function of deviation of the oscillating particle from its equilibrium position, and the energy levels denote in principle some points on the axis of the total energy (traditionally and arbitrarily shown as horizontal sections). (a) Energy levels of an oscillator with a large force constant k. (b) Energy levels of an oscillator with a smaller force constant k.

The eigenfunctions of the Schrödinger equation can be also shown explicitly, i.e.,

$$\boxed{\Psi_v(q) = N_v H_v(q) \exp(-\frac{q^2}{2}), \qquad (4.22)}$$

where $v = 0, 1, 2, \ldots$ is the oscillation quantum number, H_v represent the Hermite polynomials[45] (of degree v) defined as

$$H_v(q) = (-1)^v \exp(q^2) \frac{d^v \exp\left(-q^2\right)}{dq^v},$$

and N_v is the normalization constant $N_v = \sqrt{\left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} \frac{1}{2^v v!}}$.

The first Hermite polynomials are

$$H_0(q) = (-1)^0 \exp(q^2) \frac{\mathrm{d}^0(\exp(-q^2))}{\mathrm{d}^0 q} = \exp(q^2)\exp(-q^2) = 1,$$

$$H_1(q) = (-1)^1 \exp(q^2) \frac{\mathrm{d}(\exp(-q^2))}{\mathrm{d}q} = -\exp(q^2)(-2q)\exp(-q^2) = 2q,$$

---

[45] Charles Hermite was a French mathematician (1822–1901), professor of Sorbonne. The Hermite polynomials were defined half a century earlier by Pierre Laplace.

$$H_2(q) = (-1)^2 \exp(q^2) \frac{d^2(\exp(-q^2))}{dq^2} = 4q^2 - 2, \ldots.$$

Fig. 4.17 shows what the wave functions for the one-dimensional harmonic oscillator look like. Fig. 4.18 gives the plots representing some selected wave functions for a *two-dimensional* harmonic oscillator.[46]

The harmonic oscillator is one of the most useful and beautiful models in physics. When almost nothing is known, except that the particles are held by forces, then the first model considered is the harmonic oscillator. This happened for the black body problem (Chapter 1), and now it is the case with the quantum dots,[47] string theory,[48] solvated electron,[49] etc. In addition, *the solutions resulting when assuming harmonicity of oscillations very often turn out to be analytical, compact, and simple functions – a really remarkable feature*.

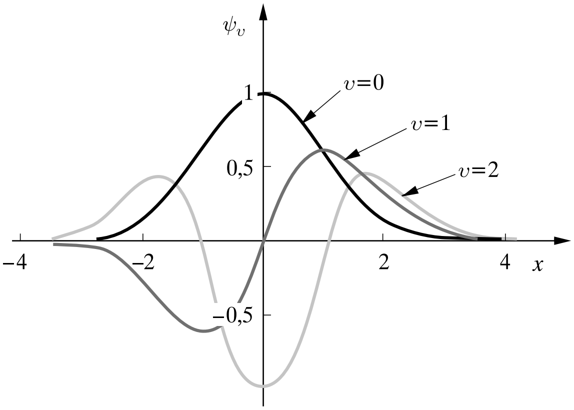


Fig. 4.17. Some of the wave functions Ψ_v for a one-dimensional oscillator. The number of nodes increases with the oscillation quantum number v. Note a striking similarity of the resulting wave functions to what we had for a particle in a box (Fig. 4.3b).

---

[46] One obtains the solution by a simple separation of variables; the wave function is a product of the two wave functions for the harmonic oscillators with x and y variables, respectively.

[47] A part of the "nanotechnology": on a solid surface some atomic clusters are placed (*quantum dots*), lines of such atoms (*nanowires*), etc. Such systems may exhibit unusual properties.

[48] Quarks interact through exchange of gluons. An attempt of separating two quarks leads to such a distortion of the gluon bond (string) that the string breaks down and separates into two strings with new quarks at their ends created from the distortion energy.

[49] Many polar molecules may lower their energy in a liquid by facing an extra electron with their positive poles of the dipole. This is the *solvated electron.*

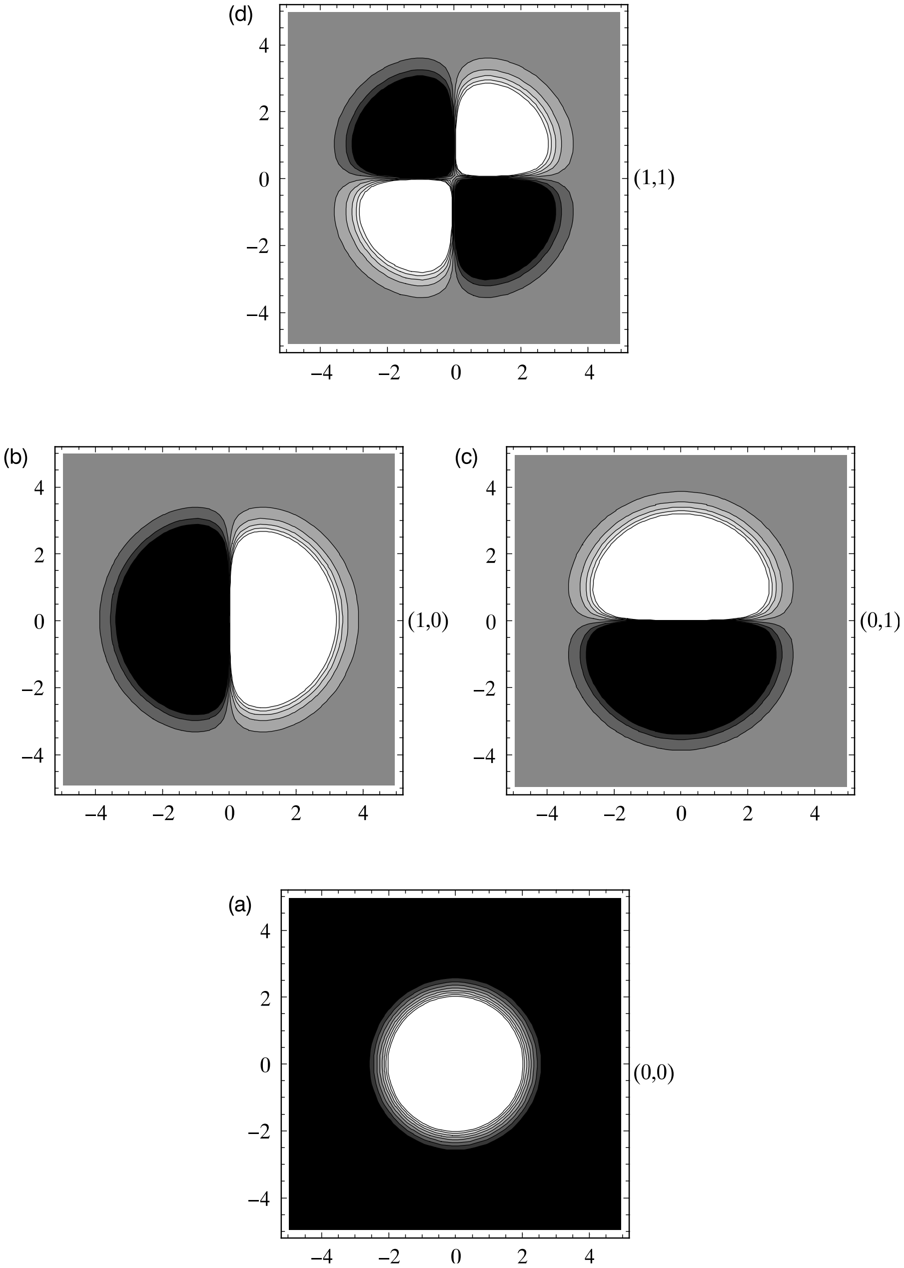


Fig. 4.18. A graphic representation of the two-dimensional harmonic oscillator wave function (isolines). (a)–(d) The wave functions labeled by a pair of oscillation quantum numbers (v_1, v_2) displayed on the right hand side of each picture. The higher the energy (upwards), the larger the number of node planes. A reader acquainted with the wave functions of the hydrogen atom will easily recognize a striking resemblance between these figures and the orbitals. Moreover, the two-dimensional oscillator wave functions are very similar to what one would have for drum vibrations.

4.8 Morse oscillator

Diatomic molecules differ from harmonic oscillators mainly in that they may dissociate. If we pull a diatomic molecule with internuclear distance R equal to the equilibrium distance R_e, then at the beginning, displacement $x = R - R_e$ is indeed proportional to the force applied, but afterwards the pulling becomes easier and easier. Finally, the molecule dissociates, i.e., we separate the two parts without any effort at all. This fundamental difference with respect to the harmonic oscillator is qualitatively captured by the potential proposed by Morse (parameter $\alpha > 0$)[50]:

$$V(x) = De^{-\alpha x}\left(e^{-\alpha x} - 2\right). \tag{4.23}$$

D represents the well depth, and the parameter α decides its width. When the displacement $x = 0$, then the function attains the minimum $V = -D$; when $x \to \infty$, then $V \to 0$ (Fig. 4.19).

Besides the abovementioned similarity, the Morse oscillator differs from real diatomics mainly by two qualitative features. First, for $R = 0$ we obtain a *finite* potential energy for the Morse oscillator; second, the asymptotic behavior of the Morse oscillator for $x \to \infty$ means exponential asymptotics, while the atomic and molecular systems at large distances interact as $\frac{1}{R^n}$.

The second derivative of $V(x)$ calculated at the minimum of the well represents the force constant k of the Morse oscillator

$$k = 2\alpha^2 D. \tag{4.24}$$

The parabola $-D + \frac{1}{2}kx^2$ best approximates $V(x)$ close to $x = 0$ and represents the harmonic oscillator potential energy (with the force constant k). The Morse oscillator is hard to squeeze – the potential energy goes up faster than that of the harmonic oscillator with the same force constant k.

Solution

One had to have courage to presume that analytical solution with such a potential energy exists. The solution was found by Morse. It represents a rare example of an exact solution to a nonlinear problem. Exact solutions exist not only for the ground (oscillation quantum number $v = 0$) but also for all the excited states ($v = 1, 2, \ldots v_{max}$) belonging to the discrete spectrum. The energy levels are nondegenerate and are given by the following formula:

---

[50] Philip McCord Morse (1903–1985) was an American theoretical physicist.

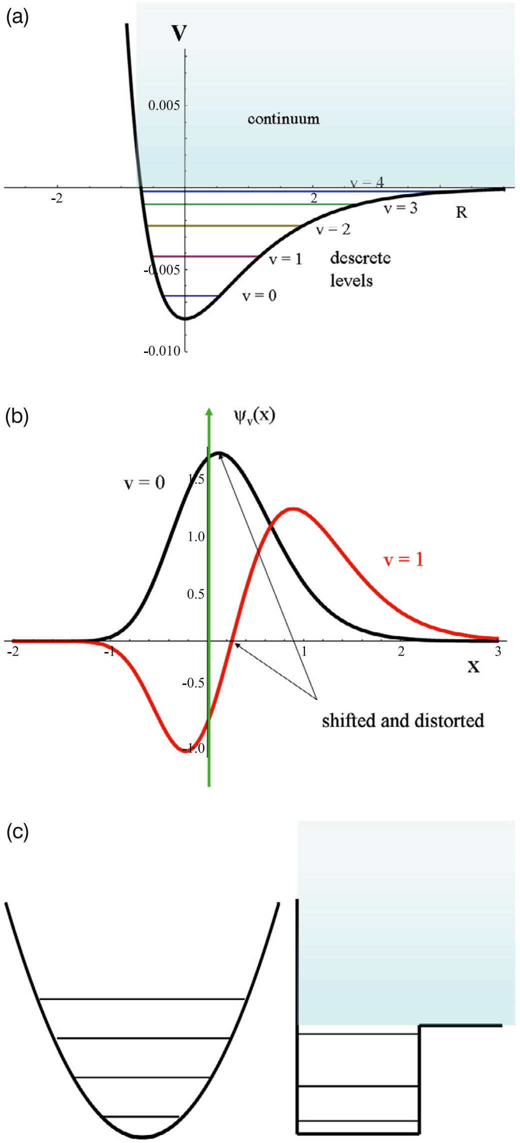


Fig. 4.19. (a) The Morse potential energy curves have the shape of a hook (D controls the well depth and α its width). This potential leads to some number of the discrete levels (on the picture $v = 0, 1, \ldots 4$) and a continuum of states above the dissociation limit. (b) The wave functions of the Morse oscillator (here for $v = 0$ and $v = 1$ states) resemble those of the harmonic oscillator, except that they are a bit shifted to the right and asymmetric with respect to the $x \to -x$ transformation. The Morse potential represents a kind of hybrid of the harmonic potential (c1) and a rectangular well (c2). Both potentials lead to exact solutions of the Schrödinger equation. Model c2 gives the discrete spectrum as well as the continuum and the resonance states. The latter ones are only very rarely considered for the Morse oscillator, but they play an important role in scattering phenomena (primarily in reactive collisions).

$$E_v = -D + h\nu(v + \frac{1}{2}) - h\nu(v + \frac{1}{2})^2\beta,$$
$$v = 0, 1, 2, ..., v_{max}, \tag{4.25}$$

where using atomic units we obtain

$$h\nu = 2\alpha\left(\frac{D}{2\mu}\right)^{\frac{1}{2}}. \tag{4.26}$$

This formula follows from the parabolic approximation of the Morse potential (valid for small displacements x),[51] while

$$\beta = \frac{\hbar\omega}{4D}, \tag{4.27}$$

where μ is the mass of the oscillating particle. When the Morse oscillator serves as a model of a diatomic molecule, μ stands for the reduced mass of both nuclei $\mu = (1/m_1 + 1/m_2)^{-1}$ (Appendix J on p. 691). As we can see, the energy of the oscillator never equals zero (similar to the harmonic oscillator) and

the separation between consecutive energy levels decreases.

The wave functions are slightly more complicated than those for the harmonic oscillator and are given by the following formula:

$$\psi_v(z) = N_v e^{-\frac{z}{2}} z^{b_v} L_v^{2b_v}(z), \tag{4.28}$$

where $v = 0, 1, ...v_{max}$, variable z is a real number related to displacement x by the formula $z = 2ae^{-\alpha x}$, and the normalization coefficient $N_v = \sqrt{\frac{2b_v v!}{\Gamma(2b_v+v+1)}}$ with $\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt$, while

---

[51] Let us recall that for the harmonic oscillator $2\pi\nu = \sqrt{\frac{k}{\mu}}$; therefore, from Eq. (4.24) $h\nu = \hbar\alpha\sqrt{\frac{2D}{\mu}}$, while $\hbar = 1$ a.u.

$$a = \frac{2D}{\hbar\omega}, \tag{4.29}$$

$$b_v = a - \frac{1}{2} - v > 0. \tag{4.30}$$

The above condition gives maximum $v = v_{max}$ and, therefore, $v_{max} + 1$ is the number of eigenfunctions. Thus, we always have a finite number of energy levels.

The term L stands for the polynomial given by the formula

$$L_n^c(z) = \frac{1}{n!} e^z z^{-c} \frac{d^n}{dz^n}\left(e^{-z} z^{n+c}\right), \tag{4.31}$$

where $n = 0, 1, 2, \ldots$ is the polynomial degree.[52] A short exercise gives

$$L_0^c(z) = 1,$$
$$L_1^c(z) = (c+1) - z,$$
$$L_2^c(z) = \frac{1}{2}z^2 - (c+2)z + \frac{1}{2}(c+1)(c+2),$$
$$\ldots$$

This means the number of nodes in a wave function is equal to v (as in the harmonic oscillator). *The wave functions resemble those of the harmonic oscillator, but are slightly deformed* (have a bit larger values for $x > 0$), Fig. 4.19b.

For very large well depths (D), the parameter β of Eq. (4.27) becomes very small. This results in E_v approaching the corresponding formula for the harmonic oscillator $-D + h\nu(v + 1/2)$, and the energy levels become equidistant with the nearest neighbor separation equal to $h\nu$. The potential is highly anharmonic (of the "hook type"), but the energy levels would be equidistant as in the harmonic oscillator. Is it possible? Yes, it is. The key is that, for small values of v, the term $-h\nu(v + 1/2)^2\beta$ does not yet enter into play and low energy levels correspond to small amplitudes (x) of vibrations. For small x, the potential is close to parabolic[53] – as for the harmonic oscillator with force constant k.

Example 2 (Hydrogen molecule). The hydrogen molecule has been investigated in detail. As will be seen in Chapters 6 and V2-2 the theory challenges there some very subtle experiments.

---

[52] Indeed, n-time derivation gives $e^{-x}x^{n+c}$ as a term with the highest power of x. Multiplication by $e^x x^{-c}$ gives x^n.

[53] As witnessed by a Taylor expansion of $V(x)$ for $x = 0$.

Let us approximate the most accurate theoretical potential energy curve[54] (as a function of the internuclear distance R) by a Morse curve.

Is such an approximation reasonable? Let us see. From Wolniewicz's calculations we may take the parameter $D = 109.52$ kcal/mol $= 38\,293$ cm^{-1}, while the parameter α is chosen in such a way as to reproduce the theoretical binding energy for $R = R_e + 0.4$ a.u.,[55] where $R_e = 1.4$ a.u. is the position of the minimum binding energy. It turns out that, say, "by chance" this corresponds to $\alpha = 1$. From Eqs. (4.29) and (4.30) we obtain $a = 17.917$ and the allowed v are those satisfying the inequality $b_v = 17.417 - v > 0$. We expect, therefore, 18 energy levels with $v = 0, 1, ..., 17$ for H_2 and 25 energy levels for T_2 (in the last case $b_v = 24.838 - v > 0$). Accurate calculations of Wolniewicz give 14 vibrational levels for H_2 and 25 levels for T_2. Thus, decreasing the reduced mass makes the vibrational levels less dense and some vibrational levels even disappear. This means that isotope substitution by a heavier isotope leads to stabilization. Moreover, from Eq. (4.26) we obtain for H_2 $h\nu = 0.019476$ a.u. $= 4274$ cm^{-1}, while from Eq. (4.27) we have $\beta = 0.0279$. From these data one may calculate the energetic gap between the ground ($v = 0$) and the first excited state ($v = 1$) for H_2, $\Delta E_{0\to1}$, as well as between the first and the second excited state, $\Delta E_{1\to2}$. We get

$$\Delta E_{0\to1} = h\nu - h\nu[(1+1/2)^2 - (0+1/2)^2]\beta = h\nu(1-2\beta),$$

$$\Delta E_{1\to2} = h\nu - h\nu[(2+1/2)^2 - (1+1/2)^2]\beta = h\nu(1-4\beta).$$

Inserting the calculated $h\nu$ and β gives $\Delta E_{0\to1} = 4155$ cm^{-1} and $\Delta E_{1\to2} = 3797$ cm^{-1}. The first value agrees very well with the experimental value[56] of 4161 cm^{-1}; comparison of the second value with the measured 3926 cm^{-1} is a little bit worse, although it is still not bad for our simple theory. The quantity D represents the *binding energy*, i.e., the energy difference between the well bottom and the energy of the dissociated atoms. In order to obtain the dissociation energy we have to consider that the system does not start from the energy corresponding to the bottom of the curve, but from the level with $v = 0$ and energy $\frac{1}{2}h\nu$, hence our estimation of the *dissociation energy* is $E_{\text{diss}} = D - \frac{1}{2}h\nu = 36\,156$ cm^{-1}, while the experimental value amounts to 36 118 cm^{-1}.

Example 3 (Two water molecules). The above example pertains to a chemical bond. Let us take in the same way a quite different situation, where we have relatively weak intermolecular interactions, namely, the hydrogen bond between two water molecules. The binding energy in such a case is of the order of $D = 6$ kcal/mol $= 0.00956$ a.u. $= 2097$ cm^{-1}, i.e., about twenty times smaller than before. To stay within a single oscillator model, let us treat each

---

[54] L. Wolniewicz, *J. Chem. Phys.*, 103(1995)1792.

[55] This choice is of course arbitrary.

[56] I. Dabrowski, *Can. J. Phys.*, 62(1984)1639.

water molecule as a point-like mass. Then $\mu = 16\,560$ a.u. Let us stay with the same value of $\alpha = 1$. We obtain (p. 227) $a = 17.794$ and hence $b_0 = 17.294$, $b_1 = 16.294$, ..., $b_{17} = 0.294$, $b_{n>17} < 0$. Thus, (accidentally) we also have 18 vibrational levels.

This time from Eq. (4.26), we have $h\nu = 0.001074$ a.u. $= 235$ cm^{-1} and $\beta = 0.02810$ a.u.; therefore $\Delta E_{0\to 1} = 222$ cm^{-1} and $\Delta E_{1\to 2} = 209$ cm^{-1}. These numbers have the same order of magnitude as those appearing in experiments (cf. p. 425).

4.9 Rigid rotator

A rigid rotator is a system of two point-like masses, m_1 and m_2, with a constant distance R between them. The Schrödinger equation may be easily separated giving two equations, one for the center-of-mass motion and the other for the relative motion of the two masses (see Appendix J on p. 691). We are interested only in the second equation, which describes the motion of a particle of mass μ equal to the reduced mass of the two particles, and position in space given by the spherical coordinates R, θ, ϕ, where $0 \le R < \infty, 0 \le \theta \le \pi, 0 \le \phi \le 2\pi$. The kinetic energy operator is equal to $-\frac{\hbar^2}{2\mu}\Delta$, where the Laplacian Δ represented in the spherical coordinates is given in Appendix H on p. 683. Since R is a constant, the part of the Laplacian which depends on the differentiation with respect to R is absent.[57] In this way we obtain the equation (equivalent to the Schrödinger equation) for the motion of a particle on a sphere:

$$-\frac{\hbar^2}{2\mu R^2}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{(\sin\theta)^2}\frac{\partial^2}{\partial\phi^2}\right\}Y = EY, \tag{4.32}$$

where $Y(\theta,\phi)$ is the wave function to be found and E represents the energy. This equation may be rewritten as

$$\hat{J}^2 Y = 2\mu R^2 E Y, \tag{4.33}$$

where the square of the angular momentum operator is equal to

$$\hat{J}^2 = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right].$$

The equation may be rewritten as

$$\frac{1}{Y}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y}{\partial\theta}\right)+\frac{1}{(\sin\theta)^2}\frac{\partial^2 Y}{\partial\phi^2}\right\} = \lambda,$$

---

[57] This reasoning has a heuristic character, but the conclusions are correct. Removing an operator is a subtle matter. In the correct solution to this problem we have to consider the two masses with a variable distance R with the full kinetic energy operator and potential energy in the form of the Dirac delta function (see Appendix E on p. 659) $-\delta(R - R_0)$.

where $\lambda = -\frac{2\mu R^2}{\hbar^2}E$. The solution of the equation is known in mathematics as a *spherical harmonic*[58]; it exists only if $\lambda = -J(J+1)$, $J = 0, 1, 2, \ldots$:

$$Y_J^M(\theta,\phi) = N_{JM} \cdot P_J^{|M|}(\cos\theta) \cdot \frac{1}{\sqrt{2\pi}} \exp(iM\phi), \tag{4.34}$$

where $N_{JM} = \sqrt{\frac{2J+1}{2}\frac{(J-|M|)!}{(J+|M|)!}}$ is the normalization coefficient and P is the *associated Legendre polynomial*[59] defined as

$$P_J^{|M|}(x) = (1-x^2)^{\frac{|M|}{2}} \frac{d^{|M|}}{dx^{|M|}} P_J(x) \tag{4.35}$$

with the *Legendre polynomial*

$$P_J(x) = \frac{1}{2^J J!} \frac{d^J}{dx^J}\left(x^2-1\right)^J. \tag{4.36}$$

From the uniqueness of the solution (Fig. 2.6g,h) it follows that M has to be an integer[60]. The solution exists if $J = 0, 1, 2, 3, \ldots$, and from the derivative in Eq. (4.35) it follows that M cannot exceed[61] J (a very important constraint), because otherwise $Y = 0$. The energetic levels are given by

$$E_J = J(J+1)\frac{\hbar^2}{2\mu R^2} \tag{4.37}$$

for $J = 0, 1, 2, \ldots$.

It is seen that the lowest energy level ($J = 0$) corresponds to $Y_0^0 =$ const (the function is, of course, nodeless, Fig. 4.20a). This means that all orientations of the rotator are of equal

---

[58] There are a few definitions of the spherical harmonics in the literature (see E.O. Steinborn, K. Ruedenberg, *Advan. Quantum Chem.*, 7(1973)1). The Condon–Shortley convention is often used, which is related to the definition given above in the following way: $Y_J^M = \varepsilon_M \left[Y_J^M\right]_{CS}$, $Y_J^J = (-1)^J \left[Y_J^J\right]_{CS}$, where $\varepsilon_M = i^{|M|+M}$.

[59] Adrien Legendre (1752–1833), a French mathematician, professor at the École Normale Supérieure – an elite school in France founded by Napoleon Bonaparte.

[60] Indeed, since ϕ is an angle, we have $\exp(iM\phi) = \exp[iM(\phi+2\pi)]$. Hence, $\exp(iM2\pi) = 1$, and, therefore, $\cos(2\pi M) = 1$ and $\sin(2\pi M) = 0$. This is fulfilled only if M is an integer.

[61] $P_J(x)$ is a polynomial of J-th degree, while $\frac{d^{|M|}}{dx^{|M|}}$ in $P_J^{|M|}(x)$ decreases the degree by M. If M exceeds J, then $P_J^{|M|}(x)$ automatically becomes equal to zero.

probability. The first excited state corresponds to $J = 1$ and is triply degenerate, since $M = 0, \pm 1$. The corresponding wave functions are (within accuracy of a multiplicative constant) $Y_1^0 \propto \cos\theta$, $Y_1^1 \propto \sin\theta \exp(i\phi)$, $Y_1^{-1} \propto \sin\theta \exp(-i\phi)$. The first function, being real, may be easily plotted (Fig. 4.20b), while the second and the third are not (they are complex). Since they both correspond to the same eigenvalue of the Hamiltonian, their arbitrary linear combination is an equally good eigenfunction of this operator. We may take, therefore, Y_1^1 and Y_1^{-1} as $\psi_1 = \frac{1}{2}\left(Y_1^1 + Y_1^{-1}\right) \propto \sin\theta\cos\phi$ and $\psi_1 = \frac{1}{2i}\left(Y_1^1 - Y_1^{-1}\right) \propto \sin\theta\sin\phi$. Both functions are real, and they are shown in Fig. 4.20c,d. Note that again we have the usual situation: the ground state is nodeless, the first excited state has a single node, etc.

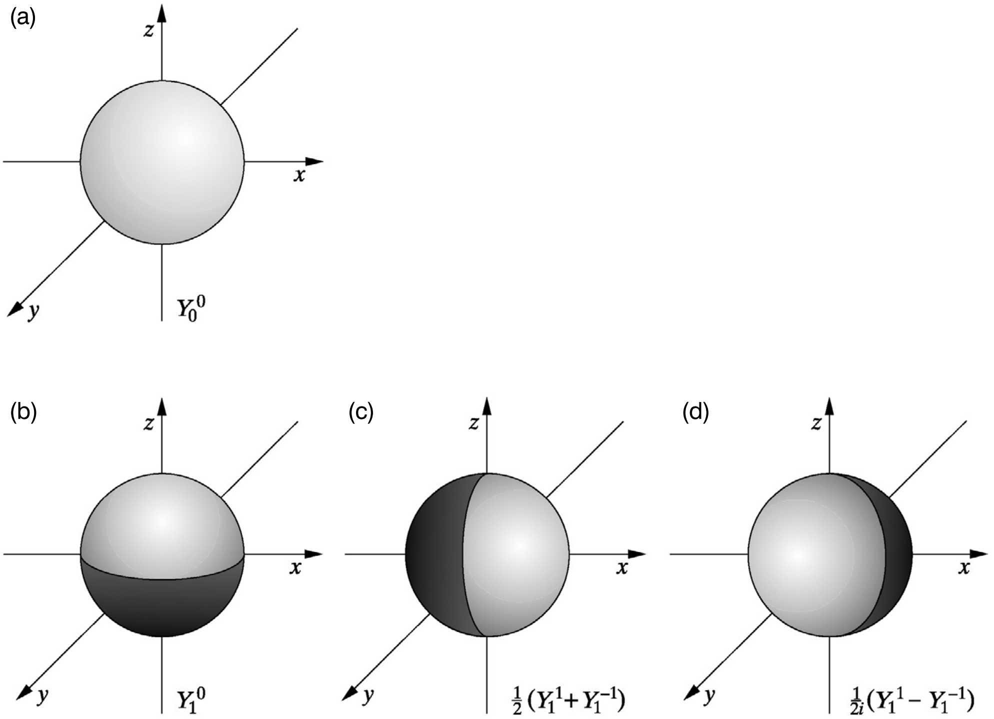


Fig. 4.20. A schematic representation of the nodes for rigid rotator wave functions: (a) Ground state (nodeless). (b) Triply degenerate first excited state (single node).

The term Y_J^M is not only the eigenfunction of the Hamiltonian $\hat{H}$ and of the square of the angular momentum $\hat{J}^2$ but also of the z component of the angular momentum operator $\hat{J}_z = -i\hbar\frac{\partial}{\partial\phi}$, i.e.,

$$\hat{J}_z Y_J^M = M\hbar Y_J^M. \tag{4.38}$$

4.10 Hydrogen-like atom

We have two particles: an electron of mass m and charge $-e$ and a nucleus of mass M and charge $+Ze$. The Hamiltonian contains two kinetic energy operators and the Coulombic interaction $-Ze^2/r$, where r is the electron–nucleus separation. We have, therefore, six coordinates. In Appendix J on p. 691, it is shown how the center-of-mass motion can be separated (we are not interested in this motion). There remain three coordinates, x, y, z, showing where the electron is with respect to the nucleus. The resulting Schrödinger equation contains a single kinetic energy operator of a particle of reduced mass μ (almost equal to the electron mass) with coordinates x, y, z and Coulombic interaction of the electron and the nucleus (as before). Now, instead of x, y, z, we introduce the spherical coordinates r, θ, ϕ. Then, as the class Q solution we obtain

$$\boxed{\psi_{nlm}(r,\theta,\phi) = N_{nl} R_{nl}(r) Y_l^m(\theta,\phi), \qquad (4.39)}$$

where Y_l^m is identical to the solution (4.34) of a rigid rotator of length r, and the function R_{nl} has the following form in a.u.:

$$R_{nl}(r) = r^l L_{n+l}^{2l+1}\left(\frac{2Zr}{na_0}\right) \exp\left(-\frac{Zr}{na_0}\right), \tag{4.40}$$

where the *Bohr first orbit radius*

$$a_0 = \frac{1}{\mu} \simeq 1 \text{ a.u.}, \tag{4.41}$$

where

principal quantum number $n = 1, 2, 3...,$
azimuthal quantum number $l = 0, 1, 2, ..., n-1,$
magnetic quantum number $m = -l, -l+1, ..., 0, ... + l,$

and the *associated Laguerre polynomial* $L_\alpha^\beta(x)$ is defined as

$$L_\alpha^\beta(x) = \frac{d^\beta}{dx^\beta} L_\alpha(x), \tag{4.42}$$

while the *Laguerre polynomial* is given by[62]

$$L_\alpha(x) = \exp(x)\frac{d^\alpha}{dx^\alpha}[x^\alpha \exp(-x)]. \tag{4.43}$$

Since the Hamiltonian commutes with the square of the total angular momentum operator $\hat{J}^2$ and with the operator of $\hat{J}_z$ (cf. Chapter 2 and Appendix F on p. 665), the functions ψ_{nlm} are also the eigenfunctions of these two operators:

$$\hat{H}\psi_{nlm} = E_n\psi_{nlm}, \tag{4.44}$$

$$\hat{J}^2\psi_{nlm} = l(l+1)\hbar^2\psi_{nlm}, \tag{4.45}$$

$$\hat{J}_z\psi_{nlm} = m\hbar\psi_{nlm}, \tag{4.46}$$

where in a.u.

$$E_n = -\frac{Z^2}{2n^2}\left(\frac{1}{1+\frac{1}{M_p}}\right), \tag{4.47}$$

with M_p representing the proton mass (in a.u.), i.e., about 1840. The content in parentheses in the last formula turns out to be 0.999457, almost 1, which would be obtained for an infinite mass of the nucleus.

Each of the energy levels is n^2-fold degenerate. Note that the hydrogen atom energy depends solely on the principal quantum number n. The fact that the energy does not depend on the projection of the angular momentum $m\hbar$ is natural, because the space is isotropic and no direction is privileged. However, the fact that it does not depend on the length of the angular momentum $\sqrt{l(l+1)}\hbar$ is at first sight strange. The secret is in the Coulombic potential $\frac{1}{r}$ produced by the *point-like* nucleus and is connected with the notion of dynamic symmetry mentioned on p. 94. If we considered a nonpoint-like nucleus or were interested in the orbital $2s$ of a quasihydrogen atom such as lithium,[63] then the energy would depend on the quantum number l.

The one-electron wave functions (orbitals) of the hydrogen atom with $l = 0$ are traditionally denoted as ns: $1s, 2s$, ..., with $l = 1$, as np: $2p, 3p, 4p$, ..., with $l = 2, 3$, ..., as nd: $3d, 4d$, nf: $4f, 5f$, ...

---

[62] Both $L_\alpha^\beta(x)$ and L_α are indeed polynomials of $\alpha - \beta$ degree. If $\beta > \alpha$, from (4.42) it follows that $L_\alpha^\beta = 0$.

[63] In which the nucleus is screened by a cloud of two $1s$ electrons. The $2s$ electron thinks it is in a hydrogen atom with a spacious nucleus of the size of the $1s$ orbital and an effective charge $+1$.

The wave functions ψ_{nlm} can be plotted in several ways. For example, the function $(nlm) = (100)$ or $1s$ given by the formula

$$1s \equiv \psi_{100}(r, \theta, \phi) = \sqrt{\frac{Z^3}{\pi}} \exp(-Zr), \tag{4.48}$$

and can be visualized in several alternative forms shown in Fig. 4.21.

We see that what the electron likes most is to sit on the nucleus. Indeed, if we chopped the space into tiny cubes, computed the value of $(1s)^2$ in each cube (the function is real, therefore, the modulus is irrelevant), and multiplied the number obtained by the volume of the cube, the resulting number in each cube would have a meaning of the probability of finding the electron in a particular cube. Evidently, this number will be largest for the cube that contains the nucleus (the origin). We were told about the Bohr model[64] in school, about the orbits, and about the first Bohr orbit (corresponding to the atom ground state). Do we relegate all this to fairy tales? Not completely. If we changed the question: what is the *distance* at which the electron is most likely to be found, then the answer should indeed be as we were taught in school: the first Bohr orbit. This is easy to show by computing the *radial probability density* of finding the electron (i.e., integrating over all orientations, leaving the dependence on the distance): $\rho(r) = \int d\theta d\phi r^2 \sin\theta \, |\psi_{100}|^2 = 4Z^3 r^2 \exp(-2Zr)$, i.e., $(1s)^2$ function computed for a given $r = r_0$ and multiplied by the volume $4\pi r_0^2 dr$ confined between two concentric spheres, one with radius r_0, the other with radius $r_0 + dr$, gives the probability of finding the electron exactly between these spheres. The maximum of $\rho(r)$ corresponds exactly to $r = 1$ a.u. or the first Bohr orbit radius.[65]

The $2s$ orbital $(n = 2, l = 0, m = 0)$ reads as (Fig. 4.22)

$$2s \equiv \psi_{200}(r, \theta, \phi) = N_{2s}\,(Zr - 2) \exp(-Zr/2), \tag{4.49}$$

with the normalization constant $N_{2s} = \frac{Z^{\frac{3}{2}}}{4\sqrt{2\pi}}$. A sphere of radius $2/Z$ (representing the nodal sphere) contains[66] only a little more than 5% of the total electronic density (independently of Z).[67]

---

[64] Nobody is perfect. Neither are geniuses. Here is a story by John Slater: “Brillouin delivered an interesting lecture concerning his relations. When he finished, Bohr stood up and attacked him with an inhuman fury. I have never heard any adult to scold another person in public with such an emotional engagement without any reason whatsoever. After this show I have decided that my antipathy with respect to Bohr dating since 1924 continues to be justified.”

[65] The computed maximum position does not coincide with the mean value of r (cf. Appendix H on p. 683) $\langle r \rangle = \langle \psi_{100} | r \psi_{100} \rangle = \int_0^\infty dr r \rho(r) = \frac{3}{2}$ a.u.

[66] See the Mathematica files for Chapter 4 in the web annex.

[67] A sphere of the same radius encloses about 75% of the electron density for the $1s$ orbital.

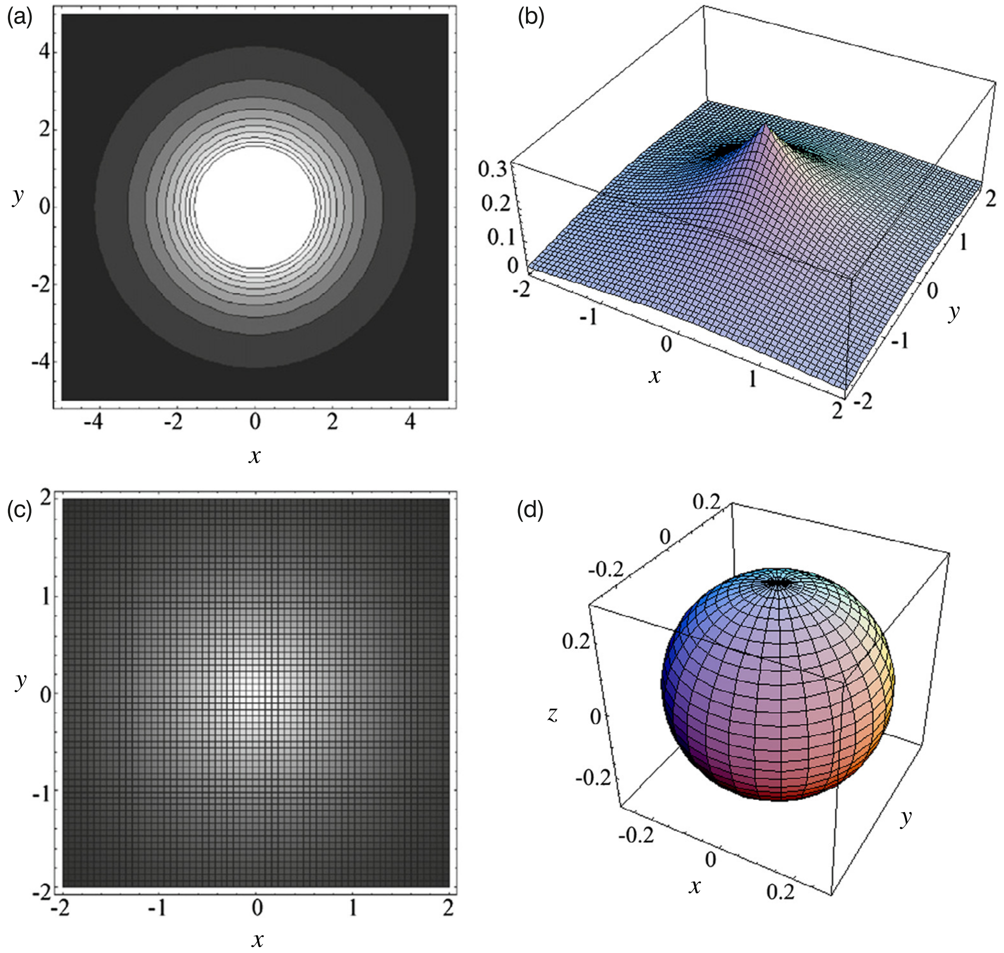


Fig. 4.21. Various ways of visualization of the $1s$ hydrogen orbital, which is a function of electron position in three-dimensional space (coordinates x, y in (a)–(c) in a.u.). (a) Isolines of the $z = 0$ section of the wave function (going through the nucleus). Black means the value zero, white means a high value. This is a map of a mountain. The center of (a) shows a large white plateau, which represents, however, an artifact. (b) In fact, the section of the $1s$ orbital as a function of r represents a mountain with a sharp summit (a discontinuity of the first derivative). (c) is similar to (a), but instead of isolines we have a white mist with the highest concentration in the center, disappearing exponentially with increasing distance r. (d) shows a spherically symmetric isosurface of the wave function.

The wave functions (orbitals) with $m \neq 0$ are difficult to draw, because they are complex. We may, however, plot the real part of ψ_{nlm} ($\text{Re}\psi_{nlm}$) by taking the sum of ψ_{nlm} and ψ_{nl-m}, i.e., $2\text{Re}\psi_{nlm}$, and the imaginary part of ψ_{nlm} ($\text{Im}\psi_{nlm}$) from the difference of ψ_{nlm} and ψ_{nl-m} equal to $2i\text{Im}\psi_{nlm}$. These functions are already real and can be easily plotted. In this way we obtain the orbitals $2p_x$ and $2p_y$ from the functions ψ_{211} and ψ_{21-1}. The orbital ψ_{210} is identical

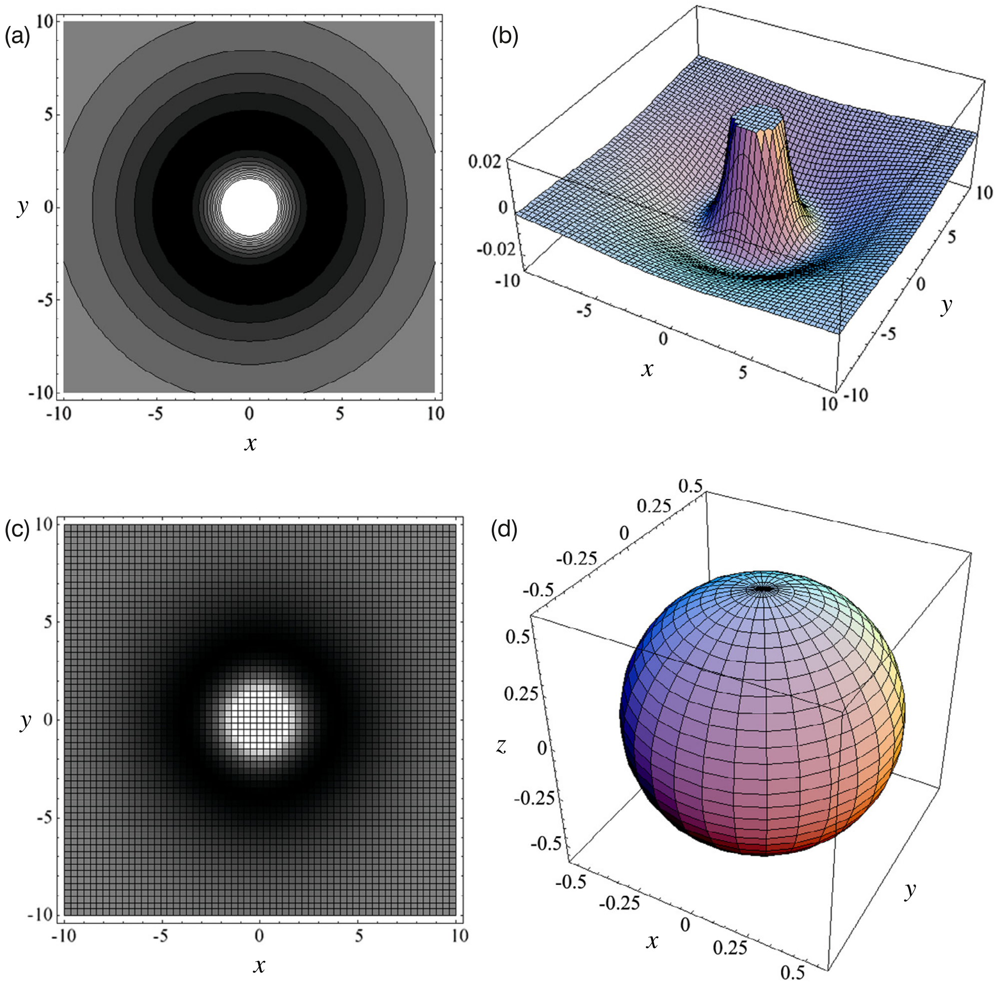


Fig. 4.22. Various graphical representations of the hydrogen $2s$ orbital (coordinates x, y in (a)–(c) in a.u.). (a) Isolines of $z = 0$ section of the orbital. Gray means the value zero, white a positive value, black a negative value. Note that gray is not only found at peripheries, but also around the center. This means that the $2s$ orbital exhibits a nodal sphere of radius 2 a.u. (see Eq. (4.49)) that would contain a little more than 5% of the electronic density (whereas for the $1s$ orbital such a sphere contains about 75% of the electron cloud). The center of (a) shows a quite large white plateau, which represents, however, an artifact. In fact, (b) shows the section of the $2s$ orbital as a function of r, which represents a mountain with a sharp peak (a discontinuity of the first derivative is shown incorrectly on the figure, instead of a sharp summit one has an artifact plateau) with a depression at its base. (c) is similar to (a), but instead of isolines one has a white mist with the largest concentration in the center, then taking negative values (black mist) and finally disappearing exponentially with increasing distance r. (d) shows a spherically symmetric isosurface of the wave function (the sphere was shown as larger than the $1s$ orbital, because the $2s$ orbital decays more slowly than $1s$).

to $2p_z$, i.e.,

$$
\begin{aligned}
2p_x &= N_{2p} x \exp(-Zr/2),\\
2p_y &= N_{2p} y \exp(-Zr/2),\\
2p_z &= N_{2p} z \exp(-Zr/2),
\end{aligned}
$$

where an easy calculation (just five lines) gives the normalization constant $N_{2p} = ZN_{2s}$. The $2p$ orbitals are shown in Fig. 4.23. Note, please, that a linear combination of eigenfunctions is not necessarily an eigenfunction. It is if the functions mixed correspond to the same eigenvalue. This is why $2p_x$ and $2p_y$ are the eigenfunctions of the Hamiltonian, of the square of the angular momentum operator, but are not eigenfunctions of $\hat{J}_z$.

Similarly we obtain the five real $3d$ orbitals the solutions of the Schrödinger equation. They can be easily obtained from Eq. (4.39) and subsequently making them real by choosing $\mathrm{Re}\psi_{nlm}$ and $\mathrm{Im}\psi_{nlm}$. As a result we have the following normalized $3d$ orbitals ($N_{3d} = \frac{Z^{\frac{7}{2}}}{81}\sqrt{\frac{2}{\pi}}$):

$$
\begin{aligned}
3d_{xy} &= N_{3d} xy \exp(-Zr/3),\\
3d_{xz} &= N_{3d} xz \exp(-Zr/3),\\
3d_{yz} &= N_{3d} yz \exp(-Zr/3),\\
3d_{x^2-y^2} &= \frac{1}{2} N_{3d} (x^2 - y^2) \exp(-Zr/3),\\
3d_{3z^2-r^2} &= \frac{1}{2\sqrt{3}} N_{3d} (3z^2 - r^2) \exp(-Zr/3).
\end{aligned}
$$

The $3d$ orbitals are shown[68] in Figs. 4.23–4.25. A summary of the low-energy hydrogen atomic orbitals is shown in Fig. 4.26.[69]

---

[68] It pays to memorize the abbreviations: $2p_x, 2p_y, 2p_z$ and the five $3d$ orbitals. Indeed, we may then easily write down their mathematical formulae (even neglecting the normalization constants). Having the formulae we may draw any section of them, i.e., we can foresee Figs. 4.24 and 4.25.

[69] A night bus ride might enrich us with some unexpected impressions. Of all atomic orbitals you may most easily see orbital $1s$. Just look through the condensation on a bus window at a single street lamp. You will see a gleam that decays to black night. You may also quite easily find a double lamp that will offer you a $2p$ orbital and sometimes have the chance to see some of the $3d$ orbitals. Once I have even found the $2s$ orbital, but I do not understand how it was possible. I was looking at a single lamp, which made an intense gleam in the center, which gradually decayed and then again an annular gleam appeared that finally vanished. This is what the square of the $2s$ orbital looks like.

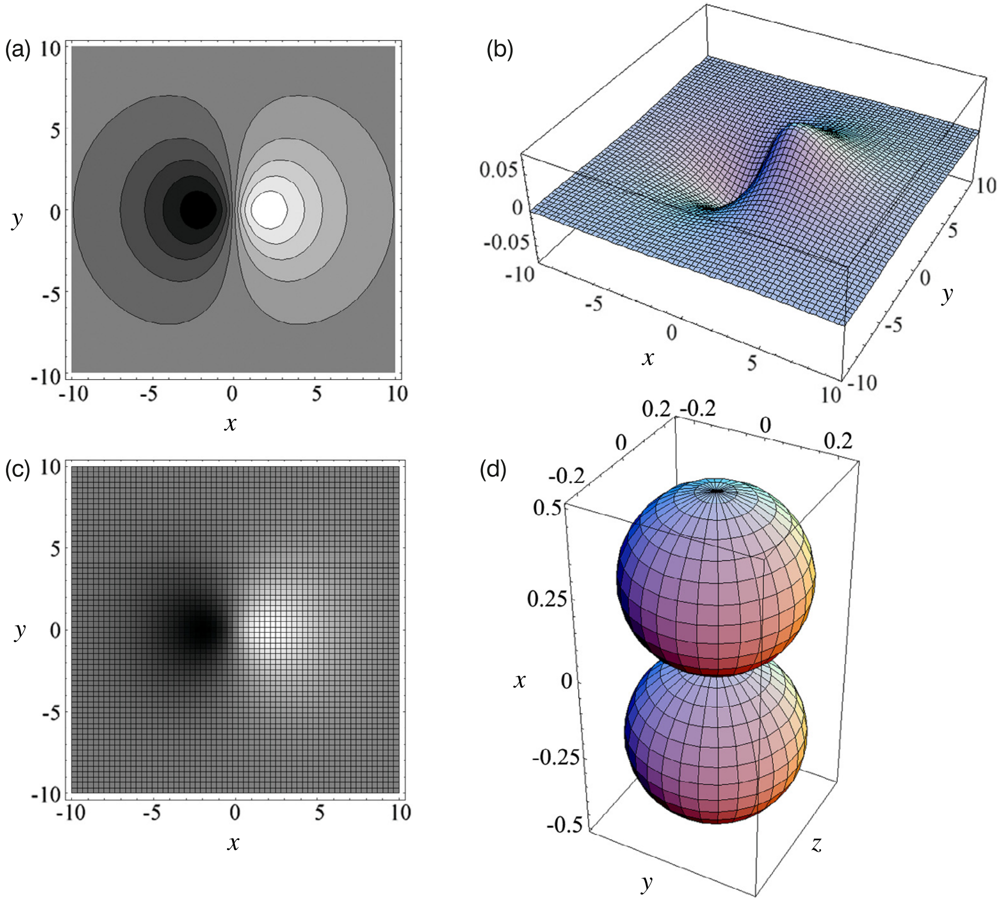


Fig. 4.23. Various graphical representations of the hydrogen $2p_x$ orbital (coordinates x (horizontal) and y (perpendicular) in (a)–(c) in a.u.). The two other $2p$ orbitals, $2p_y$ and $2p_z$, look the same as $2p_x$, but are oriented along axes y and z, respectively. Note that for the hydrogen atom all four orbitals $2s, 2p_x, 2p_y, 2p_z$ correspond to the same energy, and all have a single nodal surface. For $2s$ the surface (Fig. 4.22) is a sphere of radius 2, for the $2p_x$, $2p_y$, and $2p_z$ orbitals the nodal surfaces are the planes $x, y, z = 0$. (a) Isolines of the $z = 0$ section of the orbital. Gray means zero, white means a high value, black means a negative value. (b) The values of the section $z = 0$. Note (a,b) that the right-hand side of the orbital is positive, the left-hand side is negative. The maximum (minimum) value of the orbital is found at $x = 2$ ($x = -2$) a.u. (c) is similar to (a), but instead of isolines we have a mist with the largest concentration (positive, white) on the right and the smallest (and negative, black) concentration on the left. The orbital finally disappears exponentially with increasing distance r from the nucleus. (d) shows an isosurface of the absolute value of the angular part of the wave function $|Y_1^0|$. As for Y_1^0 itself, one of the lobes takes negative, the other positive values, and they touch each other at the position of the nucleus. To obtain the orbital, we have to multiply this angular function by a spherically symmetric function of r. This means that an isosurface of the absolute value of the wave function will also have two lobes (for the wave function itself, one will be positive, the other negative), but they will not touch each other in full accordance with (a).

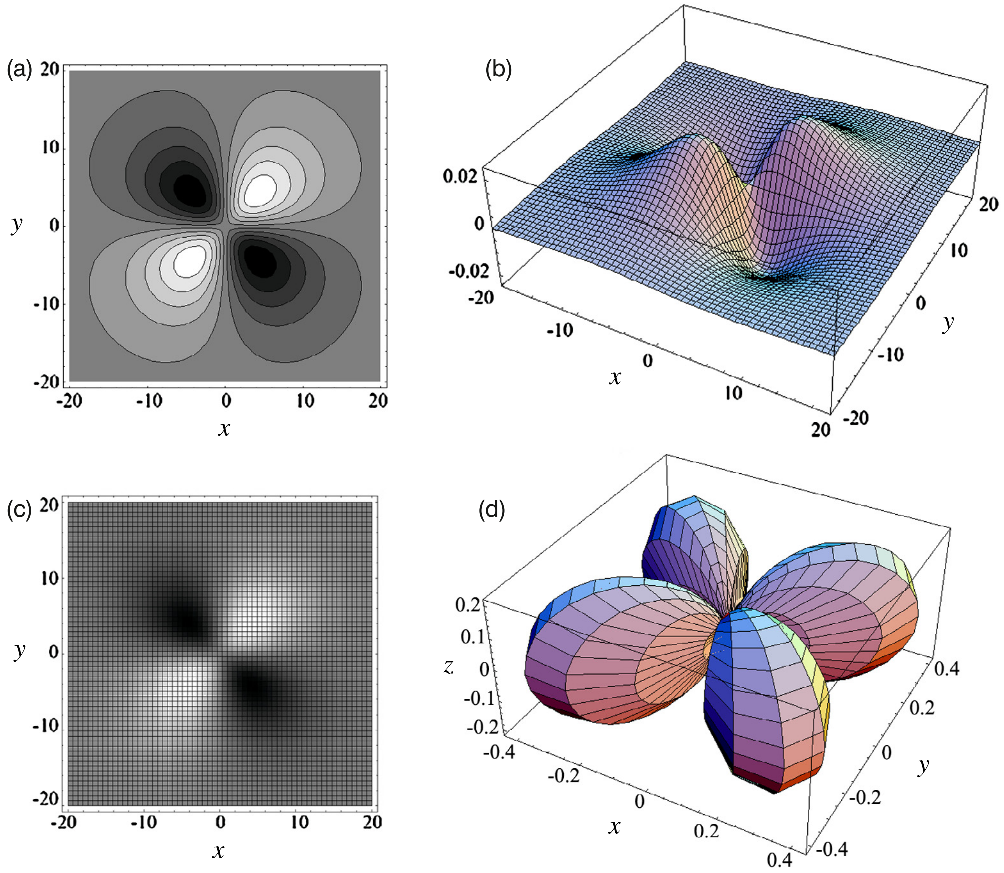


Fig. 4.24. Various graphical representations of the hydrogen $3d_{xy}$ orbital (coordinates x (horizontal) and y (perpendicular) in (a)–(c) in a.u.). The three other $3d$ orbitals, $3d_{yz}$, $3d_{xz}$, and $3d_{x^2-y^2}$, look the same as $3d_{xy}$, but are oriented in space according to their indices (in the last case along x and y axes). (a) Isolines of $z = 0$ section of the orbital. Gray means zero, white means a positive value, black means a negative value. Note (a,b) that $3d$ orbitals are symmetric with respect to inversion. (b) One may imagine the $z = 0$ section of $3d_{xy}$ as two hills and two valleys. (c) is similar to (a), but instead of isolines one has a white mist with the highest concentration on the North–East line and the smallest (and negative) concentration on the North–West line (black mist). The orbital finally disappears exponentially with increasing distance r from the nucleus. (d) shows an isosurface of the absolute value of the angular part of the wave function (just ignore the nasty edges): $|Y_2^2 - Y_2^{-2}|$. As for $Y_2^2 - Y_2^{-2}$ itself two of the lobes take the negative, the other two take positive values, they touch each other at the nucleus. To obtain the orbital one has to multiply this angular function by a spherically symmetric function of r. This means that an isosurface of the absolute value of the wave function will also have four lobes (for the wave function itself two will be positive, the other two negative), but they will not touch in full accordance with (a).

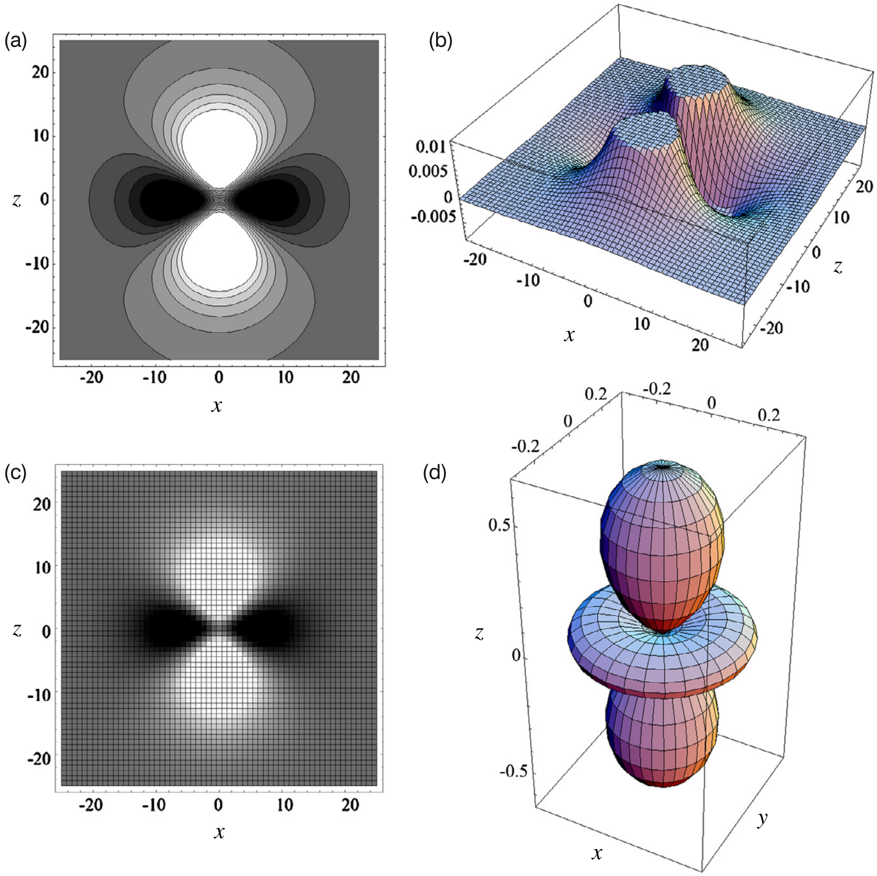


Fig. 4.25. Various graphical representations of the hydrogen $3d_{3z^2-r^2}$ orbital (coordinates x (horizontal) and z (perpendicular) in (a)–(c) in a.u.). The z axis is vertical, the x axis is horizontal. (a) Isolines of the xz section of the orbital. Gray means zero, white means a high positive value, black means a negative value. Note that in (a) and (b) $3d_{3z^2-r^2}$ orbitals are symmetric with respect to inversion. (b) We may imagine the xz section of the $3d_{3z^2-r^2}$ as two hills and two valleys, where the hills are higher than the depth of the valleys (the plateaus in (b) are artificial). (c) is similar to (a), but instead of isolines one has a mist with the highest concentration (white) on the North–South line and the smallest (and negative, black mist) concentration on the East–West line. The orbital finally disappears exponentially with increasing distance r from the nucleus. (d) shows an isosurface of the absolute value of the angular part of the wave function ($|Y_2^0|$). As for Y_2^0 itself there are two positive lobes and a negative ring; they touch each other at the nucleus. To obtain the orbital, we have to multiply this angular function by a spherically symmetric function of r. This means that an isosurface of the absolute value of the wave function will have also two lobes along the z axis as well as the ring, but they will not touch in accordance with (a). The lobes along the z axis are positive, the ring is negative. A peculiar form of $3d_{3z^2-r^2}$ becomes more familiar when one realizes that it simply represents a sum of two "usual" $3d$ orbitals. Indeed, $3d_{3z^2-r^2} = \frac{1}{2\sqrt{3}} N_1 [2z^2 - \left(x^2+y^2\right)] \exp(-Zr/3) = \frac{1}{2\sqrt{3}} N_1 [\left(z^2-x^2\right) + \left(z^2-y^2\right)] \exp(-Zr/3) = \frac{1}{2\sqrt{3}} N_1 \frac{2}{N_1} \left(3d_{z^2-x^2} + 3d_{z^2-y^2}\right) = \frac{1}{\sqrt{3}} \left(3d_{z^2-x^2} + 3d_{z^2-y^2}\right)$.

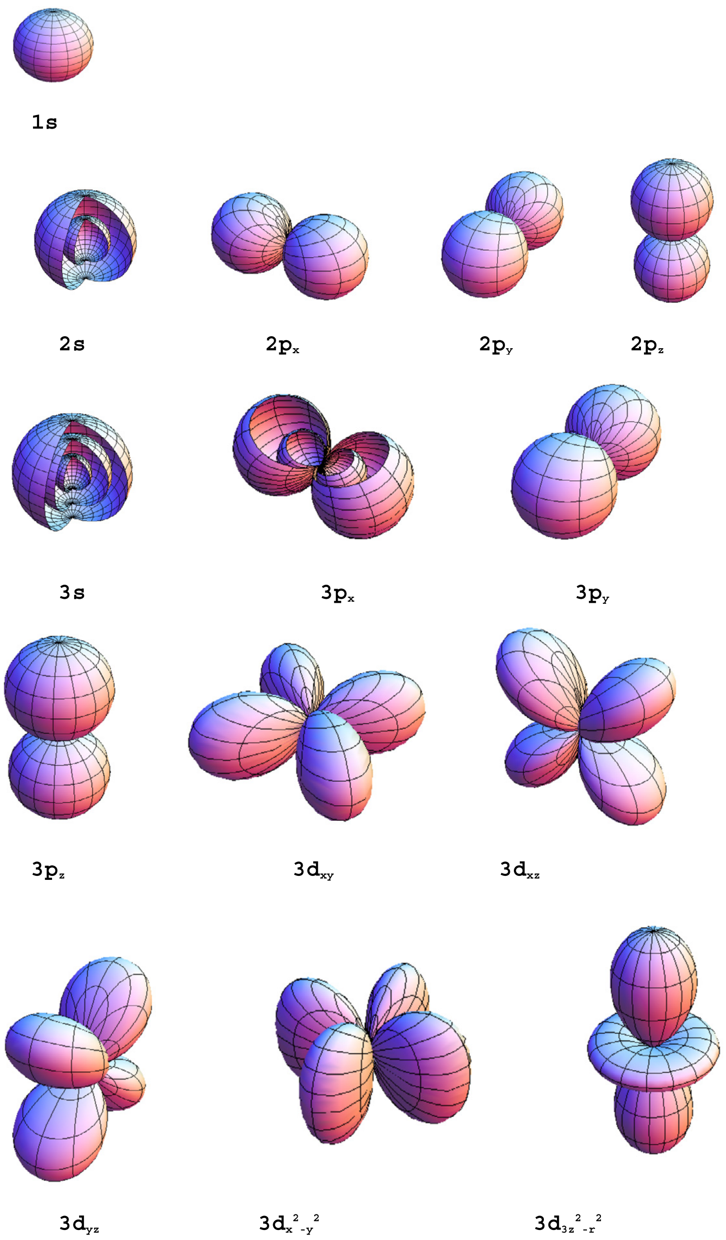


Fig. 4.26. A review of all orbitals with the quantum number $n = 1, 2, 3$. The x, y, z axes (not shown) are oriented in the same way (as the directions of $2p_x, 2p_y, 2p_z$ orbitals, respectively). The figures $2s$, $3s$ are schematic, their cross-sections are intended to underline that these orbitals are spherically symmetric, but possess a certain "internal structure." The $1s$ orbital decays monotonically with r (this is shown by a limiting sphere), but $2s$ and $3s$ change sign one and two times, respectively. The internal spheres displayed symbolize the corresponding nodal spheres. The orbital $3p_x$ (representing also $3p_y$ and $3p_z$) is shown in a similar convention: there is an extra nodal surface inside (besides the plane $x = 0$), resembling a smaller orbital p.

4.10.1 Positronium and its short life in molecules

A bound pair of an electron and a positron (the "positive electron" predicted by Dirac and discovered by Anderson, see Chapter 3) is known as positronium. The positronium can be detected in radiation of some radioactive elements, and is a product of some collisions of atomic nuclei. Its lifetime is of the order of a millisecond. After such a time a mutual annihilation of the two objects takes place with the emission of γ photons.

The Schrödinger equation gives only the stationary states (the ground state and the excited ones), which live forever. It cannot describe, therefore, the annihilation process, but it is able to describe the positronium before the annihilation takes place. This equation for the positronium is identical to that for the hydrogen atom, except that the reduced mass changes from $\mu \approx 1$ a.u. to $\mu = \frac{1}{2}$ a.u. ($\frac{1}{\mu} = \frac{1}{1} + \frac{1}{1} = 2$). As a result (see the definition of a_0 on p. 232) one may say that "the size the orbitals of the positronium doubles" when compared to that of the hydrogen atom, since their exponential decay coefficient is now multiplied by $\frac{1}{2}$. It is irrelevant whether in the center of the positronium (where the positronium orbitals are centered) one puts the positron or the electron, the view of the positronium from such a center is the same.[70] In both cases what moves is a "quasiparticle" of the opposite charge to that kept in the center, its mass being $\frac{1}{2}$.

A millisecond is a short time for us, but a very long time in the molecular world. Let us take a look of a system: proton + two electrons + positron, where the positron might be seen to play the role of an extremely light second nucleus in such a "hydrogen molecule." If the corresponding "atoms" (the hydrogen atom and the positronium) did not interact, one would have the total energy: $-\frac{1}{2} - \frac{1}{4} = -0.75$ a.u., while an accurate result for the total system under consideration[71] is equal to -0.789 a.u. Therefore the binding energy of this unusual "hydrogen molecule" is of the order of 25 kcals/mol, i.e., about $\frac{1}{4}$ of the binding energy of the hydrogen molecule.

4.11 What do all these solutions have in common?

- In all the systems considered, the stationary states are similar: the number of their nodes increases with their energy (the nodeless function corresponds to the lowest energy).
- If the potential energy is a *constant* (particle in a box, rigid rotator), then the energy level (nearest neighbor) distance *increases* with the energy.[72] The energy levels become more dense for larger boxes, longer rotators, etc.

---

[70] The situation is similar to what we have seen in the case of the hydrogen atom and the observer sitting on the electron or on the proton, Appendix J, p. 691.

[71] K. Strasburger, H. Chojnacki, *J. Chem. Phys.*, 108(1998)3218.

[72] In both cases the distance goes as the square of the quantum number.

- *A parabolic* potential energy well (harmonic oscillator) reduces this tendency and the energy levels are *equidistant*. The distance decreases if the parabola gets wider (less restrictive).
- The Morse potential energy curve may be seen as a function that may be approximated (as the energy increases) by wider and wider parabolic sections. No wonder, therefore, that the level distance *decreases*. The number of energy levels is finite.[73]
- The Coulomb potential, such as that for the hydrogen atom, resembles vaguely the Morse curve. Yet its form is a little similar to the Morse potential (dissociation limit, but infinite depth). We expect, therefore, that the energy levels for the hydrogen-like atom will become closer and closer when the energy increases, and we are right. Is the number of these energy levels finite as for the Morse potential? This is a more subtle question. Whether the number is finite or not decides the asymptotics (the behavior at infinity). The Coulomb potential makes the number infinite.

4.12 Hooke helium atom (harmonium)

Two-electron atoms or molecules already represent a serious problem for quantum chemistry, because the mutual correlation of electronic motions must be carefully taken into account. As we will see in further chapters, such calculations are feasible, but the wave functions are very complicated, e.g., may represent linear combinations of thousands of terms and still only be *approximations* to the exact solution to the Schrödinger equation. This is why people were surprised when Kais et al. showed that a two-electron atomic system has *an exact analytical solution*.[74]

Unfortunately, this wonderful two-electron system is (at least partially) nonphysical. It represents a strange helium atom, in which the two electrons (their distance denoted by r_{12}) interact through the Coulombic potential, but each is attracted to the nucleus by a ... harmonic spring, i.e., satisfying the Hooke law (of equilibrium length 0 and force constant k, with electron–nucleus distances denoted by r_1 and r_2) (Fig. 4.27).

The Hamiltonian of this problem (the adiabatic approximation[75] and atomic units are used) has the following form:

$$\hat{H} = -\frac{1}{2}\Delta_1 - \frac{1}{2}\Delta_2 + \frac{1}{2}k(r_1^2 + r_2^2) + \frac{1}{r_{12}}.$$

---

[73] Such type of reasoning prepares us for confronting real situations. Practically, we will never deal with the abstract cases described in the present chapter, and yet in later years we may say something like this: "*look, this potential energy function is similar to case X in Chapter 4 of that thick boring book we were forced to study. So the distribution of energy levels and wave functions has to be similar to those given there*."

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

[75] The nucleus is assumed to have an infinite mass.

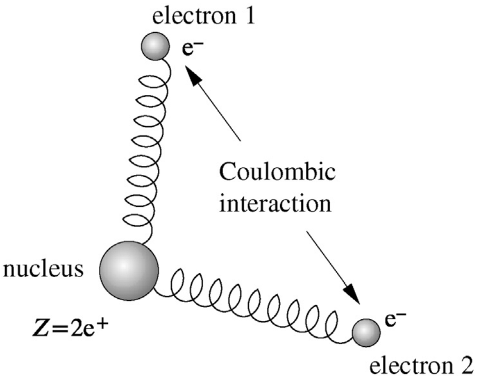


Fig. 4.27. The Hooke helium atom. The electrons repel by Coulombic forces and are attracted by the nucleus by a harmonic (not Coulombic) force.

It is amazing in itself that the Schrödinger equation for this system has an analytical solution (for $k = \frac{1}{4}$). Moreover, it could be an extremely complicated analytical formula. It is a sensation that the solution is dazzlingly beautiful and simple; we have

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = N\left(1 + \frac{1}{2}r_{12}\right)\exp[-\frac{1}{4}\left(r_1^2 + r_2^2\right)],$$

where

$$|N|^2 = \frac{\pi^{\frac{3}{2}}}{8 + 5\sqrt{\pi}}.$$

The wave function represents the product of the two harmonic oscillator wave functions (Gaussian functions), but also an additional extremely simple correlation factor $\left(1 + \frac{1}{2}r_{12}\right)$. As we will see in Chapter V2-2, exactly such a factor is required for the ideal solution. In this exact function there is nothing else, just what is absolutely necessary.[76]

4.13 Hooke molecules

The readers must, together with the author, be astonished by the exact solution of the Schrödinger equation for the harmonium. It is just a miracle: we have two electrons, but still we get an exact (and simple) solution! Is it the end of such wonders?

Well, let us stay for a while with two particles. The Hamiltonian for a system of two point masses m_1 and m_2 with positions shown by vectors $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ and interacting according to the

---

[76] We might have millions of complicated terms.

potential energy term $V_{12}(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)$ (therefore, being a function of their distance only) has the form $\hat{H}(\boldsymbol{r}_1, \boldsymbol{r}_2)$, i.e.,

$$\hat{H}(\mathbf{r}_1, \mathbf{r}_2) \equiv \hat{H}_{12}(\mathbf{r}_1, \mathbf{r}_2) = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 + V_{12}(|\mathbf{r}_1 - \mathbf{r}_2|). \tag{4.50}$$

By introducing the center-of-mass coordinates

$$\mathbf{R}_{CM} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} \tag{4.51}$$

and the coordinates describing the position of particle 1 with respect to the position of particle 2,

$$\boldsymbol{\rho} = \mathbf{r}_1 - \mathbf{r}_2, \tag{4.52}$$

we get[77] the Hamiltonian in the new coordinates (see Appendix J, Example 1), i.e.,

$$\hat{H}(\mathbf{R}_{CM}, \boldsymbol{\rho}) = -\frac{\hbar^2}{2M}\Delta_{CM} - \frac{\hbar^2}{2\mu}\Delta_\rho + V_{12}(\rho), \tag{4.53}$$

where $M = m_1 + m_2$ and the reduced mass $\mu = \frac{m_1 m_2}{M}$.

The resulting Schrödinger equation can be separated into two equations: one describing the motion of the center of mass (with the Hamiltonian $-\frac{\hbar^2}{2M}\Delta_{CM}$) and the other describing the relative motion of the particles (with the Hamiltonian $-\frac{\hbar^2}{2\mu}\Delta_\rho + V_{12}(\rho)$). This looks as if someone described the whole motion as *independent motion of two "particles"*: one of them has the position of the center of mass and the mass equal to M, while the other "particle" has the position of particle 1, seen from our armchair placed on particle 2, and has a mass equal to the reduced mass μ.

This result has already been used by us, e.g., when describing the rigid rotator. We now just want to recall it and to introduce a useful notation. We will soon sail much further while navigating in the same direction.

Three particles

Is a similar separation of the total Schrödinger equation into three one-particle independent equations still possible? It turns out that in principle it is, but under a certain condition. It is possible if the Hamiltonian has the form

---

[77] We replace a set of six coordinates (the components of $\mathbf{r}_1$ and $\mathbf{r}_2$) by a new set of six coordinates.

$$\hat{H}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \hat{H}_{12}(\mathbf{r}_1, \mathbf{r}_2) - \frac{\hbar^2}{2m_3}\Delta_3 + a_{13}r_{13}^2 + a_{23}r_{23}^2, \tag{4.54}$$

where $r_{ab} = |\boldsymbol{r}_a - \boldsymbol{r}_b|$ for $a, b = 1, 2, 3$, and $a_{13} = km_1m_3$, $a_{23} = km_2m_3$, $k \geq 0$!

Therefore, the separation is possible at any potential[78] V_{12} describing the interaction of particles 1 and 2, but particle 3 has to interact pairwise with the other particles by harmonic forces and, in addition, the harmonic force constants a_{13} and a_{23} must be proportional to the corresponding masses.

Now the separation should be demonstrated, but we will not do that, because all steps resemble very much what is shown in Appendix J, p. 691. The separation is achieved by introducing the following new coordinates (each of the vectors given below has three components treated as new coordinates):

$$\mathbf{R}_0 = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2 + m_3\mathbf{r}_3}{m_1 + m_2 + m_3},$$

$$\boldsymbol{\rho} = \mathbf{r}_1 - \mathbf{r}_2,$$

$$\mathbf{R}_1 = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2} - \mathbf{r}_3.$$

The procedure of transforming the Laplacians from the ones expressed in the old coordinates to those given in the new coordinates, the same as shown in Appendix J, leads to three mutually independent Hamiltonians,

$$\hat{H} = \hat{h}_0(\mathbf{R}_0) + \hat{h}_1(\boldsymbol{\rho}) + \hat{h}_2(\mathbf{R}_1), \tag{4.55}$$

where

$$\hat{h}_0(\mathbf{R}_0) = -\frac{\hbar^2}{2M}\Delta_{R_0} \tag{4.56}$$

corresponds to the center-of-mass motion,

$$\hat{h}_1(\boldsymbol{\rho}) = -\frac{\hbar^2}{2\mu_1}\Delta_\rho + V(\rho) + k\mu_1 m_3\rho^2 \tag{4.57}$$

describes the relative motion of particles 1 and 2,

$$\hat{h}_2(\mathbf{R}_1) = -\frac{\hbar^2}{2\mu_2}\Delta_{R_1} + k(m_1 + m_2)m_3\, R_1^2 \tag{4.58}$$

represents the spherical harmonic oscillator Hamiltonian, which in this case describes the motion of particle 3 with respect to the center of mass of particles 1 and 2 (or *vice versa*). In the

---

[78] The potential has to depend on the interparticle distance *only*. Such a situation is, however, common in physics.

above equations the following abbreviations have been used: $M = m_1 + m_2 + m_3$, $\mu_1 = \frac{m_1 m_2}{m_1+m_2}$, $\mu_2 = \frac{(m_1+m_2)m_3}{M}$.

We have something really interesting! Indeed we can manipulate the masses arbitrarily and therefore create links among some apparently unrelated systems. Note that if V is chosen as the Coulombic repulsion, Eq. (4.50) describes either the two-electron Hooke atom, i.e., harmonium[79] (then we put $m_1 = m_2 \ll m_3$) or a one-electron Hooke diatomic ($m_1 = m_2 \gg m_3$).

Four particles

In case of four particles a separable Hamiltonian has the form

$$\hat{H}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \hat{H}_{12}(\mathbf{r}_1, \mathbf{r}_2) + \hat{H}_{34}(\mathbf{r}_3, \mathbf{r}_4) \\ + k\left(m_1 m_3 r_{13}^2 + m_1 m_4 r_{14}^2 + m_2 m_3 r_{23}^2 + m_2 m_4 r_{24}^2\right).$$

New coordinates (left-hand side) are introduced:

$$\begin{aligned} \boldsymbol{\rho}_1 &= \mathbf{r}_1 - \mathbf{r}_2, \\ \boldsymbol{\rho}_2 &= \mathbf{r}_3 - \mathbf{r}_4, \\ \boldsymbol{\rho}_{34} &= \boldsymbol{\rho}_3 - \boldsymbol{\rho}_4, \\ \mathbf{R}_0 &= \frac{M_1 \boldsymbol{\rho}_3 + M_2 \boldsymbol{\rho}_4}{M_1 + M_2}, \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{\rho}_3 &= \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{M_1}, \\ \boldsymbol{\rho}_4 &= \frac{m_3 \mathbf{r}_3 + m_4 \mathbf{r}_4}{M_2}, \end{aligned}$$

where $M_1 = m_1 + m_2$, $M_2 = m_3 + m_4$. After the corresponding changes in the Laplacians we get

$$\hat{H}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \hat{h}_0(\mathbf{R}_0) + \hat{h}_1(\boldsymbol{\rho}_1) + \hat{h}_2(\boldsymbol{\rho}_2) + \hat{h}_{34}(\boldsymbol{\rho}_{34}), \tag{4.59}$$

where

$$\hat{h}_i(\boldsymbol{\rho}_i) = -\frac{\hbar^2}{2\mu_i}\Delta_{\rho_i} + V(\rho_i) + k\mu_i M_{3-i}\, \rho_i^2, \quad i = 1, 2, \tag{4.60}$$

---

[79] Harmonium represents the two-electron Hooke atom. A Hooke diatomic molecule means two heavy particles (nuclei) interacting by Coulomb forces, the same with electrons, but the heavy particle–light particle interactions are harmonic.

$$\hat{h}_{34}(\boldsymbol{\rho}_{34}) = -\frac{\hbar^2}{2\mathfrak{M}}\Delta_{\rho_{34}} + kM_1M_2\rho_{34}^2, \tag{4.61}$$

with $\mu_i = \frac{m_{2i-1}m_{2i}}{m_{2i-1}+m_{2i}}$, $\mathfrak{M} = \frac{M_1M_2}{M_1+M_2}$, and $\hat{h}_0(\boldsymbol{R}_0)$ describes the center-of-mass motion. If $m_1 = m_2 \gg m_3 = m_4$, we obtain a model of a homonuclear Hooke diatomic.

We are tempted now to go further and think about even larger numbers of particles. Maybe step by step we will conquer more and more until we reach a number of particles for which no separation is possible? Then we might say we did our best, but this is the limit... For the reader who shares our worry we have good news: *the separation is sometimes possible even for an infinite number of particles*!

N particles

Let us suppose we have K pairs of particles in a system of N particles, their interaction within each pair given by potential energy expressions[80] $V_i(\rho_i)$, $i = 1, 2, \ldots, K$ (where $\boldsymbol{\rho}_i = \boldsymbol{r}_{2i-1} - \boldsymbol{r}_{2i}$). The other interactions are described by quadratic forms of coordinates[81] (satisfying some conditions ensuring separability). The separation of variables gives K one-particle Hamiltonians,

$$\hat{h}_i(\boldsymbol{\rho}_i) = -\frac{\hbar^2}{2\mu_i}\Delta_{\rho_i} + V_i(\rho_i) + A_i\,\rho_i^2, \quad i = 1, 2, \ldots, K, \tag{4.62}$$

where A_i is a constant that depends on particles' masses and with $N - K - 1$ Hamiltonians describing spherical harmonic oscillators with properly defined coordinates (known as normal coordinates, see Chapter 7). We also get, of course, the one-particle Hamiltonian for the center-of-mass motion. In such a case of separability we obtain N Schrödinger equations, each one for one "particle."Solution of the problem of N interacting particles and a systematic procedure leading to the separation of variables has been given by my friend Jacek Karwowski, professor at Nicolaus Copernicus University (Torun, Poland).[82] The solutions already known in the liter-

---

[80] V_i with index i means that each of the two-particle potentials may have a different mathematical form.

[81] The assumption of the quadratic form may lead (provided some conditions are satisfied) to harmonic motions. We see once more a peculiar role of harmonic approximation in physics. They often lead to unexpectedly simple (and therefore beautiful) expressions.

[82] J. Karwowski, *Int. J. Quantum Chem.*, 108(2008)2253. Jacek Karwowski was the first to notice that the separability can be performed sequentially (by pairs of particles interacting through V_i), in this way including any number of particle pairs. He proved then that the remaining interactions being quadratic forms may in some cases be reduced to a canonical form (i.e., to normal modes, see Chapter 7) in such a way that does not pertain to the variables of already separated pairs. The first step is always feasible, the second one only if the quadratic form satisfies some conditions.

ature turn out to be special cases of this general one: the already described harmonium ($N = 3$), Hooke's H_2^+ molecule,[83] and Hooke's H_2 molecule.[84]

4.14 Charming SUSY and new solutions

How to find exact solutions for new systems? Our exciting story about SUSY is reserved for real fans of her quantum beauty, and certainly may be omitted by the readers with a more pragmatic than romantic attitude towards quantum mechanics.

We start from some banal observations (tension in our story will be built up gradually): two identical molecules have the same energy levels. What about two *different* systems? Can the sets of energy levels of these systems be identical? Well, it would be as if a zither and a jar would have the same set of vibrational normal modes. If this strange thing happened, we would guess that this comes maybe from some invisible common feature of these two objects. This feature might be called supersymmetry (SUSY) just to stress that the two systems are related to one another by a hidden symmetry operation.[85]

Theoretical physics already has met such a problem. There was a possibility to reduce the number of different kinds of interactions by introducing a relation between theoretical description of a fermion of spin s and a boson of spin $s \pm \frac{1}{2}$. This relation has been named supersymmetry. Such an idea looks very courageous, because the particles have very different properties (due to antisymmetry of wave functions for fermions and symmetry for bosons, see Chapter 1). In spite of that, such a relation has been introduced.[86]

---

[83] X. Lopez, J.M. Ugalde, L. Echevarria, E.V. Ludeña, *Phys. Rev.*, *A*74(2006)042504.

[84] E.V. Ludeña, X. Lopez, J.M. Ugalde, *J. Chem. Phys.*, 123(2005)024102.

[85] Quantum chemistry knows the intriguing notion of isospectral molecules, which have identical energy levels (we mean the Hueckel method, i.e., a simplified molecular orbital model, see Chapter 8). From this we have a long way to go to the same systems of realistic energy levels, and even further to the same spectra (e.g., in UV-VIS) of both molecules, which need the same transition probabilities. Nevertheless, we would not be astonished if, e.g., a piece of tiger skin and the marigold flower had the same color (coming from the same spectra), although probably the molecules responsible for this would be different.

My friend Leszek Stolarczyk remarked still another kind of symmetry in chemistry (unpublished results). Namely, the alternant hydrocarbons, defined in the Hueckel theory, despite the fact that they often do not have any spatial symmetry at all, have their energy level pattern symmetric with respect to a reference energy. We meet the same feature in the Dirac theory, if the electronic and positronic levels are considered. This suggests an underlying, not yet known, internal reason common to the alternant hydrocarbons as viewed in the Hueckel theory and the Dirac model, which seems to be related somehow to the notion of supersymmetry.

[86] It is not quite sure yet whether SUSY represents indeed a symmetry seen in Nature. Since in what is known as the standard model a rigorous SUSY does not appear, it is believed SUSY (if it exists at all) "is broken." If it would appear that SUSY exists for very high energies, this would be very important for unification of the electroweak and strong interactions in physics.

4.14.1 SUSY partners

The supersymmetry in quantum mechanics[87] will be shown in the simplest case possible: a single particle of mass m moving along the x axis. The solution of the Schrödinger equation for the harmonic oscillator by using the creation and annihilation operators will shine, exposing new and unexpected beauty.

Similarly to what we have done for the quantum oscillator, let us introduce the operator

$$\hat{A} = \frac{\hbar}{\sqrt{2m}}\frac{d}{dx} + W(x), \tag{4.63}$$

together with its Hermitian conjugate[88]

$$\hat{A}^{\dagger} = -\frac{\hbar}{\sqrt{2m}}\frac{d}{dx} + W(x). \tag{4.64}$$

Function $W(x)$ is called a *SUSY superpotential.*

From these two operators two Hamiltonians may be constructed, i.e.,

$$\hat{H}_1 = \hat{A}^{\dagger}\hat{A} \tag{4.65}$$

and

$$\hat{H}_2 = \hat{A}\hat{A}^{\dagger}. \tag{4.66}$$

After inserting (4.63) and (4.64) into (4.65) we obtain[89]

$$\hat{H}_1 f = \hat{A}^{\dagger}\hat{A} f =$$

$$\left(-\frac{\hbar}{\sqrt{2m}}\frac{d}{dx} + W(x)\right)\left(\frac{\hbar}{\sqrt{2m}}\frac{d}{dx} + W(x)\right) f =$$

$$-\frac{\hbar}{\sqrt{2m}}\frac{d}{dx}\frac{\hbar}{\sqrt{2m}}\frac{df}{dx} - \frac{\hbar}{\sqrt{2m}}\frac{d(Wf)}{dx} + W\frac{\hbar}{\sqrt{2m}}\frac{df}{dx} + W^2 f =$$

$$-\frac{\hbar^2}{2m}\frac{d^2 f}{dx^2} + W^2 f - f\frac{\hbar}{\sqrt{2m}}\frac{dW}{dx} = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V_1\right) f.$$

---

[87] The idea of supersymmetry has been introduced to quantum mechanics by C.V. Sukumar, *J. Phys. A*, 18(1985)*L*57. My inspiration for writing this section came from a beautiful 2004 lecture by Avinash Khare, who also belongs to the pioneers of SUSY in quantum mechanics (F. Cooper, A. Khare, U. Sukhatme, Supersymmetry and quantum mechanics, World Scientific, Singapore (2001)).

[88] The derivation is almost identical to that shown for the harmonic oscillator.

[89] In order to avoid a mistake, we will show all this when acting on an arbitrary function f of class Q.

Hence we have

$$\hat{H}_1 = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V_1, \tag{4.67}$$

where

$$V_1 = W^2 - \frac{\hbar}{\sqrt{2m}}W'. \tag{4.68}$$

Similarly we get

$$\hat{H}_2 = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V_2, \tag{4.69}$$

$$V_2 = W^2 + \frac{\hbar}{\sqrt{2m}}W'. \tag{4.70}$$

The potentials V_1 and V_2 are known as *SUSY partner potentials*.

4.14.2 Relation between the SUSY partners

Let us write down the Schrödinger equation for Hamiltonian $\hat{H}_1$ (the SUSY partner wave functions will be numbered by superscripts):

$$\hat{H}_1\psi_n^{(1)} = \hat{A}^\dagger\hat{A}\psi_n^{(1)} = E_n^{(1)}\psi_n^{(1)}.$$

After multiplying it from the left by $\hat{A}$, we get something interesting: $\hat{A}\hat{A}^\dagger\left(\hat{A}\psi_n^{(1)}\right) = E_n^{(1)}(\hat{A}\psi_n^{(1)})$, which means $\hat{H}_2\left(\hat{A}\psi_n^{(1)}\right) = E_n^{(1)}(\hat{A}\psi_n^{(1)})$. After introducing a symbol for the wave functions of Hamiltonian $\hat{H}_2$ we have $\hat{H}_2\psi_n^{(2)} = E_n^{(1)}\psi_n^{(2)}$, where $\psi_n^{(2)} \equiv \hat{A}\psi_n^{(1)}$.

It turns out therefore that each energy eigenvalue of $\hat{H}_1$ represents also an energy eigenvalue of $\hat{H}_2$, while the wave function corresponding to $\hat{H}_1$, after transforming it by applying operator $\hat{A}$, becomes a wave function of $\hat{H}_2$! We transform similarly the Schrödinger equation for $\hat{H}_2$,

$$\hat{H}_2\psi_n^{(2)} = \hat{A}\hat{A}^\dagger\psi_n^{(2)} = E_n^{(2)}\psi_n^{(2)},$$

through multiplying by $\hat{A}^\dagger$: $\hat{A}^\dagger\hat{A}(\hat{A}^\dagger\psi_n^{(2)}) = E_n^{(2)}(\hat{A}^\dagger\psi_n^{(2)})$. We obtain $\hat{H}_1(\hat{A}^\dagger\psi_n^{(2)}) = E_n^{(2)}(\hat{A}^\dagger\psi_n^{(2)})$. Therefore, we have the splendid beauty of SUSY right before our eyes:

$$\psi_n^{(2)} \equiv \hat{A}\psi_n^{(1)} \qquad \text{and} \qquad \psi_n^{(1)} \equiv \hat{A}^\dagger \psi_n^{(2)}. \tag{4.71}$$

An eigenvalue of $\hat{H}_1$ represents also an eigenvalue of Hamiltonian $\hat{H}_2$. Also, if $\psi_n^{(1)}$ is an eigenfunction of $\hat{H}_1$, then function $\hat{A}\psi_n^{(1)}$ is also an eigenfunction of $\hat{H}_2$. On the other hand, if $\psi_n^{(2)}$ is an eigenfunction of $\hat{H}_2$, then $\hat{A}^\dagger\psi_n^{(2)}$ represents also an eigenfunction of $\hat{H}_1$.

There is one little thing we forgot to mention: the SUSY may have a defect... Namely, the proof given above fails if for some reason $\hat{A}\psi_n^{(1)} = 0$ (case 1) or $\hat{A}^\dagger\psi_n^{(2)} = 0$ (case 2). Indeed, if this happened this would mean an eigenfunction of $\hat{H}_2$ or $\hat{H}_1$, respectively, would be 0 everywhere, which is unacceptable in quantum mechanics (wave functions must be of class Q). Will such a case have a chance to happen? First of all, please note[90] that $E_n^{(i)} \geq 0$.

- *Case* 1. From $\hat{A}\psi_n^{(1)} = 0$ it follows that also $\hat{A}^\dagger\hat{A}\psi_n^{(1)} = 0$, which means $\hat{H}_1\psi_n^{(1)} = 0 = E_n^{(1)}\psi_n^{(1)}$. Since $E_n^{(i)} \geq 0$, this may happen for $n = 0$ only (ground-state nondegeneracy), and on top of that we have to have $E_0^{(1)} = 0$. The SUSY partner of this state is $\psi_0^{(2)} = \hat{A}\psi_0^{(1)} = 0$ (because in the case considered $\hat{A}\psi_0^{(1)} \equiv 0$ everywhere). This however means that the SUSY partner $\psi_0^{(2)}$ simply does not exist in this case.
- *Case* 2. We have $\hat{A}^\dagger\psi_n^{(2)} = 0$, which leads to $\hat{A}\hat{A}^\dagger\psi_n^{(2)} = 0$, and this means $\hat{H}_2\psi_n^{(2)} = 0 = E_n^{(2)}\psi_n^{(2)}$. Since $E_n^{(i)} \geq 0$, this may happen only if $n = 0$ and in addition we should have $E_0^{(2)} = 0$. The SUSY partner of this state $\psi_0^{(1)} = \hat{A}^\dagger\psi_0^{(2)} = 0$ (because by definition $\hat{A}^\dagger\psi_0^{(2)} \equiv 0$). This means the SUSY partner $\psi_0^{(1)}$ does not exist.

Therefore,

at any complication (either $\hat{A}\psi_0^{(1)} = 0$ or $\hat{A}^\dagger\psi_0^{(2)} = 0$), if it happens, the ground-state Schrödinger equation has a *single* SUSY partner only. The conditions $\hat{A}\psi_0^{(1)} = 0$ and $\hat{A}^\dagger\psi_0^{(2)} = 0$ cannot be satisfied simultaneously.

---

[90] This follows from the equation for the eigenvalues: $\hat{H}_1\psi_n^{(1)} = \hat{A}^\dagger\hat{A}\psi_n^{(1)} = E_n^{(1)}\psi_n^{(1)}$. Indeed, let us make the scalar product of both sides with function $\psi_n^{(1)}$: $E_n^{(1)} = \left\langle \psi_n^{(1)} | \hat{H}_1\psi_n^{(1)} \right\rangle = \left\langle \psi_n^{(1)} | \hat{A}^\dagger\hat{A}\psi_n^{(1)} \right\rangle = \left\langle \hat{A}\psi_n^{(1)} | \hat{A}\psi_n^{(1)} \right\rangle \geq 0$, because the square of the length of vector $\hat{A}\psi_n^{(1)}$ is nonnegative.

Equation $\hat{A}\psi_0^{(1)} = 0$ may be used to calculate $\psi_0^{(1)}$ if $W(x)$ is known or to find $W(x)$ if $\psi_0^{(1)}$ is known.[91] In the first case we have (N stands for the normalization constant)[92]

$$\psi_0^{(1)} = N \exp[-\frac{\sqrt{2m}}{\hbar} \int_{-\infty}^{x} W(y)]\mathrm{d}y. \tag{4.72}$$

In the second case we have[93]

$$W(x) = -\frac{\hbar}{\sqrt{2m}} \frac{\frac{\mathrm{d}}{\mathrm{d}x}\psi_0^{(1)}}{\psi_0^{(1)}}. \tag{4.73}$$

Example 4 (Harmonic oscillator ⟷ harmonic oscillator). Let us see what kind of SUSY partner has a harmonic oscillator. First, let us shift the energy scale[94] so the ground state has energy 0. This means the shifted potential has the form

$$V_1(x) = \frac{1}{2}kx^2 - \frac{1}{2}h\nu = \frac{1}{2}kx^2 - \frac{1}{2}\hbar\omega. \tag{4.74}$$

Now, from Eq. (4.73) and the expression for the ground-state wave function for the harmonic oscillator, we find the superpotential from the formula (4.73)

$$W(x) = -\frac{\hbar}{\sqrt{2m}} \frac{\frac{\mathrm{d}}{\mathrm{d}x}\psi_0^{(1)}}{\psi_0^{(1)}} = \tag{4.75}$$

$$-\frac{\hbar}{\sqrt{2m}} \frac{-\sqrt{\frac{km}{\hbar^2}}x \exp(-\frac{1}{2}\sqrt{\frac{km}{\hbar^2}}x^2)}{\exp(-\frac{1}{2}\sqrt{\frac{km}{\hbar^2}}x^2)} = \sqrt{\frac{k}{2}}x. \tag{4.76}$$

Then from Eqs. (4.68) and (4.70) we find the SUSY partners: $V_1 = W^2 - \frac{\hbar}{\sqrt{2m}}W' = \frac{1}{2}kx^2 - \frac{\hbar}{\sqrt{2m}}\sqrt{\frac{k}{2}} = \frac{1}{2}kx^2 - \frac{1}{2}\hbar\omega$, which agrees with Eq. (4.74), and a new partner potential

$$V_2 = W^2 + \frac{\hbar}{\sqrt{2m}}W' = \frac{1}{2}kx^2 + \frac{\hbar}{\sqrt{2m}}\sqrt{\frac{k}{2}} = \frac{1}{2}kx^2 + \frac{1}{2}\hbar\omega.$$

---

[91] In such a case we will force one of the SUSY partners to have one level more than the other.

[92] Indeed, let us check: $\hat{A}\psi_0^{(1)} = \left(\frac{\hbar}{\sqrt{2m}}\frac{d}{dx} + W(x)\right)\psi_0^{(1)} =$
$N\frac{\hbar}{\sqrt{2m}} \exp[-\frac{\sqrt{2m}}{\hbar}\int_{-\infty}^{x} W(y)]\mathrm{d}y \left(-\frac{\sqrt{2m}}{\hbar}\right) W(x) + W(x)\psi_0^{(1)} = -W(x)\psi_0^{(1)} + W(x)\psi_0^{(1)} = 0.$

[93] We check: the right-hand side is equal to
$-\frac{\hbar}{\sqrt{2m}}\frac{\frac{\mathrm{d}}{\mathrm{d}x}\psi_0^{(1)}}{\psi_0^{(1)}} = -\frac{\hbar}{\sqrt{2m}}\frac{-\frac{\sqrt{2m}}{\hbar}W(x)\psi_0^{(1)}}{\psi_0^{(1)}} = W(x)$, which is equal to the left-hand side. This fascinating formula says: “*Show me the logarithmic derivative of the ground-state wave function and I will tell you what system we are talking about.*”.

[94] The world looks the same after someone decides to count energy with respect to another point on the energy axis.

As we see the potential V_2 also represents a harmonic oscillator potential! The only difference, a shift on the energy scale, is not important.[95] We see once more an outstanding role of harmonic oscillators in physics.

Example 5 (Particle in a box ⟷ cotangent potential). At the beginning of this chapter (p. 189) the problem of the particle in a box was solved. Now we will treat this as one of the SUSY partners, say, the first one. Our goal will be to find the second SUSY partner. Similarly as for the harmonic oscillator a while ago, we will shift the energy (this time by $-\frac{h^2}{8mL^2}$) in order to have the ground state of the particle in a box at energy 0. The formula for the energy levels in the new energy scale and using the SUSY-type notation for the eigenvalues and the eigenfunctions (as well as counting the ground state as corresponding to $n=0$) has the form

$$E_n^{(1)} = \frac{(n+1)^2 h^2}{8mL^2} - \frac{h^2}{8mL^2} = \frac{n\,(n+2)\,h^2}{8mL^2}, \quad n = 0, 1, 2, \ldots,$$

while the normalized wave function is

$$\psi_n^{(1)} = \sqrt{\frac{2}{L}} \sin \frac{(n+1)\pi}{L} x.$$

From Eq. (4.73) we find the superpotential $W(x)$ as

$$W(x) = -\frac{\hbar}{\sqrt{2m}} \frac{\frac{\mathrm{d}}{\mathrm{d}x}\psi_0^{(1)}}{\psi_0^{(1)}} = -\frac{\hbar\pi}{\sqrt{2m}L} \operatorname{ctg}(\frac{\pi}{L}x),$$

while from Eqs. (4.68) and (4.70) we obtain the potential energies for the two SUSY partners

$$V_1(x) = W^2 - \frac{\hbar}{\sqrt{2m}} W' = \frac{\hbar^2\pi^2}{2mL^2}\left[\operatorname{ctg}(\frac{\pi}{L}x)\right]^2 - \frac{\hbar}{\sqrt{2m}}\left[-\frac{\hbar}{\sqrt{2m}}\frac{\pi}{L}\right]\frac{\mathrm{d}\operatorname{ctg}(\frac{\pi}{L}x)}{\mathrm{d}x} =$$

$$\frac{\hbar^2\pi^2}{2mL^2}\left[\operatorname{ctg}(\frac{\pi}{L}x)\right]^2 + \frac{\hbar^2\pi^2}{2mL^2}\frac{-1}{\sin^2\frac{\pi}{L}x} = \frac{\hbar^2\pi^2}{2mL^2}\left[\operatorname{ctg}^2(\frac{\pi}{L}x) - \frac{1}{\sin^2\frac{\pi}{L}x}\right] =$$

$$\frac{\hbar^2\pi^2}{2mL^2}\left[\frac{\cos^2\frac{\pi}{L}x - 1}{\sin^2\frac{\pi}{L}x}\right] = -\frac{\hbar^2\pi^2}{2mL^2} = -\frac{h^2}{8mL^2} = \text{const},$$

$$V_2(x) = W^2 + \frac{\hbar}{\sqrt{2m}} W' = \frac{\hbar^2\pi^2}{2mL^2}\left[\operatorname{ctg}(\frac{\pi}{L}x)\right]^2 + \frac{\hbar}{\sqrt{2m}}\left[-\frac{\hbar}{\sqrt{2m}}\frac{\pi}{L}\right]\frac{\mathrm{d}\operatorname{ctg}(\frac{\pi}{L}x)}{\mathrm{d}x} =$$

---

[95] From a very formal point of view, the second SUSY partner is devoid of the ground state of the first partner, as discussed earlier. However, since a harmonic oscillator has an infinite number of equidistant levels, a new SUSY partner means only a shift of all energy levels up by $\hbar\omega$; therefore in this particular case the two SUSY partners represent the same system.

$$\frac{\hbar^2\pi^2}{2mL^2}\left[\mathrm{ctg}(\frac{\pi}{L}x)\right]^2-\frac{\hbar^2\pi^2}{2mL^2}\frac{-1}{\sin^2\frac{\pi}{L}x}=\frac{\hbar^2\pi^2}{2mL^2}\frac{\cos^2\frac{\pi}{L}x+1}{\sin^2\frac{\pi}{L}x}=\frac{\hbar^2\pi^2}{2mL^2}\left[2\mathrm{ctg}^2\frac{\pi}{L}x+1\right].$$

The first formula says that $V_1(x)$ has a constant value in $[0,L]$ (this corresponds to the particle-in-a-box problem). The bottom of this box is at energy $-\frac{h^2}{8mL^2}$, because only then the ground-state energy will equal 0 (as was assumed). Its SUSY partner corresponds to $V_2(x)=\frac{\hbar^2\pi^2}{2mL^2}\left[2\mathrm{ctg}^2\frac{\pi}{L}x+1\right]$. As it follows from Eq. (4.71), the wave functions corresponding to $V_2(x)$ can be calculated from the particle-in-a-box wave functions by applying operator $\hat{A}$. Therefore,

$$\psi_1^{(2)}=\hat{A}\psi_1^{(1)}=\sqrt{\frac{2}{L}}\left(\frac{\hbar}{\sqrt{2m}}\frac{d}{dx}+W(x)\right)\sin\frac{2\pi}{L}x=$$

$$\sqrt{\frac{2}{L}}\left(\frac{\hbar}{\sqrt{2m}}\frac{2\pi}{L}\cos\frac{2\pi}{L}x-\frac{\hbar\pi}{\sqrt{2m}L}\mathrm{ctg}(\frac{\pi}{L}x)\sin\frac{2\pi}{L}x\right)=$$

$$\sqrt{\frac{2}{L}}\frac{\hbar}{\sqrt{2m}}\frac{2\pi}{L}\left(\cos\frac{2\pi}{L}x-\mathrm{ctg}(\frac{\pi}{L}x)\sin\frac{\pi}{L}x\cos\frac{\pi}{L}x\right)=$$

$$\sqrt{\frac{2}{L}}\frac{\hbar}{\sqrt{2m}}\frac{2\pi}{L}\left(\cos\frac{2\pi}{L}x-\cos^2\frac{\pi}{L}x\right)=N'\sin^2\frac{\pi}{L}x,$$

where N' is a normalization constant. Similarly one gets $\psi_2^{(2)}=N''\sin x\sin 2x$. Fig. 4.28 shows $V_1(x)$ and $V_2(x)$ as well as the energy levels corresponding to both SUSY partners. For each of the partners the consecutive energy levels correspond to the wave functions with increasing number of nodes. As one can see the SUSY partner potentials differ widely: $V_1(x)$ is sharp-edged, $V_2(x)$ resembles this a bit, but is cleverly rounded in order to repel the eigenstates corresponding to $V_1(x)$ (one level up), precisely to the positions of the partner's excited states.

The two SUSY partners differ by one (ground-state) level.

Our story about beautiful SUSY has a happy end: it was possible to find about a dozen of SUSY partners (for the solutions of the Schrödinger equation known earlier).

4.15 Beacons and pearls of physics

Sometimes students, fascinated by handy computers available nowadays, tend to treat the simple systems described in this chapter as primitive and out-of-date. A professor has taken them from the attic and, after dusting off, shows them in a class, whilst outside computers with high-level science, splendid programs, and colorful graphs await. This is wrong. The simple systems

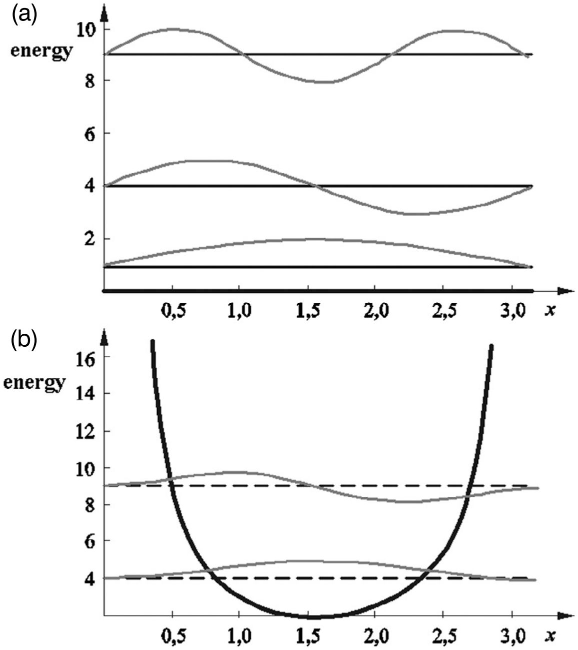


Fig. 4.28. The SUSY partners lead to the same set of energy levels, possibly except one (the ground state of one of the partners). One of the SUSY partners corresponds to the particle in a box (see p. 189), as shown in (a) together with the three lowest energy levels $1, 4, 9$. In order to obtain the simplest formulae (and therefore get rid of the unnecessary luggage), $\frac{\hbar^2\pi^2}{2mL^2}$ has been used as the energy unit and, additionally, the box length has been chosen equal to π. The original energy scale has been used in the figure (in which the potential energy for the particle in the box is equal to 0). The corresponding wave functions have been plotted for the three lowest levels (each plot put artificially at the height of the corresponding energy level). The functions have zero value beyond $(0, L)$. (b) shows $V_2(x)$, calculated as the SUSY partner, together with the two lowest energy levels 4 and 9 and the corresponding wave functions (exposed at the height of their levels).

considered in this chapter correspond to extremely rare *exact solutions of the Schrödinger equation* and are, therefore, precious pearls of physics by themselves. Nobody will give a better solution, the conclusions are a hundred percent sure. It is true that they all (except the hydrogen atom) correspond to some idealized systems.[96] There is no such a thing as an unbreakable spring (e.g., harmonic oscillator) or a rotator that does not change its length, etc. And yet these problems represent our firm ground or the beacons of our native land. Reading the present chapter we have been preparing our ship for a long voyage. When confronted with the surprises of new lands and trying to understand them

---

[96] Like Platonic ideal solids.

> the *only* points of reference or the beacons which tell us about *terra firma* will be the problems for which analytical solutions have been found.

Summary

Exact analytical solutions[97] to the Schrödinger equation play an important role as an organizer of our quantum mechanical experience. Such solutions have only been possible to obtain for some idealized objects. This is of great importance for the interpretation of *approximate* solutions for real systems. Another great feature of exact solutions is that they have an extremely wide range of applications: they are useful independently of whether we concentrate on an electron in an atom or in a molecule, a nucleon in a nucleus, a molecule as an entity, etc.

The main features of the solutions are:

- **Free particle.** The particle may be described as the superposition of the state $\exp(i\kappa x)$, corresponding to the particle moving right (positive values of x), and the state $\exp(-i\kappa x)$, corresponding to the particle moving left. Both states correspond to the same energy (and opposite momenta $\pm\hbar\kappa$).
- **Particle in a box**. We consider first a particle in a one-dimensional box, i.e., the particle is confined to section $[0, L]$ with potential energy (for a particle of mass m and coordinate x) equal to zero and ∞ outside the section. Such a potential forces the wave function to be nonzero only within the section $[0, L]$. We solve the elementary Schrödinger equation and obtain $\Psi = A \sin \kappa x + B \cos \kappa x$, where $\kappa^2 = \frac{2mE}{\hbar^2}$. Quantization appears in a natural way from the condition of continuity for the wave function at the boundaries: $\Psi(0) = 0$ and $\Psi(L) = 0$. These two conditions give the expression for the energy levels $E_n = \frac{n^2h^2}{8mL^2}$ and for the wave functions $\Psi_n = \sqrt{\frac{2}{L}} \sin \frac{n\pi}{L} x$ with quantum number $n = 1, 2, \ldots$. Conclusion: the successive energy levels are more and more distant and the wave function is simply a section of the sine function (with 0 value at the ends).
- **Tunneling effect**. We have a particle of mass m and a rectangular barrier (section $[0, a]$, width a, and height V_0). Beyond this section the potential energy is zero. The particle comes from the negative x values and has energy $E < V_0$. A classical particle would be reflected from the barrier. However, for the quantum particle:
  - the transmission coefficient is nonzero,
  - the passage of a large energy particle is easier,
  - a narrower barrier means larger transmission coefficient,
  - the higher the barrier, the smaller the transmission coefficient.

  The first feature is the most sensational, the others are intuitively quite acceptable.
- **Resonances over the barrier.** A particle with the energy $E > V_0$ either bounces off the barrier or passes over the barrier. For some particular energies there are 100% sure transmissions of the particle. These resonance states are at the energies which correspond to the *eigenvalues of a potential well with the shape of the barrier (i.e., as in the particle in a box) and placed on top of the barrier*!
- **Resonance states at double barrier.** It turns out that (for a given interbarrier distance and at energies lower than the barrier) *there are some "magic" energies of the particle (resonance energies), at*

---

[97] To distinguish from accurate solutions (i.e., received with a desired accuracy).

which the transmission coefficient is equal to 100%. The magic energies correspond to the stationary states that would be for a particle in a box a little longer than the interbarrier distance. The resonance states are also over the barrier and ensure a transmission coefficient equal to 100%, whereas other energies may lead to reflection of the particle, even if they are larger than the barrier height.

- **Harmonic oscillator**. A single particle of mass m coupled by a harmonic spring (with force constant k) corresponds to potential energy $V = \frac{kx^2}{2}$. We obtain quantization of the energy, i.e., $E_v = h\nu\left(v + \frac{1}{2}\right)$, where the vibrational quantum number $v = 0, 1, 2, \ldots$, and the angular frequency $\omega = 2\pi\nu = \sqrt{\frac{k}{m}}$. We see that *the nondegenerate energy levels are equidistant,* and their distance is larger for a larger force constant and smaller mass. The wave function has the form of a Gaussian factor and a polynomial of degree v. The polynomial ensures the proper number of nodes, while the Gaussian factor damps the plot to zero for large displacements from the particle equilibrium position. The harmonic oscillator may be viewed (Chapter 6) as a model (for small displacements) for two masses bound by a harmonic spring (like a diatomic molecule).
- **Morse oscillator**. The harmonic oscillator does not allow for the breaking of the spring connecting two particles, while the Morse oscillator admits dissociation. This is very important, because real diatomic molecules resemble the Morse rather than the harmonic oscillator. The solution for the Morse oscillator has the following features:
  - energy levels are nondegenerate,
  - their number is finite,
  - for large well depths the low energy levels tend to the energy levels of the harmonic oscillator (the levels are nearly equidistant),
  - the higher the energy level, the larger the displacement from the equidistant situation (the energy levels get closer),
  - the wave functions, especially those corresponding to deep-lying levels, are very similar to the corresponding ones of the harmonic oscillator,[98] but they do not exhibit the symmetry.[99]
- **Rigid rotator**. This is a system of two masses m_1 and m_2 that keeps their distance R fixed. After separating the center-of-mass motion (Appendix J on p. 691) we obtain an equation of motion for a single particle of mass equal to the reduced mass μ moving on a sphere of radius R (position given by angles θ and ϕ). The energy is determined by the quantum number $J = 0, 1, 2, \ldots$ and is equal to $E_J = J(J+1)\frac{\hbar^2}{2\mu R^2}$. As we can see:
  - there is an infinite number of energy levels,
  - the separation of the energy levels increases with the energy (similar to the particle-in-a-box problem),
  - the separation is smaller for larger masses,
  - the separation is smaller for longer rotators.

  The wave functions are the spherical harmonics $Y_J^M(\theta, \phi)$, which for low J are very simple and for large J complicated trigonometric functions. The integer quantum number M satisfies the relation $|M| \leq J$. The energy levels are, therefore, $(2J+1)$-tuply degenerate.
- **Hydrogen-like atom**. We have an electron and a nucleus of charges $-e$ and $+Ze$, respectively, or -1 and $+Z$ in a.u. The wave function is labeled by three quantum numbers: principal $n = 1, 2, \ldots$,

---

[98] Despite the fact that the formula itself is very different.

[99] The wave functions for the harmonic oscillator are either even or odd with respect to the inversion operation ($x \rightarrow -x$).

azimuthal $l = 0, 1, \ldots(n-1)$, and magnetic $m = -l, (-l+1), \ldots, 0, \ldots l$. The energy in a.u. is given by the formula[100] $E_n = -Z^2/(2n^2)$. The wave function represents the product of a polynomial (of r), an exponential function decreasing with r, and a spherical harmonic $Y_l^m(\theta, \phi)$, where r, θ, ϕ are the spherical coordinates of the electron, and the nucleus is at the origin. The wave functions that correspond to low energies are denoted by the symbols nl_m (with s for $l = 0$, p for $l = 1$, etc.): $1s$, $2s$, $2p_0$, $2p_1$, $2p_{-1}$, ... The degeneracy of the n-th level is equal to n^2.

- **Hooke helium atom**. In this peculiar helium atom the electrons are attracted to the nucleus by harmonic strings (of equal strength) of equilibrium length equal to zero. For $k = \frac{1}{4}$ an exact analytical solution exists. The exact wave function is a product of two Gaussian functions and a simple factor, $\left(1 + \frac{1}{2}r_{12}\right)$, that correlates the motions of the two electrons.
- **Hooke molecules**. If one has K pairs of particles (with the intrapair interaction of any kind) in a system of N particles and the other interactions are described by quadratic forms of coordinates satisfying some conditions, the separation of variables is possible and gives K one-particle Hamiltonians and $N - K - 1$ Hamiltonians describing spherical harmonic oscillators with properly defined coordinates (known as normal coordinates, see Chapter 7). Together with the one-particle Hamiltonian for the center-of-mass motion we obtain N Schrödinger equations, each one for one “particle” only.
- **Supersymmetry (SUSY)**. For a known exact solution of the Schrödinger equation for a system one can define another system with the same spectrum of the eigenvalues (possibly except the ground-state eigenvalue). Such a pair of the systems is known as the supersymmetry (SUSY) partners.

Main concepts, new terms

annihilation operator (p. 219)
associated Laguerre polynomials (p. 232)
associated Legendre polynomials (p. 229)
binding energy (p. 228)
Bohr first orbit (p. 234)
box with ends (p. 189)
correlation factors (p. 244)
creation operator (p. 219)
current (p. 205)
cyclic box (p. 193)
dissociation energy (p. 228)
FEMO (p. 192)
free particle (p. 189)
harmonic oscillator (p. 217)
Hermite polynomials (p. 217)
Hooke atom (p. 243)
Hooke molecule (p. 243)
hydrogen-like atom (p. 232)
Laguerre polynomials (p. 232)
Legendre polynomials (p. 229)
Morse oscillator (p. 224)
nanotube (p. 200)
particle in a box (p. 189)
phonon (p. 220)
phonon vacuum (p. 220)
positronium (p. 242)
radial density (p. 234)
resonance state (p. 205)
rigid rotator (p. 229)
spherical harmonics (p. 229)
SUSY partner (p. 251)
SUSY superpotential (p. 250)
transmission coefficient (p. 205)
tunneling effect (p. 203)

---

[100] An infinite number of levels; the mass of the nucleus is set to infinity.

From the research front

A field like the one discussed in the present chapter seems to be definitely closed. We have been lucky enough to solve some simple problems that could be solved, but others are just too complicated. This is not true. For several decades it has been possible to solve a series of nonlinear problems, thought in the past to be hopeless. What decides success is: choice of the problem, quality of researchers, courage, etc.[101] It is worth noting that there are also attempts at a systematic search for promising systems to solve.

Ad futurum

It seems that the number of exactly solvable problems will continue to increase, although the pace of such research will be low. If exactly solvable problems were closer to practice of physics, this might greatly accelerate research.

Additional literature

J. Dvořák, L. Skála, "Analytical Solutions of the Schrödinger Equation. Ground State Energies and Wave Functions," *Collect. Czech. Chem. Commun.*, 63(1998)1161.

Very interesting article with the character of a synthesis. Many potentials[102] leading to exactly solvable problems are presented following a uniform theoretical approach. The authors also give their own two generalizations.

F. Cooper, A. Khare, U. Sukhatme, "Supersymmetry and quantum mechanics," World Scientific, Singapore (2001).

Questions

1. Consider a free particle confined on the x axis (chosen from left to right). The wave function $\Psi(x) = A\exp(i\kappa x) + B\exp(-i\kappa x)$, where $A, B, \kappa > 0$ represent constants:
   a. describes a particle which moves right with probability A and left with probability B.
   b. when you measure the particle's momentum, you will always obtain $\hbar\kappa$.
   c. the constants A and B ensure the normalization of wave function Ψ.
   d. describes a particle which moves right with probability proportional to $|A|^2$ and left with probability proportional to $|B|^2$.
2. Consider a particle in a box of length L:
   a. the wave function normalization constant $-\sqrt{\frac{2}{L}}$ is incorrect.
   b. the wave function is single-valued at any point of space.

---

[101] Already the Morse potential looks very difficult to manage, to say nothing about the harmonic helium atom.

[102] Among them six are not discussed in the present textbook.

   c. the energy quantization follows from the fact that the wave function represents an eigenfunction of the Hamiltonian.
   d. the quantization of energy follows from the continuity of those wave functions that satisfy the Schrödinger equation.

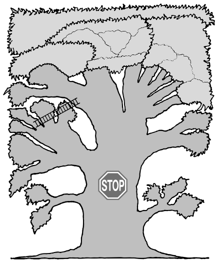
3. Consider the energy eigenvalues (E_n) and eigenfunctions (ψ_n) for the particle in a box:
   a. a heavier particle has lower energy.
   b. $\int \psi_n^* \psi_m \, dx = 0$ for $n \neq m$.
   c. increasing box length results in lower energy levels and smaller distances between them.
   d. the number of the energy eigenvalues is infinite.
4. If E stands for the kinetic energy of the particle encountering a rectangular barrier of height V_0:
   a. the particle may bounce off the barrier, but only if $E < V_0$.
   b. when $E < V_0$ the particle will pass more easily through the barrier for larger E.
   c. the smaller the barrier's width, the more easily the particle will pass through the barrier.
   d. a heavier particle will pass more easily through the barrier.
5. Consider a harmonic oscillator:
   a. the square of the modulus of the wave function does not vanish except for $x \to \pm\infty$.
   b. it has the infinite number of equidistant nondegenerate energy levels.
   c. its wave functions are all symmetric with respect to the transformation $x \to -x$.
   d. if a deuteron oscillates in a parabolic energy well instead of a proton, the separation of the energy levels doubles.
6. Consider a rigid rotator ($J = 0, 1, 2, \ldots$ is the rotational quantum number):
   a. its energy levels are equidistant and degenerate for $J \neq 0$.
   b. the degeneracy of the energy level corresponding to J is equal to 2.
   c. the rotational energy is proportional to $J(J+1)$.
   d. if the two rotating masses double their mass, the separation between the energy levels decreases by a factor of 2.
7. Consider a hydrogen atom:
   a. the highest probability density of finding the electron described by the wave function $1s$ is on the nucleus.
   b. the number of the node surfaces for the wave functions $1s, 2s, 3s, 2p_y, 3d_{xz}$ is equal to $0, 1, 2, 1, 2$.
   c. the wave functions $1s, 2s, 3s$ are spherically symmetric, while $2p_x$ represents an eigenfunction of the Hamiltonian, of the square of the angular momentum, and of the z component of the angular momentum.
   d. the orbital $3d_{3z^2-r^2}$ has cylindrical symmetry.
8. Consider a hydrogen atom (n, l, m stand for the quantum numbers, ψ_{nlm} represent orbitals):
   a. the wave function $2\psi_{431} - \frac{i}{2}\psi_{42-1}$ is an eigenfunction of the Hamiltonian.
   b. the wave function $\frac{1}{2i}\psi_{320} + \frac{1}{2}\psi_{42-1}$ is an eigenfunction of the square of the angular momentum.
   c. the wave function $\psi_{310} - 10\psi_{320}$ is an eigenfunction of the Hamiltonian and of the z component of the angular momentum.
   d. the wave function $\psi_{31-1} + 12\psi_{320} - i\psi_{300}$ represents an eigenfunction of the Hamiltonian.
9. Harmonium (r_i for $i = 1, 2$ stands for the nucleus–electron distance for the i-th electron, r_{12} represents the electron–electron distance, N denotes the normalization constant):

  a. represents the helium atom with all the Coulomb interactions replaced by the harmonic attraction.
  b. for the harmonic spring force constant $k = \frac{1}{4}$ the Schrödinger equation has an exact solution.
  c. the number of mistakes made in the wave function (for $k = \frac{1}{4}$) formula $\psi(\mathbf{r}_1, \mathbf{r}_2) = N\left(1 - \frac{1}{2}r_{12}\right)\exp[+\frac{1}{4}\left(r_1^2 + r_2^2\right)]$ is equal to 1.
  d. the probability of finding the two electrons on the opposite sides of the nucleus is larger than on the same side.

10. Consider SUSY:
  a. the supersymmetry relation pertains to two systems with the Hamiltonians $\hat{H}_1 = \hat{A}^\dagger \hat{A}$ and $\hat{H}_2 = \hat{A}\hat{A}^\dagger$.
  b. an eigenvalue of $\hat{H}_1$ represents always an eigenvalue of $\hat{H}_2$.
  c. if $\psi_n^{(1)}$ is an eigenfunction of $\hat{H}_1$, then $\hat{A}\psi_n^{(1)} \neq 0$ is an eigenfunction of $\hat{H}_2$.
  d. $\hat{H}_1$ and $\hat{H}_2$ differ by the SUSY partner potentials.

Answers

1d, 2bd, 3abcd, 4bc, 5b, 6cd, 7abd, 8abcd, 9bd, 10acd



CHAPTER 5

Three Fundamental Approximate Methods

> *Even the upper end of the river believes in the ocean.*
> ***William Stafford***

Where are we?

We are moving upwards in the central parts of the TREE trunk.

An example

Suppose we are interested in properties of the ammonia molecule in its ground state, e.g., we would like to know the mean value of the nitrogen–hydrogen distance. Only quantum mechanics gives a method for calculating this value (p. 29): we have to calculate the mean value of an operator with the ground-state wave function. But where could this function be taken from? As a solution of the Schrödinger equation? Impossible. This equation is too difficult to solve (14 particles, cf. problems with exact solutions, Chapter 4).

The only possibility is somehow to obtain an *approximate* solution to this equation.

What is it all about?

We need mathematical methods which *will allow us to obtain approximate solutions* of the Schrödinger equation. These methods are the variational method, the method of moments and the perturbational approach.

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00013-6

Perturbational method **p. 274**

- Rayleigh–Schrödinger approach (△◆)
- Hylleraas variational principle (◆)
- Hylleraas equation (◆)
- Degeneracy (◆)
- Convergence of the perturbational series (◆)

Virial theorem as a probe of wave function quality (△) **p. 287**

Why is this important?

We have to know how to compute wave functions. The exact wave function is definitely out of our reach; therefore in this chapter we will learn how to calculate its approximations.

What is needed?

- Postulates of quantum mechanics (Chapter 1, necessary),
- Hilbert space (Appendix B, p. 595, necessary),
- matrix algebra (Appendix A, p. 589, necessary),
- Lagrange multipliers (Appendix O, p. 719, necessary),
- orthogonalization (Appendix K, p. 697, occasionally used),
- matrix diagonalization (Appendix L, p. 703, necessary),
- group theory (Appendix C, p. 605, occasionally used).

Classical works

A letter (to Euler) written in 1755 by the 19-year-old Italian mathematician Giuseppe Lodovico Lagrangia, known later as Joseph-Louis Lagrange, represented a foundation of the variational method in mathematics. ★ In Lagrange's "*Essay on the Problem of Three Bodies*," published in 1772, one may also find the beginnings of the virial theorem. ★ The (all important in the present book) variational method of linear combinations of functions was formulated first by Walther Ritz in the paper "*Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik*," published in *Zeitschrift für Reine und Angewandte Mathematik*, 135(1909)1. ★ The method was applied by Erwin Schrödinger in his first works "*Quantisierung als Eigenwertproblem*" in *Annalen der Physik*, 79(1926)361, *ibid.* 79(1926)489, *ibid.* 80(1926)437, and *ibid.* 81(1926)109. ★ Schrödinger also used the perturbational approach when developing the theoretical results of Lord Rayleigh for vibrating systems (hence the often used term Rayleigh–Schrödinger perturbation theory). ★ The quantum mechanical formulation of the virial theorem dates back to a paper "*Bemerkung zum Virialsatz*," published in *Zeitschrift für Physik A*, 63(1930)855 by the Russian physicist Vladimir Fock. ★ Egil Andersen Hylleraas, in the work "*Über den Grundterm der Zweielektronenprobleme von* H^-*, He,* Li^+*,* Be^{++} *usw.*," published in *Zeitschrift der Physik*, 65(1930)209 showed for the first time that the variational principle may also be used for separate terms of perturbational series.

5.1 Variational method

5.1.1 Variational principle

Let us write the Hamiltonian $\hat{H}$ of the system under consideration[1] and take an *arbitrary* function Φ, which satisfies the following conditions:

- it depends on the same coordinates as the solution to the Schrödinger equation,
- it is of class Q (which enables it to be normalized).

We calculate the number ε, which depends on Φ (i.e., ε is the mean value of the Hamiltonian and a functional of Φ),

$$\varepsilon[\Phi] = \frac{\left\langle \Phi \middle| \hat{H} \middle| \Phi \right\rangle}{\langle \Phi | \Phi \rangle}.$$

The variational principle states:

- $\epsilon \geq E_0$, where E_0 is the *ground-state* energy of the system (the lowest eigenvalue of $\hat{H}$),
- in the above inequality $\epsilon = E_0$ happens if and only if Φ equals the exact ground-state wave function Ψ_0 of the system.

Proof (expansion into eigenfunctions)

The eigenfunctions $\{\Psi_i\}$ of the Hamiltonian $\hat{H}$, even if unknown, certainly represent a complete orthonormal set (see Appendix B on p. 595) in the Hilbert space of our system.[2] This means that any function belonging to this space can be represented as a linear combination of the functions of this set,

$$\Phi = \sum_{i=0}^{\infty} c_i \Psi_i, \tag{5.1}$$

where c_i ensure the normalization of Φ, i.e., $\sum\limits_{i=0}^{\infty} |c_i|^2 = 1$, because $\langle \Phi | \Phi \rangle = \sum_{i,j} c_j^* c_i \left\langle \Psi_j | \Psi_i \right\rangle = \sum_{i,j} c_j^* c_i \delta_{ij} = \sum_i c_i^* c_i = 1$. Let us insert this into the expression for the mean value of the en-

---

[1] We focus here on the nonrelativistic case (Eq. (2.1)), where the lowest eigenvalue of $\hat{H}$ is bound from below ($> -\infty$). As we remember from Chapter 3, this is not fulfilled in the relativistic case (Dirac's electronic sea), and may lead to serious difficulties in applying the variational method.

[2] The functions are and will remain unknown; we use here only their property of forming a complete set.

ergy $\varepsilon = \left\langle \Phi | \hat{H} \Phi \right\rangle$. Then,

$$\varepsilon - E_0 = \left\langle \Phi | \hat{H} \Phi \right\rangle - E_0 = \left\langle \sum_{j=0}^{\infty} c_j \Psi_j | \hat{H} \sum_{i=0}^{\infty} c_i \Psi_i \right\rangle - E_0 =$$

$$\sum_{i,j=0}^{\infty} c_j^* c_i E_i \left\langle \Psi_j | \Psi_i \right\rangle - E_0 = \sum_{i,j=0}^{\infty} c_j^* c_i E_i \delta_{ij} - E_0 =$$

$$\sum_{i=0}^{\infty} |c_i|^2 E_i - E_0 \cdot 1 = \sum_{i=0}^{\infty} |c_i|^2 E_i - E_0 \sum_{i=0}^{\infty} |c_i|^2 = \sum_{i=0}^{\infty} |c_i|^2 (E_i - E_0) \geq 0.$$

Note that the *equality* (in the last step) is satisfied only if $\Phi = \Psi_0$. This therefore proves the variational principle: $\varepsilon \geq E_0$.

Proof using Lagrange multipliers

In several places in this book we will need similar proofs using Lagrange multipliers. This is why we will demonstrate how to prove the same theorem using this technique (Appendix O on p. 719).

Take the functional

$$\varepsilon[\Phi] = \left\langle \Phi | \hat{H} \Phi \right\rangle. \tag{5.2}$$

We want to find a function that ensures a minimum of the functional and satisfies the normalization condition

$$\langle \Phi | \Phi \rangle - 1 = 0. \tag{5.3}$$

We will change the function Φ a little (the change will be called "variation") and see how this will change the value of the functional $\varepsilon[\Phi]$. In the functional, we have, however, Φ and Φ^*. It seems that we have, therefore, to take into account in Φ^* the variation made in Φ. *However, in reality there is no need to do that:* it is sufficient to make the variation either in Φ or in Φ^* (the result does not depend on the choice[3]). *This makes the formulae simpler*. We decide to choose the variation of Φ^*, i.e., $\delta\Phi^*$.

---

[3] Let us show this, because we will use it several times in this book. In all our cases the functional (which depends here on a single function $\phi(x)$, but later we will also deal with several functions in a similar procedure) might be rewritten as

$$\varepsilon[\phi] = \left\langle \phi | \hat{A} \phi \right\rangle, \tag{5.4}$$

Now we apply the machinery of the Lagrange multipliers (Appendix O on p. 719). Let us multiply (5.3) by (for the time being) an unknown Lagrange multiplier E and subtract afterwards from the functional ε, resulting in an auxiliary functional $G[\Phi]$, i.e.,

$$G[\Phi] = \varepsilon[\Phi] - E(\langle\Phi|\Phi\rangle - 1).$$

The variation of G (which is analogous to the differential of a function) represents a linear term in $\delta\Phi^*$. For an extremum the variation has to be equal to zero, i.e.,

$$\delta G = \left\langle \delta\Phi|\hat{H}\Phi \right\rangle - E\,\langle\delta\Phi|\Phi\rangle = \left\langle \delta\Phi|(\hat{H} - E)\Phi \right\rangle = 0.$$

Since this has to be satisfied for *any* variation $\delta\Phi^*$, it can follow only if

$$(\hat{H} - E)\Phi_{opt} = 0, \tag{5.5}$$

which means that the optimal $\Phi \equiv \Phi_{opt}$ is a solution of the Schrödinger equation[4] with E as the energy of the stationary state, i.e., Φ_{opt} together with the normalization condition.

Now let us multiply (5.5) by Φ^*_{opt} and integrate. We obtain

$$\left\langle \Phi_{opt}|\hat{H}\Phi_{opt} \right\rangle - E\left\langle \Phi_{opt}|\Phi_{opt} \right\rangle = 0, \tag{5.6}$$

---

where $\hat{A}$ is a Hermitian operator. Let us write $\phi(x) = a(x) + ib(x)$, where $a(x)$ and $b(x)$ are *real functions*. The change of ε is equal to

$$\begin{aligned}\varepsilon[\phi+\delta\phi] - \varepsilon[\phi] &= \left\langle a+\delta a+ib+i\delta b|\hat{A}(a+\delta a+ib+i\delta b) \right\rangle - \left\langle a+ib|\hat{A}(a+ib) \right\rangle \\ &= \left\langle \delta a+i\delta b|\hat{A}\phi \right\rangle + \left\langle \phi|\hat{A}(\delta a+i\delta b) \right\rangle + \text{quadratic terms} = \\ &= \left\langle \delta a|\hat{A}\phi + (\hat{A}\phi)^* \right\rangle + i\left\langle \delta b|(\hat{A}\phi)^* - \hat{A}\phi \right\rangle + \text{quadratic terms.}\end{aligned}$$

The variation of a functional only represents a linear part of the change, and therefore $\delta\varepsilon = \left\langle \delta a|\hat{A}\phi + (\hat{A}\phi)^* \right\rangle + i\left\langle \delta b|(\hat{A}\phi)^* - \hat{A}\phi \right\rangle$. At the extremum the variation has to equal zero at *any* variations of δa and δb. This may happen only if $\hat{A}\phi + (\hat{A}\phi)^* = 0$ and $(\hat{A}\phi)^* - \hat{A}\phi = 0$. This means $\hat{A}\phi = 0$ or, *equivalently*, $(\hat{A}\phi)^* = 0$.
The first of the conditions would be obtained if in ε we made the variation in ϕ^* *only* (the variation in the extremum would then be $\delta\varepsilon = \left\langle \delta\phi|\hat{A}\phi \right\rangle = 0$), hence, from the arbitrariness of $\delta\phi^*$ we would get $\hat{A}\phi = 0$; the second if we made the variation in ϕ *only* (then $\delta\varepsilon = \left\langle \phi|\hat{A}\delta\phi \right\rangle = \left\langle \hat{A}\phi|\delta\phi \right\rangle = 0$ and $(\hat{A}\phi)^* = 0$) and the result is exactly the same. This is what we wanted to show*: we may vary either ϕ or ϕ^* and the result is the same.*

[4] In variational calculus the equation for the optimum Φ, or the conditional minimum of a functional ε, is called the Euler equation. As one can see, in this case the Euler equation is identical with the Schrödinger equation.

or

$$E = \varepsilon[\frac{1}{\sqrt{\langle\Phi_{opt}|\Phi_{opt}\rangle}}\Phi_{opt}], \tag{5.7}$$

which means that the conditional minimum of $\varepsilon[\Phi]$ is $E = min(E_0, E_1, E_2, ...) = E_0$ (the ground state). Hence, for any other Φ we obtain $\varepsilon \geq E_0$.

The same was obtained when we expanded Φ into the eigenfunction series.

Variational principle for excited states

The variational principle has been proved for an approximation to the ground-state wave function. What about excited states? If the variational function Φ is *orthogonal to exact solutions to the Schrödinger equation that correspond to all the states of lower energy than the state we are interested in*, the variational principle is still valid.[5] Thus, if the wave function k being sought represents the *lowest* state among those belonging to a given irreducible representation of the symmetry group of the Hamiltonian, then the orthogonality mentioned above is automatically guaranteed (see Appendix C on p. 605). For other excited states, the variational principle cannot be satisfied, except that function Φ does not contain lower-energy wave functions, i.e., is orthogonal to them, e.g., because the wave functions have been cut out of it earlier.

Watch out: mathematical states

We mentioned in Chapter 1 that not all solutions of the Schrödinger equation are acceptable. Only those are acceptable which satisfy the symmetry requirements with respect to the exchange of labels corresponding to identical particles (Postulate V). The other solutions are called *mathematical*. If, therefore, an incautious scientist takes a variational function Φ with a nonphysical symmetry, the variational principle following our derivation exactly (p. 266) will still be valid, but with respect to the *mathematical ground state*. The mathematical states may correspond to energy eigenvalues *lower* than the physical ground state (they are called then the underground states, cf. p. 95). All this would end up as a catastrophe, because the mean value of the Hamiltonian would tend towards the nonphysical underground mathematical state.

---

[5] The corresponding proof will only be slightly modified. Simply in the expansion (5.1) of the variational function Φ, the wave functions Ψ_i that correspond to lower-energy states (than the state in which we are interested) will be absent. We will therefore obtain $\sum_{i=1} |c_i|^2 (E_i - E_k) \geq 0$, because state k is the lowest among all the states i.

5.1.2 Variational parameters lead to the variational method

The variational principle may seem a little puzzling. We insert an *arbitrary* function Φ into the integral and obtain a result related to the ground state of the *system under consideration*. And yet the arbitrary function Φ may have absolutely nothing to do with the molecule we consider. The problem is that the integral contains the most important information about our system. The information resides in $\hat{H}$. Indeed, if someone wrote down the expression for $\hat{H}$, we would know right away that the system contains N electrons and M nuclei; we would also know the charges of the nuclei, i.e., the chemical elements of which the system is composed.[6] This is important information.

The variational method represents an application of the variational principle. The trial wave function Φ is taken in an analytical form (with the variables denoted by the vector $\mathbf{x}$ and automatically satisfying Postulate V). In the "key positions" in the formula for Φ we introduce the parameters $\mathbf{c} \equiv (c_0, c_1, c_2, ..., c_P)$, which we may change smoothly. The parameters play the role of tuning, their particular values listed in vector $\mathbf{c}$ result in a certain shape of $\Phi(\mathbf{x}; \mathbf{c})$. The integration in the formula for ε pertains to the variables $\mathbf{x}$, and therefore the result depends uniquely on $\mathbf{c}$. Our function $\varepsilon(\mathbf{c})$ has the form

$$\varepsilon(c_0, c_1, c_2, ...c_P) \equiv \varepsilon(\mathbf{c}) = \frac{\left\langle \Phi(\mathbf{x}; \mathbf{c}) | \hat{H} \Phi(\mathbf{x}; \mathbf{c}) \right\rangle}{\langle \Phi(\mathbf{x}; \mathbf{c}) | \Phi(\mathbf{x}; \mathbf{c}) \rangle}.$$

Now the problem is to find the minimum of the function $\varepsilon(c_0, c_1, c_2, ...c_P)$.

In a general case the task is not simple, because what we are searching for is the *global* minimum. The relation

$$\frac{\partial \varepsilon(c_0, c_1, c_2, ...c_P)}{\partial c_i} = 0 \quad \text{for } i = 0, 1, 2, ...P$$

therefore represents only a *necessary* condition for the global minimum.[7] This problem may be disregarded when:

- the number of minima is small,
- in particular, when we use Φ with the linear parameters $\mathbf{c}$ (in this case we have a single minimum, see below).

---

[6] And yet we would be unable to decide whether we have to do with matter or antimatter, or whether we have to perform calculations for the benzene molecule or for six CH radicals (cf. Chapter 2).

[7] More about global minimization can be found in Chapter 6.

The above equations enable us to find the optimum set of parameters $\mathbf{c} = \mathbf{c}_{opt}$. Then,

> in a given class of the trial functions Φ the best possible approximation to Ψ_0 is $\Phi\left(\mathbf{x}; \mathbf{c}_{opt}\right)$, and the best approximation to E_0 is $\varepsilon\left(\mathbf{c}_{opt}\right)$.

Fig. 5.1 shows the essence of the variational method.[8]

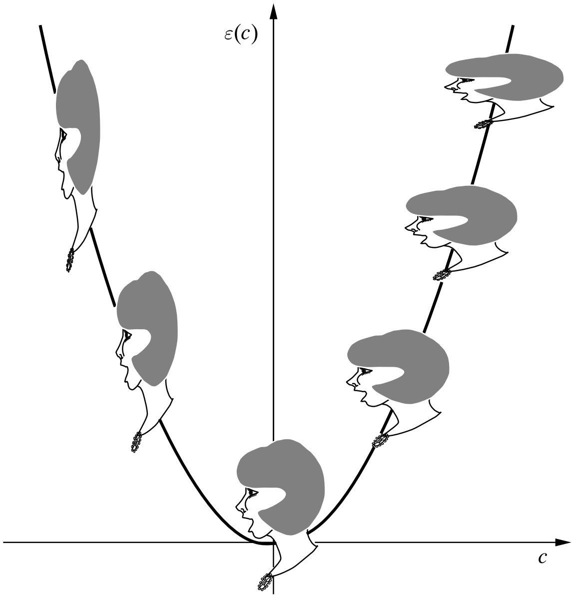


Fig. 5.1. The philosophy behind the variational method. A parameter c is changed in order to obtain the best solution possible. Any commentary would obscure the very essence of the method.

Example 1. Let us assume that someone does not know that the hydrogen-like atom (the nucleus has a charge Z) problem has an exact solution.[9] Let us apply the simplest version of the variational method to see what kind of problem we will be confronted with.

An important first step will be to decide which class of trial functions to choose. We decide to take the class[10] (for $c > 0$) $\exp(-cr)$ and after normalization of the function $\Phi(r, \theta, \phi; c) = \sqrt{\frac{c^3}{\pi}} \exp(-cr)$. The calculation $\varepsilon[\Phi] = \left\langle \Phi \middle| \hat{H} \middle| \Phi \right\rangle$ is shown in Appendix H on p. 683. We

---

[8] The variational method is used in everyday life. Usually we determine the target (say, cleaning the car), and then by trial-and-error we approach the ideal, but never fully achieve it.

[9] For a larger system we will not know the exact solution either.

[10] A particular choice is usually made by scientific discussion. The discussion might proceed as follows.

obtain $\varepsilon\left(c\right)=\frac{1}{2}c^2-Zc$. We very easily find the minimum of $\varepsilon(c)$ and the optimum c is equal to $c_{opt}=Z$, which, as we know from Chapter 4, represents the exact result. This is exceptional; in practice (for atoms or molecules) we would never know the exact result. The optimal ε might then be obtained after many days of number crunching.[11]

5.1.3 Linear variational parameters or the Ritz method[12]

> The Ritz method represents a special kind of variational method. The trial function Φ is represented as a *linear* combination of the *known* basis functions $\{\Psi_i\}$ with the (for the moment) *unknown* variational coefficients c_i,

$$\Phi=\sum_{i=0}^{P}c_i\Psi_i.$$

Then

$$\varepsilon=\frac{\left\langle\sum\limits_{i=0}^{P}c_i\Psi_i|\hat{H}\sum\limits_{i=0}^{P}c_i\Psi_i\right\rangle}{\left\langle\sum\limits_{i=0}^{P}c_i\Psi_i|\sum\limits_{i=0}^{P}c_i\Psi_i\right\rangle}=\frac{\sum\limits_{i=0}^{P}\sum\limits_{j=0}^{P}c_i^*c_jH_{ij}}{\sum\limits_{i=0}^{P}\sum\limits_{j=0}^{P}c_i^*c_jS_{ij}}=\frac{A}{B}. \tag{5.8}$$

The electron and the nucleus attract each other; therefore they will be close in space. This ensures many classes of trial functions, e.g., $\exp\left(-cr\right)$, $\exp\left(-cr^2\right)$, $\exp\left(-cr^3\right)$, etc., where $c>0$ is a single variational parameter. In the present example we pretend not to know which class of functions is most promising (i.e., which will give lower ε). Let us begin with class $\exp\left(-cr\right)$, and other classes will be investigated in the years to come. The decision made, we rush to do the calculations.

[11] For example, for $Z=1$ we had to decide a starting value of c, say, $c=2$; $\varepsilon\left(2\right)=0$. Let us try $c=1.5$; we obtain a lower (i.e., better) value $\varepsilon\left(1.5\right)=-0.375$ a.u.; the energy goes down. This is the good direction; let us try, therefore $c=1.2$; $\varepsilon\left(1.2\right)=-0.48$ a.u. Indeed, a good direction. However, when we continue and take $c=0.7$, we obtain $\varepsilon=-0.455$, i.e., a higher energy. We would continue this way and finally obtain something like $c_{opt}=1.0000000$. We might be satisfied by 8 significant figures and decide to stop the calculations. We would never be sure, however, whether other classes of trial functions would provide still better (i.e., lower) energies. In our particular case this, of course, would never happen, because "accidentally" we have taken a class which contains the exact wave function.

[12] Walther Ritz was a Swiss physicist and a former student of Poincaré. His contributions, beside the variational approach, include perturbation theory and the theory of vibrations. Ritz is also known for his controversy with Einstein on the time flow problem ("time flash"), concluded by their joint article "*An agreement to disagree*" (W. Ritz, A. Einstein, *Phys. Zeit.*, 10(1909)323).

In the formula above $\{\Psi_i\}$ represents the chosen basis set.[13] The basis set functions are often nonorthogonal, and therefore

$$\langle \Psi_i | \Psi_j \rangle = S_{ij}, \tag{5.9}$$

where **S** stands for the *overlap matrix*, and the integrals

$$H_{ij} = \left\langle \Psi_i | \hat{H} \Psi_j \right\rangle \tag{5.10}$$

are the matrix elements of the Hamiltonian. Both matrices (**S** and **H**) are calculated once and for all. The energy ε becomes a function of the linear coefficients $\{c_i\}$. The coefficients $\{c_i\}$ and the coefficients $\{c_i^*\}$ are not independent (c_i can be obtained from c_i^*). Therefore, as the linearly independent coefficients, we may treat either $\{c_i\}$ or[14] $\{c_i^*\}$. When used for the minimization of ε, both choices would give the same result. We decide to treat $\{c_i^*\}$ as variables. For each $k = 0, 1, \ldots P$ we have to have in the minimum

$$0 = \frac{\partial \varepsilon}{\partial c_k^*} = \frac{\left(\sum_{j=0}^{P} c_j H_{kj}\right) B - A \left(\sum_{j=0}^{P} c_j S_{kj}\right)}{B^2} =$$

$$\frac{\left(\sum_{j=0}^{P} c_j H_{kj}\right)}{B} - \frac{A}{B} \frac{\left(\sum_{j=0}^{P} c_j S_{kj}\right)}{B} = \frac{\left(\sum_{j=0}^{P} c_j \left(H_{kj} - \varepsilon S_{kj}\right)\right)}{B},$$

which leads to the *secular equations*

$$\left(\sum_{j=0}^{P} c_j \left(H_{kj} - \varepsilon S_{kj}\right)\right) = 0 \quad \text{for } k = 0, 1, \ldots P. \tag{5.11}$$

The unknowns in the above equation are the coefficients c_j and the energy ε. With respect to the coefficients **c**, Eq. (5.11) represent a homogeneous set of linear equations. Such a set has a nontrivial solution if the *secular determinant* is equal to zero (see Appendix A), i.e.,

---

[13] And have nothing to do with the eigenfunctions used in Eq. (5.1) to prove the variational principle. Such basis sets are available in the literature. A practical problem arises as to how many such functions should be used. In principle we should have used $P = \infty$. This, however, is unfeasible. We are restricted to a finite, *preferably small number*. And this is the moment when it turns out that some basis sets are more effective than others, that this depends on the problem considered, etc. This is how a new science emerges, which might facetiously be called basology.

[14] See footnote on p. 266.

$$\det\left(H_{kj} - \varepsilon S_{kj}\right) = 0. \tag{5.12}$$

This happens however only for some particular values of ε satisfying the above equation. Since the rank of the determinant is equal to $P+1$, we therefore obtain $P+1$ solutions $\varepsilon_i, i = 0, 1, 2, \ldots P$. Due to the Hermitian character of operator $\hat{H}$, the matrix $\mathbf{H}$ will be also Hermitian. In Appendices K on p. 697 and M on p. 705, we show that the problem reduces to the diagonalization of some transformed $\mathbf{H}$ matrix (also Hermitian). This guarantees that all ε_i will be real numbers.[15] Let us denote the lowest ε_i as ε_0, to represent an approximation[16] to the ground state energy E_0. The other $\varepsilon_i, i = 1, 2, \ldots P$, will approximate the excited states of the system with energies $E_1, E_2, E_3, \ldots$. We obtain an approximation to the i-th wave function by inserting the calculated ε_i into (5.11), and then, after including the normalization condition, we find the corresponding set of c_i. The problem is solved.

5.2 Method of moments

In the present book we will have to do quite often with a manifold of functions $G(\mathbf{x}; c_0, c_1, \ldots c_P)$, where G belongs to the Hilbert space, $\mathbf{x}$ stands for coordinates, while $c_0, c_1, \ldots c_P$ are some coefficients (parameters). The special values of the coefficients $c_0, c_1, \ldots c_P$ that imply the equality

$$G(\mathbf{x}; c_0, c_1, \ldots c_P) = 0 \tag{5.13}$$

represent the target of the method of moments. The coefficients may enter G in a complex way and searching the solution of Eq. (5.13) in the Hilbert space may be very difficult. We practically never do such a search.

Instead the method of moments may be applied. In this method one has to have a complete set of some normalized basis functions $\{\varphi_i\}$, of which only $(P+1)$ will be used for producing $(P+1)$ scalar products of both sides[17] of Eq. (5.13):

$$\langle \varphi_i | G(c_0, c_1, \ldots c_P) \rangle = 0 \tag{5.14}$$

for $i = 0, 1, \ldots P$. Eq. (5.14) represents an algebraic set of equations for unknown coefficients $c_0, c_1, \ldots c_P$, which is usually much easier to solve than Eq. (5.13).

---

[15] This is very good, because the energy of the photon, which is required for excitation from one state to the other, will be a real number.

[16] Assuming we used the basis functions that satisfy Postulate V (on symmetry).

[17] This simply means fixing the $(P+1)$ projections of the function Φ on the first $(P+1)$ axes in the Hilbert space.

The Ritz method as a method of moments

If in Eq. (5.14) one inserts as the function $G \rightarrow (\hat{H} - \varepsilon)\Phi$ and as the basis functions $\varphi_i \rightarrow \Psi_i$ of Eq. (5.1), we obtain $\sum_{j=0}^{P} c_j \left(H_{kj} - \varepsilon S_{kj}\right) = 0$, which is identical to the Ritz equation.

5.3 Perturbational method

5.3.1 Rayleigh–Schrödinger approach

The idea of the perturbational approach is very simple. We know everything about a certain nonperturbed problem. Then we slightly perturb the system and everything changes. If the perturbation is small, it seems there is a good chance that there will be no drama: the wave function and the corresponding energy will change only a little (if the changes were large, the perturbational approach would simply be inapplicable). The whole perturbational procedure aims at finding these tiny changes with satisfactory precision.

Perturbational theory is notorious for quite clumsy equations. Unfortunately, there is no way round if we want to explain how to calculate things. *However, in practice only a few of these equations will be used, and they will be highlighted in frames*.

Let us begin our story. We would like to solve the Schrödinger equation

$$\hat{H}\psi_k = E\psi_k, \tag{5.15}$$

and as a rule we will be interested in a single state k, most often the ground state ($k = 0$). This particular state will play an exceptional role in the formulae of the perturbational theory.

We apply a perturbational approach when[18]

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

where the so-called *unperturbed operator* $\hat{H}^{(0)}$ is "large," while the *perturbation operator* $\hat{H}^{(1)}$ is "small".[19] We assume that there is no problem whatsoever with solving the unperturbed Schrödinger equation (also for $n = k$)

$$\hat{H}^{(0)}\psi_n^{(0)} = E_n^{(0)}\psi_n^{(0)}. \tag{5.16}$$

---

[18] We assume all operators are Hermitian.

[19] In the sense that the energy spectrum of $\hat{H}^{(0)}$ is only slightly changed after the perturbation $\hat{H}^{(1)}$ is switched on.

We assume that $\psi_n^{(0)}$ form an orthonormal set, which is natural.[20] We are interested in the fate of the wave function $\psi_k^{(0)}$ after the perturbation is switched on (when it changes to ψ_k). We choose the *intermediate normalization*, i.e.,

$$\left\langle \psi_k^{(0)} | \psi_k \right\rangle = 1. \tag{5.17}$$

The intermediate normalization means that ψ_k, as a vector of the Hilbert space (see Appendix B on p. 595), has the normalized $\psi_k^{(0)}$ as the component along the unit basis vector $\psi_k^{(0)}$. In other words, $\psi_k = \psi_k^{(0)} +$ terms orthogonal to $\psi_k^{(0)}$. The intermediate normalization is convenient, but not necessary. Although convenient for derivation of perturbational equations, it leads to some troubles when the mean values of operators are to be calculated.

We are all set to proceed. First, we introduce the *perturbational parameter* $0 \leq \lambda \leq 1$ in the Hamiltonian $\hat{H}$, making it, therefore, λ-dependent[21]:

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

When $\lambda = 0$, $\hat{H}(\lambda) = \hat{H}^{(0)}$, while $\lambda = 1$ gives $\hat{H}(\lambda) = \hat{H}^{(0)} + \hat{H}^{(1)}$. In other words, we tune the perturbation at will from 0 to $\hat{H}^{(1)}$. It is worth noting that $\hat{H}(\lambda)$ for $\lambda \neq 0, 1$ may not correspond to any physical system. It does not need to. We are interested here in a mathematical trick; we will come back to reality by putting $\lambda = 1$ at the end.

We are interested in the Schrödinger equation satisfied for all values $\lambda \in [0, 1]$. We have

$$\hat{H}(\lambda)\, \psi_k(\lambda) = E_k(\lambda)\, \psi_k(\lambda).$$

Now this is a key step in the derivation. We expect that both the energy and the wave function can be expanded in a power series[22] of λ, i.e.,

$$E_k(\lambda) = E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \ldots, \tag{5.18}$$

---

[20] We can always do that, because $\hat{H}^{(0)}$ is Hermitian (see Appendix B).

[21] Its role is technical. It will enable us to transform the Schrödinger equation into a sequence of perturbational equations, which must be solved one by one. Then the parameter λ *disappears from the theory*, because we put $\lambda = 1$.

[22] It is in fact a Taylor series with respect to λ. The physical meaning of these expansions is the following: $E_k^{(0)}$ and $\psi_k^{(0)}$ are good approximations of $E_k(\lambda)$ and $\psi_k(\lambda)$. The rest will be calculated as a sum of small correction terms. Both series are based on a very simple idea. Imagine a rose flower on a long stalk. We are interested in the distance of the flower center from the ground. Suppose the corresponding measurement gives $h^{(0)}$ (an analog of $E_k^{(0)}$ or $\psi_k^{(0)}$). Now we are going to conceive a theory that predicts the distance of the flower center from the ground (h), when a fly of mass λ sits on the flower center. The first idea to come to mind is that the bending *should be* proportional to λ. Therefore, we have our first guess that $h(\lambda) - h^{(0)} \approx \lambda h^{(1)}$, where $h^{(1)}$ stands for some suitable constant of proportionality. We find this constant with a high precision either experimentally by weighing some small weights (less ambitious) or theoretically (more ambitious) by applying a model of

$$\psi_k(\lambda) = \psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \ldots, \tag{5.19}$$

where $E_k^{(i)}$ stand for some (unknown for the moment) coefficients, and $\psi_k^{(i)}$ represents the functions to be found. We expect the two series to converge (Fig. 5.2).

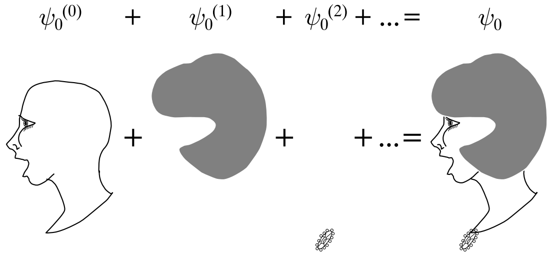


Fig. 5.2. Philosophy of the perturbational approach (optimistic version). The ideal ground-state wave function ψ_0 is constructed as a sum of a good zero-order approximation ($\psi_0^{(0)}$) and consecutive small corrections ($\psi_0^{(n)}$). The first-order correction ($\psi_0^{(1)}$) is still quite substantial, but fortunately the next corrections amount to only small cosmetic changes. I did not dare to continue the pictures in case of possible divergence of the perturbational series.

In practice we calculate only $E_k^{(1)}$, $E_k^{(2)}$, and, quite rarely, $E_k^{(3)}$, and for the wave function we usually limit ourselves to $\psi_k^{(1)}$.

How are these corrections calculated?

We insert the two perturbational series for $E_k(\lambda)$ and $\psi_k(\lambda)$ into the Schrödinger equation. Then we have

$$\left(\hat{H}^{(0)} + \lambda \hat{H}^{(1)}\right)\left(\psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \ldots\right)$$
$$= \left(E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \ldots\right)\left(\psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \ldots\right)$$

---

the flower. Then we are proud to see our theory working, when some small flies are sitting on the rose. After some time we are bored by weighing small flies, and we begin to test the formula $h(\lambda) - h^{(0)} \approx \lambda h^{(1)}$ for bees. Unfortunately, our theory ceases to work well (effect of the nonlinearity). We watch our formula as function of λ to fail and, all of a sudden, we get an idea. The idea is to add a new term $\lambda^2 h^{(2)}$ with a suitably chosen $h^{(2)}$ constant: $h(\lambda) - h^{(0)} \approx \lambda h^{(1)} + \lambda^2 h^{(2)}$. And indeed, we have again success until we consider heavier insects. Then we introduce $h^{(3)}$, etc. It is a good time to stop improving our theory and say that we have a good theory working for small insects. This is reflected by the formulae for $E_k(\lambda)$ and $\psi_k(\lambda)$ given above.
An attempt to use the theory for heavy weights is bound to finish badly. These formulae will not work (even if we take into account an infinite number of terms) for large weights, e.g., so large that the stalk breaks (an abrupt change of its structure). In such cases the series for $E_k(\lambda)$ and $\psi_k(\lambda)$ will turn out to diverge and we will witness a catastrophe of the perturbation theory.

and, since the equation has to be satisfied for *any* λ belonging to $0 \le \lambda \le 1$, this may happen only if

> the coefficients at the same powers of λ on the left- and right-hand sides must be equal.

This gives a sequence of an infinite number of perturbational equations to be satisfied by the unknown $E_k^{(n)}$ and $\psi_k^{(n)}$. *These equations may be solved consecutively*, allowing us to calculate $E_k^{(n)}$ and $\psi_k^{(n)}$ with larger and larger n. We have, for example,

$$\text{for } \lambda^0: \qquad \hat{H}^{(0)}\,\psi_k^{(0)} = \; E_k^{(0)}\,\psi_k^{(0)},$$

for λ^1:

$$\hat{H}^{(0)}\psi_k^{(1)} + \hat{H}^{(1)}\psi_k^{(0)} = E_k^{(0)}\psi_k^{(1)} + E_k^{(1)}\psi_k^{(0)}, \tag{5.20}$$

$$\text{for } \lambda^2: \hat{H}^{(0)}\psi_k^{(2)} + \hat{H}^{(1)}\psi_k^{(1)} = E_k^{(0)}\psi_k^{(2)} + E_k^{(1)}\psi_k^{(1)} + E_k^{(2)}\psi_k^{(0)},$$

$$\ldots$$

etc.[23]

Doing the same with the intermediate normalization (Eq. (5.17)), we obtain

$$\left\langle \psi_k^{(0)} | \psi_k^{(n)} \right\rangle = \delta_{0n}. \tag{5.21}$$

The first equation of (5.20) is evident (the unperturbed Schrödinger equation does not contain any unknown). The second equation involves two unknowns, $\psi_k^{(1)}$ and $E_k^{(1)}$. There is a way to eliminate $\psi_k^{(1)}$ by using the Hermitian character of the operators. Indeed, by making the scalar product of the equation with $\psi_k^{(0)}$ we obtain

$$\left\langle \psi_k^{(0)} | \left(\hat{H}^{(0)} - E_k^{(0)}\right)\psi_k^{(1)} + \left(\hat{H}^{(1)} - E_k^{(1)}\right)\psi_k^{(0)} \right\rangle =$$
$$\left\langle \psi_k^{(0)} | \left(\hat{H}^{(0)} - E_k^{(0)}\right)\psi_k^{(1)} \right\rangle + \left\langle \psi_k^{(0)} | \left(\hat{H}^{(1)} - E_k^{(1)}\right)\psi_k^{(0)} \right\rangle =$$
$$0 + \left\langle \psi_k^{(0)} | \left(\hat{H}^{(1)} - E_k^{(1)}\right)\psi_k^{(0)} \right\rangle = 0,$$

i.e.,

---

[23] We see the construction principle of these equations: we write down all the terms which give a given value of the sum of the upper indices.

the formula for the *first-order correction to the energy* is

$$E_k^{(1)} = H_{kk}^{(1)}, \tag{5.22}$$

where we defined

$$H_{kn}^{(1)} = \langle \psi_k^{(0)} | \hat{H}^{(1)} | \psi_n^{(0)} \rangle. \tag{5.23}$$

We conclude that the first-order correction to the energy, $E_k^{(1)}$, represents the mean value of the perturbation with the unperturbed wave function of the state in which we are interested (usually the ground state).[24]

Now, from the perturbation equation (5.20) corresponding to $n = 2$ we get the following through a scalar product with $\psi_k^{(0)}$:

$$\left\langle \psi_k^{(0)} | \left(\hat{H}^{(0)} - E_k^{(0)} \right) \psi_k^{(2)} \right\rangle + \left\langle \psi_k^{(0)} | \left(\hat{H}^{(1)} - E_k^{(1)} \right) \psi_k^{(1)} \right\rangle - E_k^{(2)} =$$
$$\left\langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_k^{(1)} \right\rangle - E_k^{(2)} = 0,$$

and hence

$$E_k^{(2)} = \left\langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_k^{(1)} \right\rangle. \tag{5.24}$$

For the time being we cannot compute $E_k^{(2)}$, because we do not know $\psi_k^{(1)}$, but soon we will. Let us expand $\psi_k^{(1)}$ into the complete set of the basis functions $\left\{ \psi_n^{(0)} \right\}$ with as yet unknown coefficients c_n, i.e.,

$$\psi_k^{(1)} = \sum_{n \neq k} c_n \psi_n^{(0)}. \tag{5.25}$$

---

[24] This is quite natural and we use such a perturbative estimation all the time. What it really says is: we do not know what the perturbation exactly does, but let us estimate the result by *assuming that all things are going on as they were before the perturbation was applied*. In the first-order approach, insurance estimates your loss by *averaging* over similar losses of others. Every morning you assume that the traffic on the highway will be *the same as it is everyday*, etc.

Note that because of the intermediate normalization (5.17) and (5.21), we did not take the term with $n = k$. We get

$$\left(\hat{H}^{(0)} - E_k^{(0)}\right) \sum_{n \neq k} c_n \psi_n^{(0)} + \hat{H}^{(1)} \psi_k^{(0)} = E_k^{(1)} \psi_k^{(0)},$$

and transform this into

$$\sum_{n \neq k} c_n \left(E_n^{(0)} - E_k^{(0)}\right) \psi_n^{(0)} + \hat{H}^{(1)} \psi_k^{(0)} = E_k^{(1)} \psi_k^{(0)}.$$

We find c_m by making the scalar product of both sides of the last equation with $\psi_m^{(0)}$. In our case, due to the orthonormality of functions $\left\{\psi_n^{(0)}\right\}$, we obtain

$$c_m = \frac{H_{mk}^{(1)}}{E_k^{(0)} - E_m^{(0)}},$$

which from Eq. (5.25) gives the following formula for the *first-order correction to the wave function*[25]:

$$\boxed{\psi_k^{(1)} = \sum_{n \neq k} \frac{H_{nk}^{(1)}}{E_k^{(0)} - E_n^{(0)}} \psi_n^{(0)},} \tag{5.26}$$

and the following formula for the *second-order correction to the energy*:

$$\boxed{E_k^{(2)} = \sum_{n \neq k} \frac{|H_{kn}^{(1)}|^2}{E_k^{(0)} - E_n^{(0)}}.} \tag{5.27}$$

---

[25] This formula may be seen as a foundation of almost all we do in our everyday life, consciously or unconsciously. We start from an initial situation (say, state) $\psi_k^{(0)}$. We want to change this state, but all the time monitoring whether the results are OK for us. Caution advices to apply a very small perturbation (in this situation the change is practically equal to $\psi_k^{(1)}$) and watch carefully how the system reacts upon it. The effects are seen mainly on $\psi_k^{(1)}$ and $E_k^{(2)}$. This is reflected in the first-order formulae. In practical life we apparently forget about higher-order perturbational formulae. Instead we keep applying all the time the first-order perturbation theory (in this way the higher-order corrections reenter implicitly). We decide whether *in the new situation* $\psi_k^{(0)} + \psi_k^{(1)}$ (which now is again treated as "unperturbed") the applied perturbation should be kept the same, decreased, or increased. This is how we enter a curve when driving car: we adapt the steering wheel position to what we see after having its previous position, and we keep doing this again and again.

All terms on the right-hand sides of these formulae are known (or can be calculated) from the unperturbed Eq. (5.16). For $k = 0$ (ground state) we have

$$E_0^{(2)} \leq 0. \tag{5.28}$$

From (5.26) we see that the contribution of function $\psi_n^{(0)}$ to the wave function deformation is large if the coupling between states k and n (i.e., $H_{nk}^{(1)}$) is large and these states are close in the energy scale.

We will limit ourselves to the most important corrections in the hope that the perturbational method converges fast (we will see in a moment how surprising the perturbational series behavior can be), and we assume further corrections are much less important.[26] The formulae for higher corrections become more and more complicated.

5.3.2 *Hylleraas variational principle*[27]

The derived formulae are rarely employed in practice, because we very rarely have at our disposal all the necessary solutions of Eq. (5.16). The eigenfunctions of the $\hat{H}^{(0)}$ operator appeared as a consequence of using them as the complete set of functions (e.g., in expanding $\psi_k^{(1)}$). There are, however, some numerical methods that enable us to compute $\psi_k^{(1)}$ using the complete set of functions $\{\phi_i\}$, which are not the eigenfunctions of $\hat{H}^{(0)}$.

Hylleraas noted[28] that the functional

$$\mathcal{E}\left[\tilde{\chi}\right] = \left\langle \tilde{\chi} \right| \left(\hat{H}^{(0)} - E_0^{(0)}\right) \tilde{\chi} \Big\rangle + \tag{5.29}$$

$$\left\langle \tilde{\chi} \right| \left(\hat{H}^{(1)} - E_0^{(1)}\right) \psi_0^{(0)} \Big\rangle + \left\langle \psi_0^{(0)} \right| \left(\hat{H}^{(1)} - E_0^{(1)}\right) \tilde{\chi} \Big\rangle \tag{5.30}$$

exhibits its minimum at $\tilde{\chi} = \psi_0^{(1)}$ and for this function the value of the functional is equal to $E_0^{(2)}$. Indeed, inserting $\tilde{\chi} = \psi_0^{(1)} + \delta\chi$ into Eq. (5.30) and using the Hermitian character of the operators, we obtain

---

[26] Some scientists have been bitterly disappointed by this assumption.

[27] See his biographic note in Chapter V2-2.

[28] E.A. Hylleraas, *Zeit. Phys.*, 65(1930)209.

$$\mathcal{E}\left[\psi_0^{(1)}+\delta\chi\right]-\mathcal{E}\left[\psi_0^{(1)}\right]=\left\langle\psi_0^{(1)}+\delta\chi|\left(\hat{H}^{(0)}-E_0^{(0)}\right)(\psi_0^{(1)}+\delta\chi)\right\rangle+$$
$$\left\langle\psi_0^{(1)}+\delta\chi|\left(\hat{H}^{(1)}-E_0^{(1)}\right)\psi_0^{(0)}\right\rangle+\left\langle\psi_0^{(0)}|\left(\hat{H}^{(1)}-E_0^{(1)}\right)(\psi_0^{(1)}+\delta\chi)\right\rangle=$$
$$\left\langle\delta\chi|\left(\hat{H}^{(0)}-E_0^{(0)}\right)\psi_0^{(1)}+\left(\hat{H}^{(1)}-E_0^{(1)}\right)\psi_0^{(0)}\right\rangle+$$
$$\left\langle\left(\hat{H}^{(0)}-E_0^{(0)}\right)\psi_0^{(1)}+\left(\hat{H}^{(1)}-E_0^{(1)}\right)\psi_0^{(0)}|\delta\chi\right\rangle+$$
$$\left\langle\delta\chi|\left(\hat{H}^{(0)}-E_0^{(0)}\right)\delta\chi\right\rangle=\left\langle\delta\chi|\left(\hat{H}^{(0)}-E_0^{(0)}\right)\delta\chi\right\rangle\geq 0.$$

This proves the Hylleraas variational principle. The last equality follows from the first-order perturbational equation, and the last inequality from the fact that $E_0^{(0)}$ is assumed to be the lowest eigenvalue of $\hat{H}^{(0)}$ (see the variational principle).

What is the minimal value of the functional under consideration? Let us insert $\tilde{\chi}=\psi_0^{(1)}$. We obtain

$$\mathcal{E}\left[\psi_0^{(1)}\right]=$$
$$\left\langle\psi_0^{(1)}|\left(\hat{H}^{(0)}-E_0^{(0)}\right)\psi_0^{(1)}\right\rangle+\left\langle\psi_0^{(1)}|\left(\hat{H}^{(1)}-E_0^{(1)}\right)\psi_0^{(0)}\right\rangle+\left\langle\psi_0^{(0)}|\left(\hat{H}^{(1)}-E_0^{(1)}\right)\psi_0^{(1)}\right\rangle=$$
$$\left\langle\psi_0^{(1)}|\left(\hat{H}^{(0)}-E_0^{(0)}\right)\psi_0^{(1)}+\left(\hat{H}^{(1)}-E_0^{(1)}\right)\psi_0^{(0)}\right\rangle+\left\langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\right\rangle=$$
$$\left\langle\psi_0^{(1)}|0\right\rangle+\left\langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\right\rangle=\left\langle\psi_0^{(0)}|\hat{H}^{(1)}\psi_0^{(1)}\right\rangle=E_0^{(2)}.$$

5.3.3 Hylleraas equation

The first-order perturbation equation (p. 277) after inserting

$$\psi_0^{(1)}=\sum_j^N d_j\phi_j \tag{5.31}$$

takes the form

$$\sum_j^N d_j(\hat{H}^{(0)}-E_0^{(0)})\phi_j+(\hat{H}^{(1)}-E_0^{(1)})\psi_0^{(0)}=0.$$

Making the scalar products of the left- and right-hand sides of the equation with functions $\phi_i, i=1,2,...,$ we obtain

$$\sum_j^N d_j(\hat{H}_{ij}^{(0)}-E_0^{(0)}S_{ij})=-(\hat{H}_{i0}^{(1)}-E_0^{(1)}S_{i0})\text{ for }i=1,2,...,N,$$

where $\hat{H}_{ij}^{(0)} \equiv \left\langle \phi_i | \hat{H}^{(0)} \phi_j \right\rangle$, and the overlap integrals $S_{ij} \equiv \langle \phi_i | \phi_j \rangle$. Using the matrix notation we may write the *Hylleraas equation*

$$(\mathbf{H}^{(0)} - E_k^{(0)} \mathbf{S})\mathbf{d} = -\mathbf{v}, \tag{5.32}$$

where the components of the vector $\mathbf{v}$ are $v_i = \hat{H}_{i0}^{(1)} - E_0^{(1)} S_{i0}$. All the quantities can be calculated and the set of N linear equations with unknown coefficients d_i remains to be solved.[29]

5.3.4 Degeneracy

There is a trap in the perturbational formulae, which may lead to catastrophe. Let us imagine that the unperturbed state k (its change represents the target of the perturbation theory) is g-tuply degenerate. Then Eq. (5.16) is satisfied by g wave functions $\psi_{k,m}^{(0)}, m = 1, 2, ..., g$, which according to Appendix B (p. 595) can be chosen as orthonormal, i.e., $\left\langle \psi_{k,m}^{(0)} | \psi_{k,n}^{(0)} \right\rangle = \delta_{mn}$:

$$\hat{H}^{(0)} \psi_{k,m}^{(0)} = E_k^{(0)} \psi_{k,m}^{(0)}. \tag{5.33}$$

If one decided to choose a given $\psi_{k,m}^{(0)}$ as the unperturbed state, we could calculate neither $E_k^{(2)}$ nor $\psi_k^{(1)}$, because the denominators in the corresponding formulae would be equal to 0 and the results would "explode to infinity."

We should somehow fix the problem. Let us focus on a particular unperturbed state $\psi_{k,1}^{(0)}$. A slightest perturbation applied would force us to consider the Ritz variational method to determine the resulting state. We can choose all $\psi_{k,m}^{(0)}$ functions as possible expansion functions. This however means also including $\psi_{k,m}^{(0)}, m = 2, 3, ..., g$, but *they may enter with potentially very large weights*.[30] This means a possibility of a giant change of the wave function resulting from even a very small perturbation.[31] Therefore, we cannot say the perturbation is small, while this is the most important requirement for perturbation theory to operate (Fig. 5.3).

---

[29] We obtain the *same equation* if in the Hylleraas functional (5.30) the variational function χ is expanded as a linear combination (5.31), and then vary d_i in a similar way to that of the Ritz variational method described on p. 271.

[30] Because they satisfy the Schrödinger equation for the unperturbed system with the same eigenvalue of energy. As a consequence any linear combination of them satisfies this equation as well as $\psi_{k,1}^{(0)}$ itself.

[31] Thinking about future: this is the reason for the richness of local spatial configurations around atoms in chemical compounds (what is called the valency). Most often such atoms offer degenerate or quasidegenerate valence orbitals. They can therefore mix (hybridization) easily even under small perturbation coming from the neighborhood. The product of such mixing (hybridized orbitals) are oriented towards the perturbing partner atoms. *On top of that* these hybrids may offer 2, 1, 0 electrons (depending on what number of electrons the partner's orbitals carry), which leads to the coordination bond with donation of the electron pair, covalent bond, or coordination bond with acceptance of the electron pair, respectively.

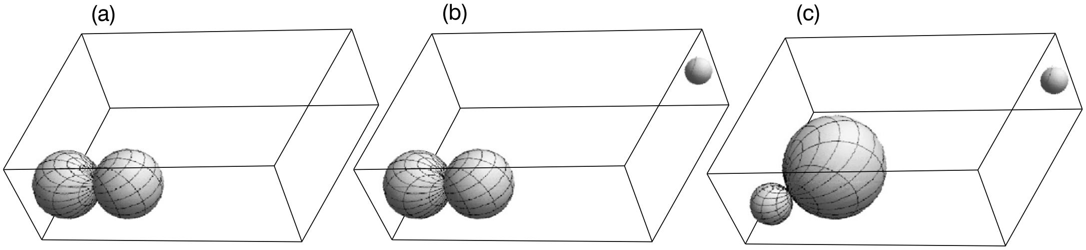


Fig. 5.3. A degenerate state may mean an instable system ("superpolarizable"), which adapts itself very easily to an external (even very small) perturbation. The figure shows how the hydrogen atom in a degenerate state changes when a proton approaches it. (a) The atom in a nonperturbed stationary state $2p_x$. (b) The proton interacts with the hydrogen atom, but the direction of the $2p_x$ orbital does not reflect this. It is true that the orbital $2p_x$ describes quite well the total energy (because the interaction is weak), but very badly the electron charge distribution in the hydrogen atom interacting with the proton. (c) Contrary to this, the function $\phi_k^{(0)} = (2s) + \frac{1}{\sqrt{2}}(2p_x + 2p_y)$ satisfies the Schrödinger equation for the unperturbed hydrogen atom as well as the function $2p_x$, but is much more reasonable after the perturbation is switched on. The function $\phi_k^{(0)}$ shows that the electron recognizes the direction to the proton and tries to approach it. The function $\phi_k^{(0)}$ represents a giant change with respect to $2p_x$. The most reasonable unperturbed function will be determined by the variational method as a linear combination of (mostly) $2s, 2p_x, 2p_y, 2p_z$. After this is done, one may keep improving the function by adding perturbational corrections.

The first thing to do is to adapt the wave function of the unperturbed system to the perturbation. To this end we will use the Ritz variational method with the expansion functions[32] $\psi_{k,m}^{(0)}, m = 1, 2, ..., g$, i.e.,

$$\phi_{k,m}^{(0)} = \sum_{i=1}^{g} c_{km}\psi_{k,m}^{(0)}. \tag{5.34}$$

The secular equation (5.11), after inserting $\hat{H} = \hat{H}^{(0)} + \hat{H}^{(1)}$ and Eq. (5.33), transforms to[33]

$$\det\left\{E_k^{(0)}\delta_{mn} + \hat{H}_{mn}^{(1)} - \varepsilon\delta_{mn}\right\} = 0. \tag{5.35}$$

---

[32] The same solution is obtained when inserting expansion (5.34) into the first-order perturbational equation, carrying out the scalar products of this equation consecutively with $\psi_{k,m}^{(0)}, m = 1, 2, ..., g$, and solving the resulting system of equations.

[33] $\hat{H}_{mn} = \left\langle\psi_{k,m}^{(0)}|\hat{H}^{(0)}\psi_{k,n}^{(0)}\right\rangle + \left\langle\psi_{k,m}^{(0)}|\hat{H}^{(1)}\psi_{k,n}^{(0)}\right\rangle = E_k^{(0)}\delta_{mn} + \hat{H}_{mn}^{(1)}$;
$S_{mn} = \delta_{mn}$.

This may be rewritten as

$$\det\left\{\hat{H}^{(1)}_{mn} - E^{(1)}_k \delta_{mn}\right\} = 0, \tag{5.36}$$

with[34] $E^{(1)}_k \equiv \varepsilon - E^{(0)}_k$. Solving this equation we obtain the unknown energies $E^{(1)}_{k,i}, i = 1, 2, ...g$. Solving afterwards the secular equation (Chapter 5 and Appendix M, p. 705) consecutively for $E^{(1)}_{k,i}, i = 1, 2, ...g$, we get the sets (numbered by $i = 1, 2, ...g$) of coefficients c_{km}. This means we have the quantity g of the unperturbed functions. We choose one of them, the one which corresponds to the target state, and the perturbational approach is applied to it, in the same way as for the nondegenerate case (except we do not include $\psi^{(0)}_{k,m}$ to the expansion functions any more).

5.3.5 *Convergence of the perturbational series*

The perturbational approach is applicable when the perturbation only slightly changes the energy levels, therefore not changing their order. This means that the unperturbed energy level separations have to be much larger than a measure of perturbation such as $\hat{H}^{(1)}_{kk} = \left\langle \psi^{(0)}_k | \hat{H}^{(1)} \psi^{(0)}_k \right\rangle$. However, even in this case we may expect complications.

The subsequent perturbational corrections need not be monotonically decreasing. However, if the perturbational series (5.19) converges, for any $\varepsilon > 0$ we may choose such N_0 that for $N > N_0$ we have $\left\langle \psi^{(N)}_k | \psi^{(N)}_k \right\rangle < \varepsilon$, i.e., the vectors $\psi^{(N)}_k$ have smaller and smaller length in the Hilbert space.

Unfortunately, perturbational series are often divergent in a sense known as *asymptotic convergence*. A divergent series $\sum_{n=0}^{\infty} \frac{A_n}{z^n}$ is called an *asymptotic series* of a function $f(z)$ if the function $R_n(z) = z^n[f(z) - S_n(z)]$, where $S_n(z) = \sum_{k=0}^{n} \frac{A_k}{z^k}$, satisfies the following condition: $\lim_{z\to\infty} R_n(z) = 0$ for any fixed n. In other words, the error of the summation, i.e., $[f(z) - S_n(z)]$ tends to 0 as $z^{-(n+1)}$ or faster.

Despite the fact that the series used in physics and chemistry are often asymptotic, i.e., divergent, we are able to obtain results of high accuracy with them provided we limit ourselves to an appropriate number of terms. The asymptotic character of such series manifests itself in practice in such a way that the partial sums stabilize and we obtain numerically a situation typical for convergence. For example, we sum up the consecutive perturbational corrections and obtain the partial sums changing on the eighth, then ninth, then tenth significant figures. This is a very

---

[34] The energies $E^{(1)}_{k,i}$ have an additional superscript "(1)" just to stress that they are proportional to perturbation. They all stem from the unperturbed state k, but in general differ for different i, i.e., we usually get a splitting of the energy level k.

good time to stop the calculations, publish the results, finish the scientific career, and move on to other business. The addition of further perturbational corrections ends up in catastrophe (cf. Appendix V2-G on p. V2-613). It begins with an innocent, very small increase in the partial sums; they just begin to change the ninth, then the eighth, then the seventh significant figure. Then, it only gets worse and worse and ends with an explosion of the partial sums to ∞ and a very bad state of mind for the researcher (I did not dare to depict it in Fig. 5.2).

In perturbation theory we assume that $E_k(\lambda)$ and $\psi_k(\lambda)$ are analytical functions of λ (p. 276). In this *mathematical* aspect of the physical problem we may treat λ as a complex number. *Then the radius of convergence ρ of the perturbational series on the complex plane is equal to the smallest $|\lambda|$ for which one has a pole of $E_k(\lambda)$ or $\psi_k(\lambda)$*. The convergence radius ρ_k for the energy perturbational series may be computed as (if the limit exists[35])

$$\rho_k = \lim_{N\to\infty} \frac{\left|E_k^{(N)}\right|}{\left|E_k^{(N+1)}\right|}.$$

For physical reasons $\lambda = 1$ is most important. It is, therefore, desirable to have $\rho_k \geq 1$. Note (Fig. 5.4) that if $\rho_k \geq 1$, then the series with $\lambda = 1$ is convergent together with the series with $\lambda = -1$.

Let us take as the unperturbed system the harmonic oscillator (with the potential energy equal to $\frac{1}{2}x^2$) in its ground state and the operator $\hat{H}^{(1)} = -0.000001 \cdot x^4$ as its perturbation. In such a case the perturbation seems to be small[36] in comparison with the separation of the eigenvalues of $\hat{H}^{(0)}$. And yet the perturbational series carries the seed of catastrophe. It is quite easy to see why a catastrophe has to happen. After the perturbation is added, the potential becomes *qualitatively* different from $\frac{1}{2}x^2$. For large x, instead of going to ∞, it will tend to $-\infty$. The perturbation is not small at all; it is a monster. This will cause the perturbational series to diverge. How will it happen in practice? Well, in higher orders we have to calculate the integrals $\left\langle \psi_n^{(0)} | \hat{H}^{(1)} \psi_m^{(0)} \right\rangle$, where n, m stand for the vibrational quantum numbers. As we recall from Chapter 4, high-energy wave functions have large values for large x, where the perturbation changes as x^4 and gets larger and larger as x increases. This is why the integrals will be large. Therefore, the better we do our job (higher orders, higher-energy states) the faster we approach catastrophe.

---

[35] If the limit does not exist, nothing can be said about ρ_k.

[36] As a measure of the perturbation we may use $\left\langle \psi_0^{(0)} | \hat{H}^{(1)} \psi_0^{(0)} \right\rangle$, which means an integral of x^4 multiplied by a Gaussian function (cf. Chapter 4). Such an integral is easy to calculate and, in view of the fact that it will be multiplied by the (small) factor 0.000001, the perturbation will turn out to be small.

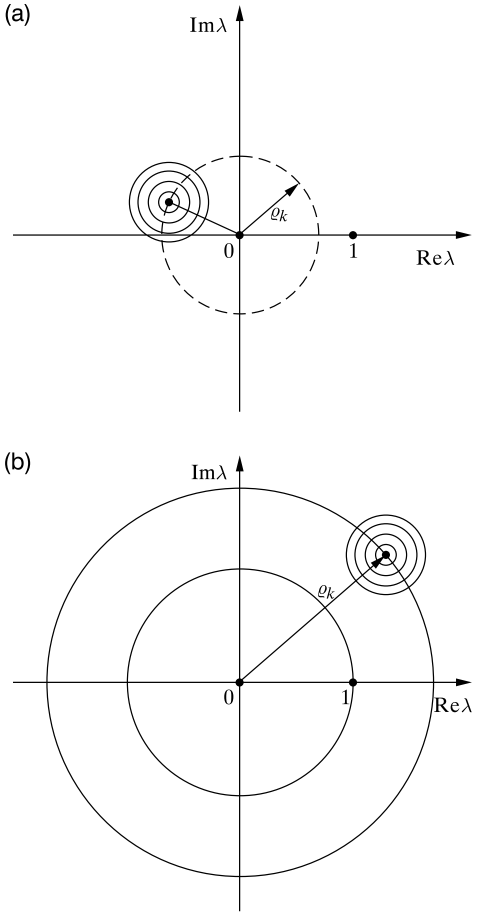


Fig. 5.4. The complex plane of the λ parameter. The physically interesting points are at $\lambda = 0, 1$. In perturbation theory we finally put $\lambda = 1$. Because of this the convergence radius ρ_k of the perturbational series has to be $\rho_k \geq 1$. However, *if any complex* λ *with* $|\lambda| < 1$ corresponds to a pole of the energy, the perturbational series will diverge in the physical situation ($\lambda = 1$). The figure shows the position of a pole by concentric circles. (a) The pole (depicted as the center of the four concentric circles) is too close ($\rho_k < 1$) and the perturbational series diverges. (b) The perturbational series converges, because $\rho_k > 1$.

Let us consider the opposite perturbation $\hat{H}^{(1)} = +0.000001 \cdot x^4$. Despite the fact that everything looks good (the perturbation does not qualitatively change the potential), the series will diverge sooner or later. It is bound to happen, because the convergence radius does not depend on the sign of the perturbation. A researcher might be astonished when the corrections begin to "explode."

Quantum chemistry experiences with perturbational theories look quite consistent:

- low orders may give excellent results,
- higher orders often make the results worse.[37]

5.4 Virial theorem as a probe of wave function quality

5.4.1 Classical mechanics – the virial

We will focus on the quantum mechanical aspect of the virial theorem, but the notion appeared first in the Newtonian classical mechanics (for the derivation see Appendix I on p. 685) and still continues to play an important role there. The name *virial* stems from the Latin name *vires*, i.e., forces, and has to do with forces $\mathbf{F}_i$ acting on the particles $i = 1, 2, \ldots N$ the system is composed of. In classical mechanics the following expression can be proved for *stably bound systems of point-like particles* (Laplace 1772, see Appendix I on p. 685):

$$2\langle T\rangle_\tau + \sum_{i=1}^{N} \langle \mathbf{r}_i \cdot \mathbf{F}_i\rangle_\tau = 0, \tag{5.37}$$

where $\langle\rangle_\tau$ stands for the *time average*, T is the kinetic energy, $\mathbf{r}_i$ is a radius vector for particle i, and $\sum_{i=1}^{N} \langle \mathbf{r}_i \cdot \mathbf{F}_i\rangle_\tau$ is known as the *virial*. If all particles interact *pairwise* through the central potential

$$\mathcal{V}(r_{ij}) = A r_{ij}^{n}, \tag{5.38}$$

where A is a constant and r is the interparticle distance, one obtains (see Appendix I, p. 685) from Eq. (5.37)

$$2\langle T\rangle_\tau = n\langle V\rangle_\tau, \tag{5.39}$$

where $V = \sum_{i<j} \mathcal{V}(r_{ij})$ represents the total potential energy. The power n is equal -1 in case of stars and planets (gravitation) as well as nuclei and electrons (Coulomb law).

The virial theorem (5.37) is of general applicability: it pertains to simple and small (like electrons and nuclei) as well as to complex and giant systems (like a liquid or galaxies). In the case of statistical mechanics the time average can be replaced by the ensemble average. Moreover, the virial theorem (5.37) is applicable for equilibrium and nonequilibrium systems as well. These features make it one of the most valuable tools for estimation of quantities inaccessible by many other methods. However, the virial theorem is known to be dangerous, because it is quite easy to cross its applicability range. Its usefulness as well as its limitation will be illustrated in the present chapter. Let us start from something really big.

---

[37] Even orders as high as 2000 have been investigated in the hope that the series will improve the results...

5.4.2 Looking at stars – the discovery of dark matter

In 1933 the Swiss astrophysicist Fritz Zwicky studied a cluster of 10 000 stars located in the Northern constellation of Berenice's Hair. Many of these stars looked however a bit too fuzzy. Strange. It turned out that in fact they did not represent stars, but galaxies. No doubt, a single galaxy (like our Milky Way) with billions of suns and planets is beyond our imagination, but 10 000 galaxies make an impression more than 10 000 times stronger. Astronomers are certainly smart and brave: by analyzing the amount of coming light Zwicky was able to estimate the cluster's mass ("visible mass").

And here enters the power of the virial theorem. By measuring the Doppler effect of the light coming from the stars Zwicky was able to obtain the galaxies' velocity distribution. Then, using the virial theorem, he performed an *independent* estimation of the cluster's mass[38] (Fig. 5.5), and the result was $4.5 \cdot 10^{13}$ in the unit of our Sun's mass. The comparison of the two masses' values found led Zwicky to the conclusion that there must be 400 times more matter than he calculated from what he could see (as "visible matter"). Contemporary estimations reduced this discrepancy to an order of magnitude, but the problem persisted. The difference between the two masses is attributed mainly to a mysterious invisible "dark matter," which interacts with the other matter and itself only gravitationally, and therefore stays close to galaxies, black holes, etc. Currently, nobody knows what a major part of the dark matter really is! It has to be somewhere around us though, because we are in one of the galaxies – our home Milky Way. We cannot see it or touch it, but nevertheless we can detect its presence...[39]

5.4.3 Quantum mechanics

Eq. (5.37) is replaced in quantum mechanics (see Appendix I, Eq. (I.16) on p. 689) by a relation between the mean value of the kinetic energy $\left\langle \hat{T} \right\rangle$ and the mean value of the virial

---

[38] A single galaxy in the Berenice's Hair cluster of $N = 10\,000$ galaxies has the mean mass m and the mean velocity v. Thus, the total mass of the cluster is $M = Nm$ and its total kinetic energy $\langle T \rangle_\tau \approx N\frac{1}{2}mv^2 = \frac{1}{2}Mv^2$. The total gravitational potential energy (G is the gravity constant) for the cluster can be estimated as $\langle V \rangle_\tau \approx -G\sum_{i<j}\frac{m^2}{r_{ij}} = -\frac{1}{2}G\sum_{i,j}\frac{m^2}{r_{ij}} \approx -\frac{1}{2}G\frac{\sum_{i,j}m^2}{R} = -\frac{1}{2}\frac{GM^2}{R}$. Therefore, from the virial theorem ($n = -1$ for gravitation) one gets $2\frac{1}{2}Mv^2 = \frac{1}{2}\frac{GM^2}{R}$ and hence $M = 2v^2\frac{R}{G}$. The mean velocity v may be computed from the Doppler effect visible in atomic spectra, while R can be estimated from the observed angular dimension and the distance (R was taken as two millions of light years). This enables one to compute the mass of the cluster of galaxies M.

[39] I recall a lecture delivered at the University of Warsaw by Professor Andrei Derevianko from the University of Nevada, USA. He described an experiment with two synchronized atomic clocks (19 significant figures accuracy, A. Derevianko, M. Pospelov, *Nature Phys.*, 10(2014)933). Apparently, with such an extraordinary accuracy the researchers were able to detect (by gravitational desynchronization) a passing of huge (∼200 m) objects – some lumps of the invisible dark matter right through the lab. With such news I am still unable to stay calm.

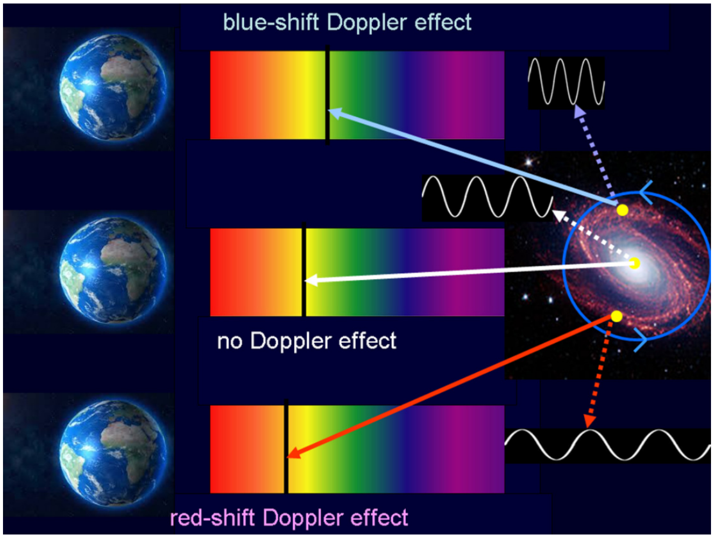


Fig. 5.5. Dark matter. A galaxy (the luminous object in the last column of the figure) is composed of billions of stars. A star's interior (very hot) is assumed to emit white light, which, using a glass prism on Earth, may be decomposed into a rainbow (shown in the second column of the figure). However, a star's outer part is relatively cold, and sodium atoms can exist there. Such an atom may absorb the outgoing interior light, thus making the rainbow incomplete (see the black line in the central part). We know very well such a sodium characteristic line from our laboratories. Since the galaxy rotates (in this figure we neglect other motions possible), the three stars indicated by small circles represent three qualitatively different scenarios among the galaxy's stars. The middle star produces the Doppler unshifted absorption line, the upper star moves towards the Earth and therefore squeezes the light wave length (blue-shifted Doppler effect), and the lower star escapes from the Earth, which results in the Doppler red-shift of the absorption line. Knowing the wave length change detected, we are able to calculate each star's velocity v (its component along the observation line) from the Doppler shift. Calculating the mean value of v and using the virial theorem of the present chapter, we are able to tell the galaxy's mass. It turns out that such a value is many times larger than the galaxy's mass estimated based on the galaxy's light intensity. This is why the astronomers were forced to postulate the existence of some dark matter (its origin currently unknown).

$\left\langle\sum_{i=1}^{N}\left(\mathbf{r}_i\cdot\hat{\mathbf{F}}_i\right)\right\rangle$ calculated for a *stationary* state ψ, e.g., $\left\langle\hat{T}\right\rangle\equiv\left\langle\psi|\hat{T}\psi\right\rangle$, etc.:

$$\boxed{2\left\langle\hat{T}\right\rangle+\left\langle\sum_{i=1}^{N}\left(\mathbf{r}_i\cdot\hat{\mathbf{F}}_i\right)\right\rangle=0. \qquad (5.40)}$$

On the other hand, as shown in Appendix I Eq. (5.39), if the potential energy comes from the pairwise interparticle interaction of the form $V = Ar^n$, the virial theorem may be rewritten[40] as

$$\frac{2\left\langle\hat{T}\right\rangle}{\langle V\rangle} = n, \tag{5.41}$$

where $\left\langle\hat{T}\right\rangle$ stands for the mean kinetic energy and $\langle V\rangle$ is the mean potential energy. For atoms and molecules $n = -1$ (because of Coulombic forces).

5.4.4 A review of examples

Example 2 (Exact solution for the harmonic oscillator). Let us take the exact ground-state solution (see p. 219) $\psi_0(q) = \frac{1}{\sqrt[4]{\pi}}\exp(-\frac{1}{2}q^2)$ and calculate $\left\langle\hat{T}\right\rangle = -\frac{1}{2}\hbar\omega\left\langle\psi_0|\frac{\mathrm{d}^2}{\mathrm{d}q^2}\psi_0\right\rangle = \frac{1}{4}\hbar\omega$ and $\langle V\rangle = \frac{1}{4}\hbar\omega$. We get correctly an exact ground-state energy $E_0 = \left\langle\hat{T}\right\rangle + \langle V\rangle = \frac{1}{2}\hbar\omega$. We see that again the virial theorem given by Eq. (5.41), $\frac{2\left\langle\hat{T}\right\rangle}{\langle V\rangle} = 2$, is exactly satisfied (for the harmonic potential $n = 2$).

Example 3 (Particle in a (rigid) box (the virial theorem not satisfied)). Let us take an exact wave function $\psi_n = \sqrt{\frac{2}{L}}\sin\frac{n\pi}{L}x$ for a particle moving along the x axis in a box $(0, L)$, where inside the box we have $V = 0$ (and $V = \infty$ outside). The particle is confined within $(0, L)$. We check Eq. (I.15). First, let us take $\frac{\mathrm{d}\left\langle\sum_{i=1}^{N}\mathbf{r}_i\cdot\hat{\mathbf{p}}_i\right\rangle}{\mathrm{d}t} = \frac{\mathrm{d}\langle x\hat{p}_x\rangle}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left\langle\psi_n|x\hat{p}_x\psi_n\right\rangle = 0$, because $\left\langle\psi_n|x\hat{p}_x\psi_n\right\rangle$ does not depend on time. We calculate $\left\langle\hat{T}\right\rangle = \left\langle\sqrt{\frac{2}{L}}\sin\frac{n\pi}{L}x| -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\sqrt{\frac{2}{L}}\sin\frac{n\pi}{L}x\right\rangle = \frac{n^2h^2}{8mL^2}$ and $\langle V\rangle = 0$, because $V = 0$ in $(0, L)$. This means that the virial theorem of Eq. (I.15) is not satisfied! What happens?

There is one problem though. At $x = 0$ and $x = L$ the potential goes to infinity, which means that the forces $\hat{\mathbf{F}}_i$ may go there to infinity! But we did not inform the virial about this... Instead it looked as if the forces were equal zero everywhere! Our result means a penalty we pay for this.

Example 4 (Particle in a box with soft walls (the virial theorem satisfied)). It looks as if to satisfy the virial theorem one should allow a *free* interplay between the particle's kinetic energy (trying to make the particle fly to infinity) while the potential energy forces the particle to stay in a region and finally stabilizes the system. Let us check this idea and change

---

[40] Note that V of Eq. (5.41) cannot be shifted in the energy scale; it has to have the form $V = Ar^n$.

a bit a box $(0, L)$ by allowing the particle to move rather easily in the middle of the box, but meeting an increasing repulsion when approaching the walls of the box. Of course we have a lot of freedom in such choice, but we could use as such a potential energy that one which we have derived for the supersymmetry (SUSY) partner on p. 256. This SUSY potential $V_2(x) = \frac{\hbar^2\pi^2}{2mL^2}(2\mathrm{ctg}^2\frac{\pi}{L}x + 1)$ goes to infinity as the box potential did, but this time more gently, as seen in Fig. 4.28b. Let us check the virial theorem for the exact ground-state wave function derived on p. 255, $\psi_1^{(2)} = N'(\sin\frac{\pi}{L}x)^2$. We need to calculate the mean value of the virial of Eq. (I.16): $\left\langle\sum_{i=1}^{N}\left(\mathbf{r}_i \cdot \hat{\mathbf{F}}_i\right)\right\rangle = \left\langle x(-\frac{\mathrm{d}V_2}{\mathrm{d}x})\right\rangle$. Using Mathematica we easily get $N' = \sqrt{\frac{8}{3L}}$, the virial $\left\langle x(-\frac{\mathrm{d}V_2}{\mathrm{d}x})\right\rangle = -\frac{4\hbar^2\pi^2}{3mL^2}$, and the mean kinetic energy $\left\langle\hat{T}\right\rangle = \frac{2\hbar^2\pi^2}{3mL^2}$. We check Eq. (I.16): $2\left\langle\hat{T}\right\rangle + \left\langle\sum_{i=1}^{N}\left(\mathbf{r}_i \cdot \hat{\mathbf{F}}_i\right)\right\rangle = 2\frac{2\hbar^2\pi^2}{3mL^2} - \frac{4\hbar^2\pi^2}{3mL^2} = 0$, i.e., the virial theorem is satisfied! Note that the system was allowed to have a free interplay of the kinetic and potential energy.

Example 5 (Rigid rotating dipole – the virial theorem not satisfied). The exact solution of the Schrödinger equation for two point masses (their reduced mass is μ) *preserving the interparticle distance* r has been given on p. 229. This means the rotator is infinitely stiff, when we try to change r. This already means that we will be unable to calculate the forces (they go to infinity) that enter the virial theorem, Eq. (I.16), and *we expect troubles with the virial theorem*. For our future analysis let us assume additionally that the two masses have opposite electric charges (forming a dipole): $+1$ and -1. This does not change anything in the solution of the problem. The only difference is that instead of $V = 0$ now we will have $V = -\frac{1}{r} = \mathrm{const}$. Therefore, in the Schrödinger equation $\hat{H}Y_J^M = E_J Y_J^M$ with $\hat{H} = \hat{T} + V$ we will have $\hat{T}Y_J^M = E'Y_J^M$, exactly as it was for the rigid rotator, but with another constant value of the energy: $E'_J = E_J - V = E_J + \frac{1}{r} = J(J+1)\frac{\hbar^2}{2\mu r^2} + \frac{1}{r}$, where the rotational quantum number $J = 0, 1, 2, \ldots$. With constant V the force is equal to zero and the mean value of the virial $\left\langle\sum_{i=1}^{N}\left(\mathbf{r}_i \cdot \hat{\mathbf{F}}_i\right)\right\rangle = 0$. While $2\left\langle\hat{T}\right\rangle = J(J+1)\frac{\hbar^2}{\mu r^2} + \frac{2}{r}$ is never equal zero, there is no virial to cancel $2\left\langle\hat{T}\right\rangle$ and the virial equation (I.16) is not satisfied. Note that there is no free interplay of the kinetic and potential energies: the rotator has been kept rigid ($r = \mathrm{const}$). What if we made r variable? Well, in such a case our rotator changes to the hydrogen atom!

Example 6 (Flexible rotating dipole which is identical to the hydrogen atom – the virial theorem satisfied). The potential energy (the Coulomb law) $V = -\frac{1}{r}$ is of the kind $V = Ar^n$, $n = -1$. Therefore, if the virial theorem is to be satisfied, we should get $\frac{2\left\langle\hat{T}\right\rangle}{\langle V\rangle} = -1$. Let us check this.

An exact ground-state solution of the Schrödinger equation for the hydrogen atom is given by Eq. (4.48) on p. 234, i.e.,

$$1s \equiv \frac{1}{\sqrt{\pi}} \exp(-r), \tag{5.42}$$

where r represents the nucleus–electron distance. We find in Appendix H $\left\langle \hat{T} \right\rangle = \frac{1}{2}$ and $\langle V \rangle = -1$ (a.u. are used). Thus, $\frac{2\left\langle \hat{T} \right\rangle}{\langle V \rangle} = -1$ and indeed the virial theorem is exactly satisfied. Let us note that the mean value of the Hamiltonian $\left\langle \hat{H} \right\rangle = \left\langle \hat{T} \right\rangle + \langle V \rangle = -\frac{1}{2}$, which represents an exact result (assuming the infinite proton mass) (see p. 233).

We have proved that the same should happen for any stationary state. Let us check the first excited state $2s$ of Eq. (4.49),

$$2s \equiv N_{2s}\,(r-2)\exp(-r/2) \tag{5.43}$$

with the normalization constant $N_{2s} = \frac{1}{4\sqrt{2\pi}}$.

We calculate using Eq. (H.1):

$$\left\langle \hat{T} \right\rangle = -\frac{1}{2}\langle \Delta \rangle =$$
$$-\frac{1}{2}N_{2s}^{2}\int (r-2)\exp(-\frac{r}{2})\left[\frac{1}{r^{2}}\frac{\partial}{\partial r}r^{2}\frac{\partial}{\partial r}\right](r-2)\exp(-\frac{r}{2})\mathrm{d}V.$$

The integration[41] gives $\left\langle \hat{T} \right\rangle = \frac{1}{8}$ a.u. and then $\langle V \rangle = -\left\langle \frac{1}{r} \right\rangle = -\frac{1}{4}$ a.u. Again, $\frac{2\left\langle \hat{T} \right\rangle}{\langle V \rangle} = -1$, the virial theorem is satisfied exactly. Also, the total energy $\left\langle \hat{H} \right\rangle = \left\langle \hat{T} \right\rangle + \langle V \rangle = -\frac{1}{8}$, exactly as it should be for the $2s$ state (see p. 233).

Again, allowing a free interplay between kinetic and potential energies leads to the virial theorem being satisfied.

5.4.5 What about the mean values calculated with an approximate solution?

5.4.5.1 A link to the variational principle – scaling the trial wave function

The examples given above are not very helpful in quantum chemistry, because they pertain to exact solutions, never met in practice. We have learned however that for the exact solutions, for *any bound system free in its motion,* the virial theorem *is* satisfied.

---

[41] For example, by using Mathematica.

Is it possible that a degree of satisfaction of the virial theorem represents a hint about the quality of an approximate wave function? An important source of approximate wave functions represents the variational method (see p. 269), where we have to do with a *class* of trial functions, each of them corresponding to a set of variational parameters. It will be shown below that

> for the trial function with the optimal set of the parameters (i.e., minimizing the mean value of the energy) the virial theorem is exactly satisfied. This satisfaction takes place independently of whether the obtained optimal wave function is an exact or an approximate one.

Suppose we decide to use a single variational parameter η in order to create a class of the normalized variational trial functions (p. 265). To do that, we start from a normalized function $\psi(\mathbf{r}_1, ...\mathbf{r}_N)$, where $\mathbf{r}_i$ stands for the position of the i-th electron, and scale each vector $\mathbf{r}_i$ by multiplying it by η: $\mathbf{r}_i \longrightarrow \boldsymbol{\rho}_i = \eta\mathbf{r}_i$, i.e., $\rho_{xi} = \eta x_i$, $\rho_{yi} = \eta y_i$, $\rho_{zi} = \eta z_i$ for $i = 1, ...N$. To guarantee that the functions are normalized for $\eta \neq 1$ we have to have

$$\psi_\eta = N'\psi(\eta\mathbf{r}_1, ...\eta\mathbf{r}_N)$$

with the normalization constant[42] $N' = \eta^{\frac{3N}{2}}$.

Now, let us calculate the mean value of energy (see p. 266):

$$\varepsilon = \left\langle\psi_\eta|\hat{H}\psi_\eta\right\rangle = \left\langle\psi_\eta|\left(\hat{T} + V\right)\psi_\eta\right\rangle = \left\langle\psi_\eta|\hat{T}\psi_\eta\right\rangle + \langle\psi_\eta|V\psi_\eta\rangle.$$

One gets[43] $\langle T\rangle_\eta \equiv \left\langle\psi_\eta|\hat{T}\psi_\eta\right\rangle = \eta^2\langle T\rangle_{\eta=1}$ and for the Coulombic potential $\langle V\rangle_\eta \equiv \langle\psi_\eta|V\psi_\eta\rangle = \eta\langle V\rangle_{\eta=1}$.

---

[42] Normalization:

$$|N'|^2\int\psi^*(\eta\mathbf{r}_1, ...\eta\mathbf{r}_N)\psi(\eta\mathbf{r}_1, ...\eta\mathbf{r}_N)\mathrm{d}\mathbf{r}_1...\mathrm{d}\mathbf{r}_N = |N'|^2\frac{1}{\eta^{3N}}\int\psi^*(\boldsymbol{\rho}_1, ...\boldsymbol{\rho}_N)\psi(\boldsymbol{\rho}_1, ...\boldsymbol{\rho}_N)\mathrm{d}\boldsymbol{\rho}_1...\mathrm{d}\boldsymbol{\rho}_N$$
$$|N'|^2\frac{1}{\eta^{3N}}\int\psi^*(\mathbf{r}_1, ...\mathbf{r}_N)\psi(\mathbf{r}_1, ...\mathbf{r}_N)\mathrm{d}\mathbf{r}_1...\mathrm{d}\mathbf{r}_N = |N'|^2\frac{1}{\eta^{3N}}\cdot 1 = 1.$$

Hence, we may take $N' = \eta^{\frac{3N}{2}}$ as the normalization constant.

[43] Note that $\frac{\partial\rho_{xi}}{\partial x_i}\frac{\partial}{\partial\rho_{xi}} = \eta\frac{\partial}{\partial\rho_{xi}}$, $\frac{\partial^2}{\partial x_i^2} = \eta^2\frac{\partial^2}{\partial\rho_{xi}^2}$ and $\triangle_{xi,yi,zi} = \eta^2\triangle_{\rho_{xi},\rho_{yi},\rho_{zi}}$. Therefore,

$$\eta^{3N}\left[-\frac{1}{2}\sum_i\langle\psi(\eta\mathbf{r}_1, ...\eta\mathbf{r}_N)|\triangle_{xi,yi,zi}\psi(\eta\mathbf{r}_1, ...\eta\mathbf{r}_N)\rangle\right] =$$

Thus, the total energy calculated with ψ_η is equal to

$$\varepsilon(\eta) = \eta^2 \langle T \rangle_{\eta=1} + \eta \langle V \rangle_{\eta=1} . \tag{5.44}$$

When minimizing $\varepsilon(\eta)$ we get

$$0 = 2\eta_{\text{opt}} \langle T \rangle_{\eta=1} + \langle V \rangle_{\eta=1} , \tag{5.45}$$

where

$$\eta_{\text{opt}} = -\frac{\langle V \rangle_{\eta=1}}{2\langle T \rangle_{\eta=1}} . \tag{5.46}$$

Using this value to obtain the lowest energy among the trial functions belonging to the chosen variational class we end up with

$$\varepsilon_{\text{opt}} = \varepsilon(\eta_{\text{opt}}) = \eta_{\text{opt}}^2 \langle T \rangle_{\eta=1} + \eta_{\text{opt}} \langle V \rangle_{\eta=1} = \left(\frac{\langle V \rangle_{\eta=1}}{2\langle T \rangle_{\eta=1}} \right)^2 \langle T \rangle_{\eta=1} - \frac{\langle V \rangle_{\eta=1}}{2\langle T \rangle_{\eta=1}} \langle V \rangle_{\eta=1} =$$
$$\frac{1}{4} \frac{\left(\langle V \rangle_{\eta=1}\right)^2}{\langle T \rangle_{\eta=1}} - \frac{1}{2} \frac{\left(\langle V \rangle_{\eta=1}\right)^2}{\langle T \rangle_{\eta=1}} = -\frac{1}{4} \frac{\left(\langle V \rangle_{\eta=1}\right)^2}{\langle T \rangle_{\eta=1}}$$

and for $\eta = \eta_{\text{opt}}$ we get the virial theorem (*for Coulombic potential*) satisfied for the optimal scaling parameter[44]

$$\eta^{3N} \left[-\frac{1}{2} \sum_i \left\langle \psi(\eta \mathbf{r}_1, \ldots \eta \mathbf{r}_N) | \eta^2 \triangle_{\rho_{xi}, \rho_{yi}, \rho_{zi}} \psi(\eta \mathbf{r}_1, \ldots \eta \mathbf{r}_N) \right\rangle \right] =$$
$$-\frac{1}{2} \eta^2 \eta^{3N} \sum_i \int \psi^*(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) | \triangle_{\rho_{xi}, \rho_{yi}, \rho_{zi}} \psi(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) \mathrm{d}\mathbf{r}_1 \ldots \mathrm{d}\mathbf{r}_N =$$
$$-\frac{1}{2} \eta^2 \eta^{3N} \sum_i \int \psi^*(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) | \triangle_{\rho_{xi}, \rho_{yi}, \rho_{zi}} \psi(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) \mathrm{d}\mathbf{r}_1 \ldots \mathrm{d}\mathbf{r}_N =$$
$$-\frac{1}{2} \eta^2 \sum_i \int \psi^*(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) | \triangle_{\rho_{xi}, \rho_{yi}, \rho_{zi}} \psi(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) \mathrm{d}\boldsymbol{\rho}_1 \ldots \mathrm{d}\boldsymbol{\rho}_N = \eta^2 \langle T \rangle_{\eta=1} .$$

Then,

$$\langle \psi_\eta | V \psi_\eta \rangle = \eta^{3N} \left\langle \psi(\eta \mathbf{r}_1, \ldots \eta \mathbf{r}_N) | \sum_{i<j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \psi(\eta \mathbf{r}_1, \ldots \eta \mathbf{r}_N) \right\rangle =$$
$$\eta^{3N} \eta^{-3N} \int \psi^*(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) | \sum_{i<j} \eta \frac{e^2}{|\boldsymbol{\rho}_i - \boldsymbol{\rho}_j|} \psi(\boldsymbol{\rho}_1, \ldots \boldsymbol{\rho}_N) \mathrm{d}\boldsymbol{\rho}_1 \ldots \mathrm{d}\boldsymbol{\rho}_N = \eta \langle V \rangle_{\eta=1} .$$

[44] $\frac{2\langle T \rangle_{\eta_{\text{opt}}}}{\langle V \rangle_{\eta_{\text{opt}}}} = \frac{2\eta^2 \langle T \rangle_{\eta=1}}{\eta \langle V \rangle_{\eta=1}} = -\frac{2\langle V \rangle_{\eta=1}}{2\langle T \rangle_{\eta=1}} \frac{\langle T \rangle_{\eta=1}}{\langle V \rangle_{\eta=1}} = -1.$

COULOMBIC POTENTIAL

$$\frac{2\langle T\rangle_{\eta_{\mathrm{opt}}}}{\langle V\rangle_{\eta_{\mathrm{opt}}}} = -1. \tag{5.47}$$

Thus, we have shown, for the first time in this book, that the virial theorem is *exactly* satisfied for an *approximate* wave function describing the system. Note that this happened for a specific wave function, which, when scaling uniformly all distances, minimizes (*among a class of trial functions*) the mean value of the total energy (variational method).

Example 7 (The hydrogen atom – an approximate wave function, but still the virial theorem satisfied!). Let us consider a set of the normalized Gaussian functions $\psi_\eta = \left(\frac{2\eta}{\pi}\right)^{\frac{3}{4}} \exp(-\eta r^2)$. *There is no chance to find an exact solution of the Schrödinger equation taking* ψ in this form. We know that for sure, because we know the exact solution and it has the form $1s = N\exp(-r)$, which is not a single Gaussian function. Still let us calculate the energy as the expectation value using ψ_η. As shown in Appendix Q on p. 729, Eqs. (Q.2) and (Q.3), we get as the mean value of energy: $\varepsilon(\eta) = \left\langle \hat{T} \right\rangle_\eta + \langle V\rangle_\eta = \frac{3\eta}{2} - 2\sqrt{\frac{2\eta}{\pi}}$, $\left\langle \hat{T} \right\rangle_\eta = \frac{3\eta}{2}$, $\langle V\rangle_\eta = -2\sqrt{\frac{2\eta}{\pi}}$. We may find the optimal η from $\frac{\mathrm{d}\varepsilon}{\mathrm{d}\eta} = 0$. This gives $0 = \frac{3}{2} - \sqrt{\frac{2}{\pi\eta_{\mathrm{opt}}}}$, hence $\eta_{\mathrm{opt}} = \frac{8}{9\pi}$. The mean value of the energy is equal to $\varepsilon(\eta_{\mathrm{opt}}) = \frac{3\frac{8}{9\pi}}{2} - 2\sqrt{\frac{2\frac{8}{9\pi}}{\pi}} = -\frac{4}{3\pi} = -0.4244132$ a.u. $> -\frac{1}{2}$ a.u. $= \varepsilon_{\mathrm{exact}}$, i.e., it is greater than the exact ground-state energy.

Let us calculate the virial ratio: $\frac{2\left\langle \hat{T} \right\rangle_\eta}{\langle V\rangle_\eta} = -\frac{3\eta}{2\sqrt{\frac{2\eta}{\pi}}} = -\frac{3\sqrt{\eta\pi}}{2\sqrt{2}}$. For $\eta = \eta_{\mathrm{opt}}$ we have $\frac{2\left\langle \hat{T} \right\rangle_{\eta_{\mathrm{opt}}}}{\langle V\rangle_{\eta_{\mathrm{opt}}}} = -1$. *For the first time in our virial story, we have an example showing that the virial theorem is satisfied exactly, even if the wave function differs from the exact one!* In the present case, similarly as in the examples given above, the system is again fully relaxed, but this time only within a fixed form of the trial functions.

Example 8 (Morse potential: the bound and nonbound states). What about a potential that is not of the kind given by Eq. (5.38)? Let us take a Morse potential (p. 224), shown in Fig. 5.6 (with parameters $D = 0.2, \alpha = 2$).

Such a potential leads to a number of bound states[45] that resemble the states of the harmonic oscillator, especially the low-energy ones. In particular, the lowest-energy solution is nodeless

---

[45] For a particle of mass $\mu = \frac{m_p}{2}$, this is a reduced mass for oscillations of the hydrogen molecule, $m_p = 1840$, which is the mass of a proton.

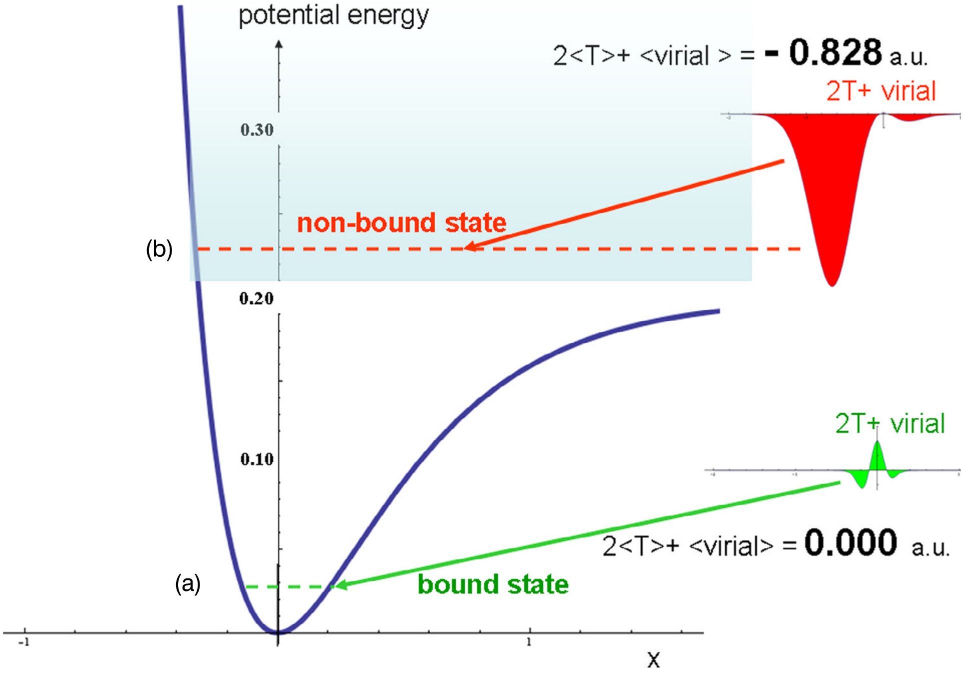


Fig. 5.6. The Morse potential has a finite number of bound states and an infinite number of non-bound ones (a continuum shown above the dissociation limit). In the figure Eq. (5.40) is examined for the two states. (a) The variationally optimized Gaussian function, which (right-hand side lower insert) corresponds to the mean energy below the dissociation limit. (b) A much more diffuse Gaussian function (right-hand side upper insert) corresponding to the mean energy above the dissociation limit. Only in the first case the virial equation (5.40) is satisfied. For illustration the corresponding integrands, 2T+virial, are plotted (in the same scale, right-hand side). The parameters of the Morse potential (see p. 224) have been chosen as $D = 0.2$ a.u., $\alpha = 2$ (the potential has been shifted upwards by D).

and therefore resembles[46] a single Gaussian function $\psi_\eta = \sqrt{\frac{2\eta}{\pi}} \exp(-\eta x^2)$, the latter being an exact ground-state function for the harmonic oscillator. The mean energy $\bar{E} = \left\langle \hat{T} + V \right\rangle$ depends on η and $\bar{E}(\eta)$ covers the energy range from its optimized value (obtained by the variational method, in the example given $\eta_{\text{opt}} = 23.5$) with $\bar{E}\left(\eta_{\text{opt}}\right) = 0.02444$ a.u. up to very high values. Therefore, $\psi_{\eta_{\text{opt}}}$ mimics the ground-state wave function, which is bound. If the potential were the harmonic one, we would get for the exact wave function $\frac{\langle \hat{T} \rangle}{\langle V \rangle} = 1$. Note that

---

[46] The corresponding exact Morse ground-state wave function is a bit distorted with respect to ψ_η.

for the Morse potential and $\psi_{\eta_{\text{opt}}}$ we get instead $\frac{\langle \hat{T} \rangle}{\langle V \rangle} = 1.09$, a quite close value. The bound state under consideration should satisfy the virial equation (5.40), and in our case it is satisfied to the accuracy of three significant figures (see Fig. 5.6). The virial theorem (5.40) does not need to be satisfied for the continuum states (above the dissociation limit shown in the figure). We will (quite naively) mimic such a state by a Gaussian function[47] with an energy higher than the dissociation limit ($\bar{E} = 0.23$ a.u.) and indeed we see that Eq. (5.40) is far from being satisfied (we obtain -0.828 a.u. instead of 0, see Fig. 5.6).

5.4.6 Quantum chemistry: how useful is the virial theorem?

In contemporary quantum chemistry one often ends up with wave functions, which contain thousands or even many millions of linear and nonlinear parameters. Typically these nonlinear parameters appear as exponential coefficients α multiplying the squares of electronic position vectors, e.g., $\exp(-\alpha r_i^2)$. It would be therefore a formidable job to correct the α coefficients found by an energy minimization in the space of all the parameters. On the other hand, such a wave function is often already close to the exact one, but still can be improved relatively easily by introducing a uniform scaling of the electronic position vectors by a *single* parameter η: $\mathbf{r}_i \longrightarrow \boldsymbol{\rho}_i = \eta \mathbf{r}_i$. The value $\eta = 1$ corresponds to the function we are going to improve, for which $\frac{2\langle \hat{T} \rangle_{\eta=1}}{\langle V \rangle_{\eta=1}}$ deviates from -1. Then, one adjusts η to satisfy $\frac{2\langle \hat{T} \rangle_{\eta_{\text{opt}}}}{\langle V \rangle_{\eta_{\text{opt}}}} = -1$ and the initial function is improved by introducing $\boldsymbol{\rho}_i = \eta_{\text{opt}} \mathbf{r}_i$.

Such scaling has usually little effect on the total energy, but the effect is two or even three orders of magnitude larger, if one considers the components of the total energy. These components are often of importance for physical interpretation of chemical phenomena; therefore the scaling procedure helps to ensure a proper balance of the kinetic and potential energies.

Summary

There are basically three numerical approaches to obtain approximate solutions to the Schrödinger equation, i.e., variational, perturbational and the method of moments. In calculations we usually apply variational and moment methods, while the perturbational approach is often applied to estimate small physical effects. The result is that most concepts (practically all we know) characterizing the reaction of a molecule to an external field come from the perturbational approach. This leads to such quantities (see Chapter V2-4) as dipole moment, polarizability, hyperpolarizability, etc. The computational role of perturbational theories may, in this context, be seen as being of the second order.

---

[47] This mimicking fails in that a single Gaussian function is normalizable, while any continuum function cannot be normalized.

- **Variational method**
  - The method is based on the variational principle, which says that if for a system with Hamiltonian $\hat{H}$ we calculate the number $\varepsilon = \frac{\left\langle \Phi | \hat{H} \Phi \right\rangle}{\langle \Phi | \Phi \rangle}$, where Φ *stands for an arbitrary function*, then the number $\varepsilon \geq E_0$, with E_0 being the ground-state eigenvalue of $\hat{H}$. If it happens that $\varepsilon[\Phi] = E_0$, then there is only one possibility: Φ represents the exact ground-state wave function ψ_0.
  - The variational principle means that to find an approximate ground-state wave function we can use the *variational method*: minimize $\varepsilon\left[\Phi\right]$ by changing (varying) Φ. The minimum value of $\varepsilon\left[\Phi\right]$ is equal to $\varepsilon\left[\Phi_{opt}\right]$ which approximates the ground-state energy E_0 and corresponds to Φ_{opt}, i.e., an approximation to the ground-state wave function ψ_0.
  - In practice the variational method consists of the following steps:
    - make a decision as to the *trial function class,* among which the $\Phi_{opt}\left(\mathbf{x}\right)$ will be sought[48];
    - introduce into the function the *variational parameters* $\mathbf{c} \equiv (c_0, c_1, ...c_P)$: $\Phi\left(\mathbf{x}; \mathbf{c}\right)$. In this way ε becomes a function of these parameters: $\varepsilon\left(\mathbf{c}\right)$;
    - minimize $\varepsilon\left(\mathbf{c}\right)$ with respect to $\mathbf{c} \equiv (c_0, c_1, ...c_P)$ and find the optimal set of parameters $\mathbf{c} = \mathbf{c}_{opt}$;
    - the value $\varepsilon\left(\mathbf{c}_{opt}\right)$ represents an approximation to E_0;
    - the function $\Phi\left(\mathbf{x}; \mathbf{c}_{opt}\right)$ is an approximation to the ground-state wave function $\psi_0\left(\mathbf{x}\right)$.
  - The Ritz procedure is a special case of the variational method, in which the parameters $\mathbf{c}$ enter Φ *linearly*: $\Phi\left(\mathbf{x}; \mathbf{c}\right) = \sum_{i=0}^{P} c_i \Psi_i$, where $\{\Psi_i\}$ are some *known* basis functions that form (or, more exactly, in principle form) the complete set of functions in the Hilbert space. This formalism leads to a set of homogeneous linear equations to solve ("secular equations"), from which we find approximations to the ground- and excited-state energies and wave functions.
- **Method of moments** Quantum chemistry produces equations to be solved in order to find the wave function Ψ. In such an equation a complicated function G (that depends on Ψ) belonging to the Hilbert space has to satisfy $G = 0$. Function Ψ depends on some parameters of unknown values, their role is to assure $G = 0$. One assumes first some (in practice finite and therefore incomplete) basis set of known functions $\{\varphi_i\}$ representing vectors of the Hilbert space, their number equal to the unknown parameters of Ψ. Then, for each of these vectors a projection[49] $\langle \varphi_i | G \rangle$ has to be forced to be 0. This means a set of equations $\langle \varphi_i | G \rangle = 0$ for the parameters to be solved. Once the optimal parameters are found the approximate wave function Ψ is ready.

  In the present book we will have to do quite often with a manifold of functions $G(\mathbf{x}; c_{0,} c_1, ...c_P)$, where G belongs to the Hilbert space, $\mathbf{x}$ stands for coordinates, while $c_{0,} c_1, ...c_P$ are some coefficients. The coefficients $c_{0,} c_1, ...c_P$ that imply the equality

$$G(\mathbf{x}; c_{0,} c_1, ...c_P) = 0 \tag{5.48}$$

---

[48] $\mathbf{x}$ symbolizes the set of coordinates (space and spin, cf. Chapter 1).

[49] Such a projection is sometimes called "a moment".

represent the target of the method of moments. The coefficients may enter G in a complex way and searching the solution of Eq. (5.48) in the Hilbert space may be very difficult. We practically never do such a search.

Instead the method of moments may be applied. In this method one has to have a complete set of the basis functions $\{\varphi_i\}$, of which only $P+1$ will be used for producing $P+1$ scalar products of both sides[50] of Eq. (5.48):

$$\left\langle \varphi_i | G(c_{0,} c_1, ...c_P) \right\rangle = 0 \tag{5.49}$$

for $i = 0, 1, ...P$. Eqs. (5.49) represent an algebraic set of equations for unknown coefficients $c_{0,} c_1, ...c_P$, which is usually much easier to solve than Eq. (5.48).

- **Perturbational method** We assume that the solution to the Schrödinger equation for the *unperturbed system is known* ($E_k^{(0)}$ for the energy and $\psi_k^{(0)}$ for the wave function, usually $k=0$, i.e., the ground state), but when a *small perturbation* $\hat{H}^{(1)}$ is added to the Hamiltonian, then the solution changes (to E_k and ψ_k, respectively) and is to be sought using the perturbational approach. Then the key assumption is $E_k(\lambda) = E_k^{(0)} + \lambda E_k^{(1)} + \lambda^2 E_k^{(2)} + \ldots$ and $\psi_k(\lambda) = \psi_k^{(0)} + \lambda \psi_k^{(1)} + \lambda^2 \psi_k^{(2)} + \ldots$, where λ is a parameter that tunes the perturbation. The goal of the perturbational approach is to compute corrections to the energy: $E_k^{(1)}, E_k^{(2)}, \ldots$, and to the wave function: $\psi_k^{(1)}, \psi_k^{(2)}, \ldots$. We assume that because the perturbation is small, only a few such corrections are to be computed, in particular, $E_k^{(1)} = \left\langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_k^{(0)} \right\rangle$, $\psi_k^{(1)} = \sum_{n \neq k} \frac{H_{nk}^{(1)}}{E_k^{(0)} - E_n^{(0)}} \psi_n^{(0)}$, $E_k^{(2)} = \sum_{n \neq k} \frac{|H_{kn}^{(1)}|^2}{E_k^{(0)} - E_n^{(0)}}$, where $H_{kn}^{(1)} = \left\langle \psi_k^{(0)} | \hat{H}^{(1)} \psi_n^{(0)} \right\rangle$.
- **The virial and the wave function quality** The virial theorem pertains to all bound systems: planets, stars, but also nuclei and electrons. It relates the mean values of the kinetic and of the potential energy. If the potential energy comes from the pairwise interparticle interaction of the form $V = Ar^n$ (A is a constant), the virial theorem may be expressed as

$$\frac{2\left\langle \hat{T} \right\rangle}{\langle V \rangle} = n, \tag{5.50}$$

where $\left\langle \hat{T} \right\rangle$ stands for the mean kinetic energy and $\langle V \rangle$ is the mean potential energy. For atoms and molecules (Coulomb potential), but also for planets, stars and galaxies (gravitation) $n = -1$. Any deviation of $\frac{2\left\langle \hat{T} \right\rangle}{\langle V \rangle}$ from the n value means that the wave function parameters may still be improved. We may watch that easily, because $\frac{2\left\langle \hat{T} \right\rangle}{\langle V \rangle}$ is more sensitive than the calculated total energy (whether variational or nonvariational).

---

[50] This simply means fixing the $P+1$ projections of the function Φ on the first $P+1$ axes in the Hilbert space.

Main concepts, new terms

asymptotic convergence (p. 284)
complete set of functions (p. 271)
corrections to energy (p. 276)
corrections to wave function (p. 276)
Hylleraas equation (p. 282)
Hylleraas functional (p. 280)
Hylleraas variational principle (p. 280)
perturbation (p. 274)
perturbational method (p. 274)
perturbed system (p. 274)
Ritz method (p. 271)
secular equation (p. 273)
secular determinant (p. 273)
trial function (p. 265)
underground states (p. 268)
unperturbed system (p. 274)
variational function (p. 265)
variational method (p. 265)
variational parameters (p. 265)
variational principle (p. 265)
variational principle for excited states (p. 268)
virial theorem (p. 287)

From the research front

In practice, the Ritz variational method is used most often. One of the technical problems to be solved is the size of the basis set. Enormous progress in computation and software development now facilitate investigations which in the year 2000 were absolutely beyond imagination. The world record in quantum chemistry means a few billion expansion functions. To accomplish this quantum chemists have had to invent some powerful methods of applied mathematics.

Ad futurum

The computational technique impetus we witness nowadays will continue in the future (maybe in a modified form). It will be no problem to find some reliable approximations to the ground-state energy and wave function for a molecule composed of thousands of atoms. We will be effective. We may, however, ask whether such effectiveness is at the heart of science. Would it not be interesting to know what these ten billion terms in our wave function are telling us about and what we could learn from this?

Additional literature

I. Mayer, "Simple Theorems, Proofs, and Derivations in Quantum Chemistry," Springer, 2003.

A useful and very well-written book.

W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, "Numerical Recipes. The Art of Scientific Computing," Cambridge University Press, 1986, p. 19–77, 274–326, 335–381.

Probably the best textbook in computational mathematics; some chapters are very closely related to the topics of this chapter (diagonalization, linear equations).

H. Margenau, G.M. Murphy, "The Mathematics of Physics and Chemistry," D. van Nostrand Co., 1956.

An excellent old book dealing with most mathematical problems which we may encounter in chemistry and physics, including the variational and perturbational methods.

J.O. Hirschfelder, W. Byers Brown, S.T. Epstein, "Recent Developments in Perturbation Theory," *Adv. Quantum Chem.*, 1(1964)255.

A long article on perturbation theory. For many years obligatory for those working in the domain.

Questions

1. The variational principle:
   a. says that the mean value of the Hamiltonian calculated with any trial wave function is greater than the experimental ground-state energy of the system.
   b. says that the number $\varepsilon = \frac{\langle \phi | \hat{H}\phi \rangle}{\langle \phi | \phi \rangle}$ computed with any trial wave function ϕ is greater than or equal to the lowest eigenvalue of the Hamiltonian $\hat{H}$.
   c. if $\frac{\langle \phi | \hat{H}\phi \rangle}{\langle \phi | \phi \rangle}$ is equal to the lowest eigenvalue of the Hamiltonian $\hat{H}$, this means ϕ is an exact eigenfunction of $\hat{H}$ corresponding to this eigenvalue.
   d. may lead to nonphysical states of the system.
2. The variational method (where $\hat{H}$ stands for the Hamiltonian, ϕ is a trial function, and ε denotes the mean value of energy):
   a. means minimization of functional $\varepsilon[\phi] = \frac{\langle \phi | \hat{H}\phi \rangle}{\langle \phi | \phi \rangle}$.
   b. variational parameters are introduced in the trial function ϕ and ε becomes a function of these parameters.
   c. we minimize ϕ as a function of variational parameters.
   d. we search for the global minimum of ε as a function of variational parameters. The set of the optimum parameters gives the best approximation to the ground-state wave function and to the energy of the ground state.
3. The variational method:
   a. if a variational trial function is orthogonal to the exact eigenfunctions of all the states with energy $E < A$, the variational principle pertains the state with the lowest eigenvalue of energy, which satisfies $E \geq A$.
   b. cannot be used if perturbation is large.
   c. if the variational wave function transforms according to the irreducible representation Γ of the symmetry group of the Hamiltonian, the variational method will find an approximation to the lowest-energy state with the wave function transforming according to Γ.
   d. is applicable for the ground state only.
4. The Ritz method (where $\hat{H}$ stands for the Hamiltonian, ϕ is a variational trial function, and Ψ_i denote the basis functions):

   a. is a variational method in which the trial function has the form $\phi = \sum_{i=1}^{P} c_i \Psi_i$, where Ψ_i are the eigenfunctions of the Hamiltonian, while c_i are unknown coefficients.
   b. is a variational method, in which the trial function has the form $\phi = \sum_{i=1}^{P} c_i \Psi_i$, where Ψ_i are known basis functions, while c_i are unknown coefficients.
   c. the optimal variational coefficients are found by solving the secular equations.
   d. approximate eigenvalues of the Hamiltonian are obtained as solutions of the equation $\det\left(H_{ij} - \varepsilon S_{ij}\right) = 0$, where $H_{ij} = \left\langle \Psi_i | \hat{H} \Psi_j \right\rangle$ and $S_{ij} = \left\langle \Psi_i | \Psi_j \right\rangle$.
5. Consider the Ritz method (where $\hat{H}$ stands for the Hamiltonian, ϕ is a variational trial function, Ψ_i denote the basis functions, and P means the number of variational parameters):
   a. the number of integrals $H_{ij} = \left\langle \Psi_i | \hat{H} \Psi_j \right\rangle$ needed in computation is equal to $P(P+1)$.
   b. functions Ψ_i have to be orthonormal.
   c. the approximate eigenvalues of the Hamiltonian are computed by diagonalization of the Hamiltonian matrix: $H_{ij} = \left\langle \Psi_i | \hat{H} \Psi_j \right\rangle$.
   d. the matrix of elements H_{ij} is Hermitian and its eigenvalues are always real numbers.
6. Consider the virial theorem in classical mechanics (T stands for the kinetic energy, V is the potential energy, $\mathbf{r}_i$ and $\mathbf{F}_i$ represent the position vector and the force acting on particle i, and N means the number of particles):
   a. the quantity $\sum_{i=1}^{N} \mathbf{r}_i \cdot \mathbf{F}_i$ represents what is known as the virial.
   b. in any mechanical system one has $2\langle T \rangle_\tau + \sum_{i=1}^{N} \langle \mathbf{r}_i \cdot \mathbf{F}_i \rangle_\tau =$ minimum, where $\langle\rangle_\tau$ means averaging over a large time τ.
   c. for gravitational as well as electrostatic interactions of point-like particles one has $2\langle T \rangle_\tau = -\langle V \rangle_\tau$.
   d. for harmonic pairwise interactions $\langle T \rangle_\tau = \langle V \rangle_\tau$.
7. Consider the virial theorem in quantum mechanics (where T stands for the kinetic energy, V is the potential energy, $\mathbf{r}_i$ and $\mathbf{F}_i$ represent the position vector and the force acting on particle i, N means the number of particles, and $\langle\rangle$ stands for the quantum mechanical averaging):
   a. for any exact state $\frac{2\langle T \rangle}{\langle V \rangle} = -1$.
   b. for any variationally optimized trial wave function $\psi_\eta = N' \psi(\eta \mathbf{r}_1, ... \eta \mathbf{r}_N)$ with the scaling parameter η: $\frac{2\langle T \rangle_{\eta_{\mathrm{opt}}}}{\langle V \rangle_{\eta_{\mathrm{opt}}}} = -1$.
   c. for exact stationary states $\frac{2\langle T \rangle}{\langle V \rangle} = -1$ provided that the Hamiltonian corresponds to the Coulomb interactions of the point-like particles.
   d. for any exact solution of the Morse oscillator one has $\frac{\langle T \rangle}{\langle V \rangle} = 1$.
8. In the perturbational approach[51]:
   a. ψ_k is an eigenfunction of $\hat{H}$, while $\psi_k^{(0)}$ is an eigenfunction of $\hat{H}^{(0)}$.

---

[51] $\hat{H}$ means the Hamiltonian of the system, $\hat{H}^{(0)}$ stands for the unperturbed Hamiltonian, $\hat{H}^{(1)}$ denotes the perturbation, ψ_k means the wave function of the perturbed state k, $\psi_k^{(0)}$ denotes the wave function of the unperturbed state k – both functions normalized, E_k and $E_k^{(0)}$ mean the corresponding energies, and $\psi_k^{(n)}$ and $E_k^{(n)}$ stand for the n-th correction to the wave function and energy, respectively.

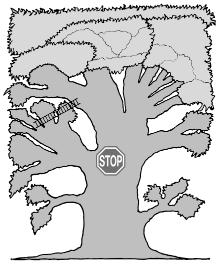
   b. intermediate normalization means that in the equality $\psi_k = c_k^{(0)}\psi_k^{(0)} + \chi$ the integral $\langle \chi | \psi_k \rangle = 0$, while $c_k^{(0)} = 1$.
   c. intermediate normalization means that in equality $\psi_k = c_k^{(0)}\psi_k^{(0)} + \chi$ the integral $\left\langle \chi | \psi_k^{(0)} \right\rangle = 0$, while $c_k^{(0)} = 1$.
   d. $E_k = E_k^{(0)} + E_k^{(1)}$.

9. In the perturbational approach (for notation, see question 6):
   a. $E_k^{(0)} + E_k^{(1)} = \left\langle \psi_k^{(0)} | \hat{H} \psi_k^{(0)} \right\rangle$.
   b. $E_0^{(2)} \leq 0$.
   c. $E_k^{(2)} \leq 0$.
   d. $E_k^{(2)} = \sum_{n(\neq k)}^{\infty} \frac{\left| \hat{H}_{kn}^{(1)} \right|^2}{E_k^{(0)} - E_n^{(0)}}$.

10. In the perturbational approach (for notation, see question 6):
   a. from the intermediate normalization it follows that $\left\langle \psi_k^{(n)} | \psi_k^{(0)} \right\rangle = 0$ for $n > 0$.
   b. $E_0^{(0)} + E_0^{(1)} > E_0$.
   c. $|E_k^{(n-1)}| \geq |E_k^{(n)}|$.
   d. to calculate $E_0^{(2)}$ one has to know $\psi_0^{(2)}$.

Answers

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



CHAPTER 6

A Key Concept of Molecular 3D Structure – Separation of Electronic and Nuclear Motions

> *Any separation is a link.*
> **Simone Weil**

Where are we?

We are on the *most important* branch of the TREE.

An example

A professor introduces his students to chemistry. He came to the lecture room with a model in his hands: a stick-and-balls three-dimensional model of the vitamin C molecule (ascorbic acid, Fig. 6.1). In doing so he tacitly imprints in his students' minds an idea, the most important one in chemistry: molecules *have* some spatial shape, similarly as has a table or a chair. This idea, extremely successful in chemistry, is derived in the present chapter, but only under the assumption that nuclei are infinitely heavier than electrons (which represents an approximation) and the electronic states are well separated in the energy scale, which in some cases may not be true. A molecule (such as ascorbic acid, shown in Fig. 6.1) rotates in space and at the same time oscillates (this will be the subject of the present chapter). *The spatial relationships that can be deduced from such a picture represent the basis of what we can say about chemistry of the ascorbic acid.* For example, its acidity comes mainly from a tendency to dissociate a particular proton 1, while only one of four isomers, differing by configurations of atoms 2 (oxygen) and 3 (hydrogen) with respect to their nearest neighbor carbon atoms, acts as vitamin C.

Still another example. In the infrared range, neither the hydrogen atom nor the chlorine atom has any electromagnetic wave absorption. However, the hydrogen chloride diatomic molecule that is formed by these two atoms *has* a very rich absorption spectrum with a quasiregular and mysterious structure, shown in Fig. 6.10 on p. 340. If the theory given in the previous chapters is correct, then it should explain every detail of such a strange spectrum. We hope also we will *understand* why such a spectrum may appear.

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00014-8

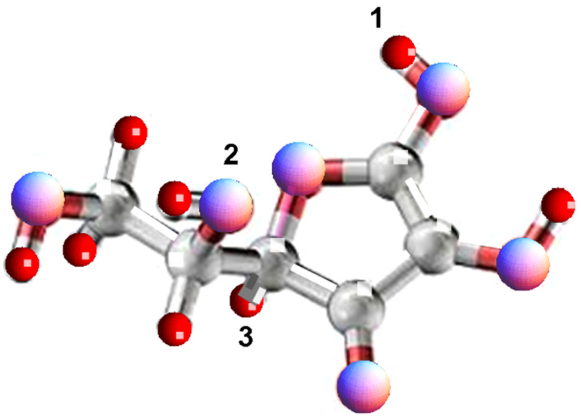


Fig. 6.1. A three-dimensional model of a molecule. The ascorbic acid molecule exhibits a spatial shape represented naively as some shiny atoms (balls – the largest are carbon atoms, the medium ones are oxygen atoms, and the smallest are hydrogen atoms) connected by some shiny rods (chemical bonds). We use such molecular models in chemistry (with great success!) as if real molecules resembled the macroscopic objects which we know from our everyday life.

What is it all about?

Nuclei are thousands of times heavier than electrons. As an example let us take the hydrogen atom. From the conservation of momentum law, it follows that the proton moves 1840 times slower than the electron. In a polyatomic system, while a nucleus moves a little, an electron travels many times through the molecule. It seems that a lot can be simplified when assuming electronic motion in a field created by immobile nuclei. This concept is behind what is called *adiabatic approximation*, in which the motions of the electrons and the nuclei are separated.[1] *Only after this approximation is introduced can we obtain the fundamental concept of chemistry: the molecular structure in three-dimensional space.*

The separation of the electronic and nuclear motions will be demonstrated in detail by taking the example of a diatomic molecule.

Why is it important?

The separation of the electronic and nuclear motions represents a fundamental approximation of quantum chemistry. Without this, chemists would lose their *basic model of the molecule*: the three-dimensional structure with the nuclei occupying some positions in three-dimensional space, with chemical bonds, etc. This is why the present chapter occupies the *central* position on the TREE.

What is needed?

- Postulates of quantum mechanics (Chapter 1, necessary),
- separation of the center-of-mass motion (Appendix J, p. 691, necessary),
- rigid rotator (Chapter 4, necessary),
- harmonic and Morse oscillators (Chapter 4, necessary),
- conclusions from group theory (Appendix C, p. 605, advised),
- dipole moment (Appendix V2-G, p. V2-613, occasionally used).

Classical works

A fundamental approximation (called the Born–Oppenheimer approximation) has been introduced in a paper "*Zur Quantentheorie der Molekeln*" by Max Born and Julius Robert Oppenheimer in *Annalen der Physik*, 84(1927)457. The approximation follows from the fact that nuclei are much heavier than electrons. ★ The conical intersection problem was first recognized by three young and congenial Hungarians: Janos (later John) von Neumann and Jenó Pál (later Eugene) Wigner in the papers "*Über merkwürdige diskrete Eigenwerte*," in *Physikalische Zeitschrift,* 30(1929)465 and "*Über das Verhalten von Eigenwerten bei adiabatischen Prozessen*," also published in *Physikalische Zeitschrift*, 30(1929)467, and later in a paper by Edward Teller published in the *Journal of Chemical Physics* 41(1937)109. ★ Gerhard Herzberg was the greatest spectroscopist of the 20th century, author of the fundamental three-volume work "*Spectra of Diatomic Molecules*" (1939), "*Infrared and Raman Spectra of Polyatomic Molecules*" (1949), and "*Electronic Spectra of Polyatomic Molecules*" (1966). ★ The world's first computational papers using a rigorous approach to go beyond the Born–Oppenheimer approximation for molecules

---

[1] It does not mean that the electrons and the nuclei move independently. We obtain two coupled equations: one for the motion of the electrons in the field of the fixed nuclei, and the other for the motion of the nuclei in the potential averaged over the electronic positions.

were two articles by Włodzimierz Kołos and Lutosław Wolniewicz, the first in *Acta Physica Polonica*, 20(1961)129 entitled "*The Coupling between Electronic and Nuclear Motion and the Relativistic Effects in the Ground State of the H_2 Molecule*" and the second in *Physics Letters*, 2(1962)222 entitled "*A Complete Non-Relativistic Treatment of the H_2 Molecule*." ★ The discovery of the conical intersection and the funnel effect in photochemistry is attributed to Howard E. Zimmerman (*Journal of the American Chemical Society*, 88(1966)1566) and to Joseph Michl (*Journal of Molecular Photochemistry*, 243(1972)). Important contributions in this domain were also made by Lionel Salem and Christopher Longuet-Higgins.

John von Neumann (1903–1957), known as Jancsi (then Johnny), was the wunderkind of a top Hungarian banker (Jancsi showed off at receptions by reciting from memory all the phone numbers after reading a page of the phone book). He attended the same famous Lutheran High School in Budapest as Jenó Pál (who later used the name Eugene) Wigner. In 1926 von Neumann received his chemistry engineering diploma, and in the same year he completed his PhD in mathematics at the University of Budapest. He finally emigrated to the USA and founded the Princeton Advanced Study Institute. John von Neumann was a mathematical genius. He contributed to the mathematical foundations of quantum theory, computers, and game theory.

Von Neumann made a strange offer of a professorship at the Advanced Study Institute to Stefan Banach from the John Casimir University in Lvov. He handed him a cheque with a handwritten figure "1" and asked Banach to add as many zeros as he wanted. "*This is not enough money to persuade me to leave Poland*," answered Banach.

Edward Teller (1908–2004), American physicist of Hungarian origin, professor at the George Washington University, the University of Chicago, and the University of California. Teller left Hungary in 1926, received his PhD in 1930 at the University of Leipzig, and fled Nazi Germany in 1935. Teller was the project leader and the top brain behind the American hydrogen bomb project in Los Alamos, believing that this was the way to overthrow communism. The hydrogen bomb patent is owned by Edward Teller and Stanisław Ulam.

Eugene Paul Wigner (1902–1995), American chemist, physicist and mathematician of Hungarian origin, professor at Princeton University (USA). At the age of 11 Wigner, a primary schoolboy from Budapest, was in a sanatorium in Austria with suspected tuberculosis. Lying for hours on a desk chair reading books, he was seduced by the beauty of mathematics (fortunately, it turned out he did not have tuberculosis).

In 1915 Wigner entered the famous Lutheran High School in Budapest. Fulfilling the wish of his father, who dreamed of having a successor in managing the familial tannery, Wigner graduated from the Technical University in Budapest as a chemist. In 1925, at the Technical University in Berlin, he defended his PhD thesis on chemical kinetics "*Bildung und Zerfall von Molekülen*" under the supervision of Michael Polanyi, a pioneer in the study of chemical reactions. In 1926 Wigner left the tannery...

Accidentally he was advised by his colleague von Neumann to focus on group theory (where he achieved the most spectacular successes). Wigner was the first to understand the main features of the nuclear forces. In 1963 he won the Nobel Prize "*for his contributions to the theory of the atomic nucleus and elementary particles, particularly through the discovery and application of fundamental symmetry principles.*"

Christopher Longuet-Higgins (1923–2004), professor at the University of Sussex, Great Britain, began his scientific career as a theoretical chemist. His main achievements are connected with conical intersection, as well as with the introduction of permutational groups in the theoretical explanation of the spectra of flexible molecules.

Longuet-Higgins was elected as a member of the Royal Society of London for these contributions. He turned to artificial intelligence at the age of 40, and in 1967 he founded the Department of Machine Intelligence and Perception at the University of Edinburgh. Longuet-Higgins investigated machine perception of speech and music. His contribution to this field was recognized by the award of an Honorary Doctorate in Music by Sheffield University. Courtesy of Professor John D. Roberts.

6.1 Separation of the center-of-mass motion

Space-fixed coordinate system (SFCS)

Let us consider first a diatomic molecule with the nuclei labeled by a, b and n electrons (with charges $\mathcal{Z}_a$ and $\mathcal{Z}_b$). Let us choose a Cartesian coordinate system in our laboratory (called the space-fixed coordinate system [SFCS]) with the origin located at an arbitrarily chosen point and with arbitrary orientation of the axes.[2] The nuclei have the following positions: $\mathbf{R}_a = (X_a, Y_a, Z_a)$ and $\mathbf{R}_b = (X_b, Y_b, Z_b)$, while electron i has the position $\mathbf{r}'_i = (x'_i, y'_i, z'_i)$ (see Fig. 6.2).

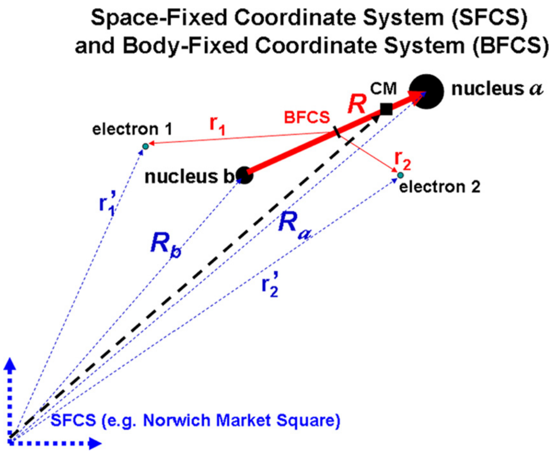


Fig. 6.2. The space-fixed coordinate system (SFCS) and the body-fixed coordinate system (BFCS) for a diatomic molecule.

We write the Hamiltonian for the system (Chapter 1),

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2M_a}\Delta_a - \frac{\hbar^2}{2M_b}\Delta_b - \sum_{i=1}^{n}\frac{\hbar^2}{2m}\Delta'_i + V, \tag{6.1}$$

where the first two terms stand for the kinetic energy operators of the nuclei, the third term corresponds to the kinetic energy of the electrons (m is the electron mass, all Laplacians are in the SFCS), and V denotes the Coulombic potential energy operator (interaction of all the particles: nuclei–nuclei, nuclei–electrons, electrons–electrons), i.e.,

$$V = \frac{\mathcal{Z}_a \mathcal{Z}_b e^2}{R} - \mathcal{Z}_a \sum_i \frac{e^2}{r_{ai}} - \mathcal{Z}_b \sum_i \frac{e^2}{r_{bi}} + \sum_{i<j} \frac{e^2}{r_{ij}}. \tag{6.2}$$

---

[2] For example, right in the center of the Norwich market square.

When we are not interested in collisions of our molecule with a wall or similar obstruction, we may consider a separation of the motion of the center of mass, then forget about the motion, and focus on the rest, i.e., on the *relative* motion of the particles.

New coordinates

The total mass of the molecule is $M = M_a + M_b + mn$. The components of the center-of-mass position vector are[3]

$$X = \frac{1}{M}\left(M_a X_a + M_b X_b + \sum_i m x_i'\right),$$
$$Y = \frac{1}{M}\left(M_a Y_a + M_b Y_b + \sum_i m y_i'\right),$$
$$Z = \frac{1}{M}\left(M_a Z_a + M_b Z_b + \sum_i m z_i'\right).$$

Now, we decide to abandon this SFCS. Instead of the old coordinates, we will choose a new set of $3n + 6$ coordinates (see Appendix J on p. 691, choice II):

- three center-of-mass coordinates X, Y, Z,
- three components of the vector $\mathbf{R} = \mathbf{R}_a - \mathbf{R}_b$ that points to nucleus a from nucleus b,
- $3n$ electronic coordinates $x_i = x_i' - \frac{1}{2}(X_a + X_b)$ and similarly for y_i and z_i, for $i = 1, 2, \ldots n$, which show the electron's position with respect to the *geometric center*[4] of the molecule.

Hamiltonian in the new coordinates

The new coordinates have to be introduced into the Hamiltonian. To this end, we need the second derivative operators in the old coordinates to be expressed by the new ones. First (similarly as in Appendix J), let us construct the *first* derivative operators, i.e.,

$$\frac{\partial}{\partial X_a} = \frac{\partial X}{\partial X_a}\frac{\partial}{\partial X} + \frac{\partial Y}{\partial X_a}\frac{\partial}{\partial Y} + \frac{\partial Z}{\partial X_a}\frac{\partial}{\partial Z} + \frac{\partial R_x}{\partial X_a}\frac{\partial}{\partial R_x} + \frac{\partial R_y}{\partial X_a}\frac{\partial}{\partial R_y} + \frac{\partial R_z}{\partial X_a}\frac{\partial}{\partial R_z} +$$

---

[3] Do not mix the coordinate Z with the nuclear charge $\mathcal{Z}$.

[4] If the origin were chosen in the center of mass instead of the geometric center, V would become mass-dependent (J. Hinze, A. Alijah, L. Wolniewicz, *Pol. J. Chem.*, 72(1998)1293), cf. also Appendix J, Example II. We want to avoid this.

$$+\sum_i \frac{\partial x_i}{\partial X_a}\frac{\partial}{\partial x_i}+\sum_i \frac{\partial y_i}{\partial X_a}\frac{\partial}{\partial y_i}+\sum_i \frac{\partial z_i}{\partial X_a}\frac{\partial}{\partial z_i}=$$
$$\frac{\partial X}{\partial X_a}\frac{\partial}{\partial X}++\frac{\partial R_x}{\partial X_a}\frac{\partial}{\partial R_x}+\sum_i \frac{\partial x_i}{\partial X_a}\frac{\partial}{\partial x_i}=\frac{M_a}{M}\frac{\partial}{\partial X}+\frac{\partial}{\partial R_x}-\frac{1}{2}\sum_i \frac{\partial}{\partial x_i},$$

and similarly for the coordinates Y_a and Z_a. For nucleus b the expression is a little bit different: $\frac{\partial}{\partial X_b}=\frac{M_b}{M}\frac{\partial}{\partial X}-\frac{\partial}{\partial R_x}-\frac{1}{2}\sum_i \frac{\partial}{\partial x_i}$.

For the first derivative operator with respect to the coordinates of the electron i we obtain

$$\frac{\partial}{\partial x_i'}=\frac{\partial X}{\partial x_i'}\frac{\partial}{\partial X}+\frac{\partial Y}{\partial x_i'}\frac{\partial}{\partial Y}+\frac{\partial Z}{\partial x_i'}\frac{\partial}{\partial Z}+\frac{\partial R_x}{\partial x_i'}\frac{\partial}{\partial R_x}+\frac{\partial R_y}{\partial x_i'}\frac{\partial}{\partial R_y}+\frac{\partial R_z}{\partial x_i'}\frac{\partial}{\partial R_z}+$$
$$+\sum_j \frac{\partial x_j}{\partial x_i'}\frac{\partial}{\partial x_j}+\sum_j \frac{\partial y_j}{\partial x_i'}\frac{\partial}{\partial y_j}+\sum_j \frac{\partial z_j}{\partial x_i'}\frac{\partial}{\partial z_j}=$$
$$\frac{\partial X}{\partial x_i'}\frac{\partial}{\partial X}+\frac{\partial x_i}{\partial x_i'}\frac{\partial}{\partial x_i}=\frac{m}{M}\frac{\partial}{\partial X}+\frac{\partial}{\partial x_i},$$

and similarly for y_i' and z_i'.

Now, let us create the second derivative operators, i.e.,

$$\frac{\partial^2}{\partial X_a^2}=\left(\frac{M_a}{M}\frac{\partial}{\partial X}+\frac{\partial}{\partial R_x}-\frac{1}{2}\sum_i \frac{\partial}{\partial x_i}\right)^2=\left(\frac{M_a}{M}\right)^2\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial R_x^2}+\frac{1}{4}\left(\sum_i \frac{\partial}{\partial x_i}\right)^2+$$
$$2\frac{M_a}{M}\frac{\partial}{\partial X}\frac{\partial}{\partial R_x}-\frac{\partial}{\partial R_x}\sum_i \frac{\partial}{\partial x_i}-\frac{M_a}{M}\frac{\partial}{\partial X}\sum_i \frac{\partial}{\partial x_i},$$
$$\frac{\partial^2}{\partial X_b^2}=\left(\frac{M_b}{M}\frac{\partial}{\partial X}-\frac{\partial}{\partial R_x}-\frac{1}{2}\sum_i \frac{\partial}{\partial x_i}\right)^2=\left(\frac{M_b}{M}\right)^2\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial R_x^2}+\frac{1}{4}\left(\sum_i \frac{\partial}{\partial x_i}\right)^2+$$
$$-2\frac{M_b}{M}\frac{\partial}{\partial X}\frac{\partial}{\partial R_x}+\frac{\partial}{\partial R_x}\sum_i \frac{\partial}{\partial x_i}-\frac{M_b}{M}\frac{\partial}{\partial X}\sum_i \frac{\partial}{\partial x_i},$$
$$\frac{\partial^2}{\partial \left(x_i'\right)^2}=\left(\frac{m}{M}\frac{\partial}{\partial X}+\frac{\partial}{\partial x_i}\right)^2=\left(\frac{m}{M}\right)^2\frac{\partial^2}{\partial X^2}+\frac{\partial^2}{\partial x_i^2}+2\frac{m}{M}\frac{\partial}{\partial X}\frac{\partial}{\partial x_i}.$$

After inserting all this into the Hamiltonian (6.1) we obtain the Hamiltonian expressed in the new coordinates[5]:

---

[5] The potential energy also has to be expressed using the new coordinates.

$$\hat{\mathcal{H}} = \left[-\frac{\hbar^2}{2M} \Delta_{XYZ} \right] + \hat{H}_0 + \hat{H}', \tag{6.3}$$

where the first term means the center-of-mass kinetic energy operator, $\hat{H}_0$ is the *electronic Hamiltonian* (*clamped nuclei Hamiltonian*)

$$\hat{H}_0 = -\sum_i \frac{\hbar^2}{2m} \Delta_i + V, \tag{6.4}$$

while $\Delta_i \equiv \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$ and

$$\hat{H}' = -\frac{\hbar^2}{2\mu} \Delta_{\mathbf{R}} + \hat{H}'' \tag{6.5}$$

with $\Delta_{\mathbf{R}} \equiv \frac{\partial^2}{\partial R_x^2} + \frac{\partial^2}{\partial R_y^2} + \frac{\partial^2}{\partial R_z^2}$, where

$$\hat{H}'' = \left[-\frac{\hbar^2}{8\mu} \left(\sum_i \nabla_i \right)^2 + \frac{\hbar^2}{2} \left(\frac{1}{M_a} - \frac{1}{M_b} \right) \nabla_{\mathbf{R}} \sum_i \nabla_i \right],$$

and μ denotes the reduced mass of the two nuclei ($\mu^{-1} = M_a^{-1} + M_b^{-1}$).

The $\hat{H}_0$ term does not contain the kinetic energy operator of the nuclei, but all the other terms (this is why it is called the electronic Hamiltonian): the first term stands for the kinetic energy operator of the electrons, and V means the potential energy corresponding to the Coulombic interaction of all particles. The first term in the operator $\hat{H}'$, i.e., $-\frac{\hbar^2}{2\mu}\Delta_{\mathbf{R}}$, denotes the kinetic energy operator of the nuclei,[6] while the operator $\hat{H}''$ couples the motions of the nuclei and electrons.[7]

---

[6] What moves is a particle of reduced mass μ and with coordinates R_x, R_y, R_z. This means that the particle has the position of nucleus a, whereas nucleus b is at the origin. Therefore, this term accounts for the vibrations of the molecule (changes in length of $\mathbf{R}$), as well as its rotations (changes in orientation of $\mathbf{R}$).

[7] The first of these two terms contains the reduced mass of the two nuclei, where ∇_i denotes the nabla operator for *electron* i, $\nabla_i \equiv \mathbf{i}\frac{\partial}{\partial x_i} + \mathbf{j}\frac{\partial}{\partial y_i} + \mathbf{k}\frac{\partial}{\partial z_i}$ with $\mathbf{i}, \mathbf{j}, \mathbf{k}$ – the unit vectors along the axes x, y, z. The second term is nonzero only for the heteronuclear case and contains the mixed product of nablas: $\nabla_R \nabla_i$, with

After separation of the center-of-mass motion

After separation of the center-of-mass motion (the first term in Eq. (6.3) is gone, see Appendix J on p. 691) we obtain the eigenvalue problem of the Hamiltonian

$$\hat{H} = \hat{H}_0 + \hat{H}'. \tag{6.6}$$

This is an exact result, fully equivalent to the Schrödinger equation.

6.2 Exact (nonadiabatic) theory

The total wave function that describes both electrons and nuclei can be proposed in the following form[8] ($\mathcal{N} = \infty$):

$$\Psi(\mathbf{r}, \mathbf{R}) = \sum_{k}^{\mathcal{N}} \psi_k(\mathbf{r}; R)\, f_k(\mathbf{R}), \tag{6.7}$$

$\nabla_{\mathbf{R}} = \mathbf{i}\frac{\partial}{\partial R_x} + \mathbf{j}\frac{\partial}{\partial R_y} + \mathbf{k}\frac{\partial}{\partial R_z}$ and R_x, R_y, R_z as the components of the vector $\mathbf{R}$.

[8] Where does such a form of the wave function come from?

If the problem were solved exactly, then the solution of the Schrödinger equation could be sought, e.g., by using the Ritz method (p. 271). Then we have to decide what kind of basis set to use. We could use two auxiliary complete basis sets: one that depends on the electronic coordinates $\{\bar{\psi}_k(\mathbf{r})\}$ or one that depends on the nuclear coordinates $\{\bar{\phi}_l(\mathbf{R})\}$. The complete basis set for the Hilbert space of our system could be constructed as a Cartesian product $\{\bar{\psi}_k(\mathbf{r})\} \times \{\bar{\phi}_l(\mathbf{R})\}$, i.e., all possible product-like functions $\bar{\psi}_k(\mathbf{r})\,\bar{\phi}_l(\mathbf{R})$. Thus, the wave function could be expanded in a series

$$\Psi(\mathbf{r}, \mathbf{R}) = \sum_{kl} c_{kl}\bar{\psi}_k(\mathbf{r})\,\bar{\phi}_l(\mathbf{R}) = \sum_{k}^{\mathcal{N}} \bar{\psi}_k(\mathbf{r}) \left[\sum_{l} c_{kl}\bar{\phi}_l(\mathbf{R})\right] =$$

$$= \sum_{k}^{\mathcal{N}} \bar{\psi}_k(\mathbf{r})\, f_k(\mathbf{R}),$$

where $f_k(\mathbf{R}) = \sum_l c_{kl}\bar{\phi}_l(\mathbf{R})$ stands for a to-be-sought coefficient depending on $\mathbf{R}$ (rovibrational function). If we had to do with complete sets, then both $\bar{\psi}_k$ and f_k should not depend on anything else, since a sufficiently long expansion of the terms $\bar{\psi}_k(\mathbf{r})\,\bar{\phi}_l(\mathbf{R})$ would be suitable to describe all possible distributions of the electrons and the nuclei.

However, we are unable to manage the complete sets; instead, we are only able to take a few terms in this expansion. We would like them to describe the molecule reasonably well, and at the same time to have only a few, or to be exact, only one such term. If so, it would be reasonable to introduce a *parametric dependence* of the function $\bar{\psi}_k(\mathbf{r})$ on the position of the nuclei, which in the present case of a diatomic molecule means the internuclear distance. *This is equivalent to telling someone how the electrons behave when the internuclear distances are such and such, and how they behave when the distances are changed.*

where $\psi_k(\mathbf{r}; R)$ are the eigenfunctions of $\hat{H}_0$

$$\hat{H}_0(R)\,\psi_k(\mathbf{r}; R) = E_k^0(R)\psi_k(\mathbf{r}; R) \tag{6.8}$$

that depend parametrically[9] on the internuclear distance R, and $f_k(\mathbf{R})$ are yet unknown rovibrational functions (describing the rotations and vibrations of the molecule).

Averaging over electronic coordinates

First, let us write down the Schrödinger equation with the Hamiltonian (6.6) and the wave function as in (6.7):

$$(\hat{H}_0 + \hat{H}')\sum_l^{\mathcal{N}} \psi_l(\mathbf{r}; R)\, f_l(\mathbf{R}) = E\sum_l^{\mathcal{N}} \psi_l(\mathbf{r}; R)\, f_l(\mathbf{R})\,. \tag{6.9}$$

Let us multiply both sides by $\psi_k^*(\mathbf{r}; R)$ and then integrate over the *electronic* coordinates $\mathbf{r}$ (which will be stressed by the subscript "e"):

$$\sum_l^{\mathcal{N}} \left\langle \psi_k|(\hat{H}_0 + \hat{H}')[\psi_l f_l]\right\rangle_e = E\sum_l^{\mathcal{N}} \langle\psi_k|\psi_l\rangle_e\, f_l. \tag{6.10}$$

On the right-hand side of (6.10) we profit from the orthonormalization condition $\langle\psi_k|\psi_l\rangle_e = \delta_{kl}$, on the left-hand side we recall that ψ_k is an eigenfunction of $\hat{H}_0$,

$$E_k^0 f_k + \sum_l^{\mathcal{N}} \left\langle \psi_k|\hat{H}'[\psi_l f_l]\right\rangle_e = E f_k. \tag{6.11}$$

Now, let us focus on the expression $\hat{H}'(\psi_l f_l) = -\frac{\hbar^2}{2\mu}\Delta_{\mathbf{R}}(\psi_l f_l) + \hat{H}''(\psi_l f_l)$, which we have in the integrand in Eq. (6.11). Let us concentrate on the first of these terms. We have

$$\begin{aligned}
&-\frac{\hbar^2}{2\mu}\Delta_{\mathbf{R}}(\psi_l f_l) = -\frac{\hbar^2}{2\mu}\nabla_{\mathbf{R}}\nabla_{\mathbf{R}}(\psi_l f_l) = -\frac{\hbar^2}{2\mu}\nabla_{\mathbf{R}}[\psi_l\nabla_{\mathbf{R}} f_l + (\nabla_{\mathbf{R}}\psi_l) f_l]\\
&= -\frac{\hbar^2}{2\mu}[\nabla_{\mathbf{R}}\psi_l\nabla_{\mathbf{R}} f_l + \psi_l\Delta_{\mathbf{R}} f_l + (\Delta_{\mathbf{R}}\psi_l) f_l + \nabla_{\mathbf{R}}\psi_l\nabla_{\mathbf{R}} f_l]\\
&= -\frac{\hbar^2}{2\mu}[2\,(\nabla_{\mathbf{R}}\psi_l)\,(\nabla_{\mathbf{R}} f_l) + \psi_l\Delta_{\mathbf{R}} f_l + (\Delta_{\mathbf{R}}\psi_l) f_l].
\end{aligned} \tag{6.12}$$

---

[9] For each value of R we have a different formula for ψ_k.

After inserting the result into $\left\langle \psi_k | \hat{H}'(\psi_l f_l) \right\rangle_e$ and recalling Eq. (6.5), we have

$$\begin{aligned}
&\left\langle \psi_k | \hat{H}'[\psi_l f_l] \right\rangle_e = 2(-\frac{\hbar^2}{2\mu})\, \langle \psi_k | \nabla_{\mathbf{R}} \psi_l \rangle_e\, \nabla_{\mathbf{R}} f_l + \langle \psi_k | \psi_l \rangle_e\, (-\frac{\hbar^2}{2\mu}) \Delta_{\mathbf{R}} f_l + \\
&+ \left\langle \psi_k | (-\frac{\hbar^2}{2\mu}) \Delta_{\mathbf{R}} \psi_l \right\rangle_e f_l + \left\langle \psi_k | \hat{H}'' \psi_l \right\rangle_e f_l = \\
&(1 - \delta_{kl})\, (-\frac{\hbar^2}{\mu})\, \langle \psi_k | \nabla_{\mathbf{R}} \psi_l \rangle_e\, \nabla_{\mathbf{R}} f_l - \delta_{kl} \frac{\hbar^2}{2\mu} \Delta_{\mathbf{R}} f_l + H'_{kl} f_l,
\end{aligned} \tag{6.13}$$

with

$$H'_{kl} \equiv \left\langle \psi_k | \hat{H}' \psi_l \right\rangle_e .$$

We obtain the following form of (6.11):

$$E_k^0 f_k + \sum_l^{\mathcal{N}} \left[(1 - \delta_{kl})\, (-\frac{\hbar^2}{\mu})\, \langle \psi_k | \nabla_{\mathbf{R}} \psi_l \rangle_e\, \nabla_{\mathbf{R}} f_l - \delta_{kl} \frac{\hbar^2}{2\mu} \Delta_{\mathbf{R}} f_l + H'_{kl} f_l \right] = E f_k$$

(we have profited from the equality $\langle \psi_l | \nabla_{\mathbf{R}} \psi_l \rangle_e = 0$, which follows from the differentiation of the normalization condition[10] for the function ψ_l).

Nonadiabatic nuclear motion

Grouping all the terms with f_l on the left-hand side we obtain a set of $\mathcal{N}$ equations,

$$\left[-\frac{\hbar^2}{2\mu} \Delta_{\mathbf{R}} + E_k^0 (R) + H'_{kk} (R) - E \right] f_k = - \sum_{l (\neq k)}^{\mathcal{N}} \Theta_{kl}\ f_l, \tag{6.14}$$

for $k = 1, 2, \ldots \mathcal{N}$, with the nonadiabatic coupling operators

$$\Theta_{kl} = -\frac{\hbar^2}{\mu}\, \langle \psi_k | \nabla_{\mathbf{R}} \psi_l \rangle_e\, \nabla_{\mathbf{R}} + H'_{kl}. \tag{6.15}$$

---

[10] We assume that the phase of the wave function $\psi_k(\mathbf{r}; R)$ does not depend on R, i.e., $\psi_k(\mathbf{r}; R) = \bar{\psi}_k(\mathbf{r}; R) \exp(i\phi)$, where $\bar{\psi}_k$ is a real function and $\phi \neq \phi(R)$. This immediately gives $\langle \psi_k | \nabla_{\mathbf{R}} \psi_k \rangle_e = \left\langle \bar{\psi}_k | \nabla_{\mathbf{R}} \bar{\psi}_k \right\rangle_e$, which is zero from differentiating the normalization condition. Indeed, we have the normalization condition $\int \psi_k^2 \mathrm{d}\tau_e = 1$. Hence, $\nabla_{\mathbf{R}} \int \psi_k^2 \mathrm{d}\tau_e = 0$, or $2 \int \psi_k \nabla_{\mathbf{R}} \psi_k \mathrm{d}\tau_e = 0$.
Without this approximation we would surely be in trouble.

Note that the operator H'_{lk} depends on the length of the vector $\mathbf{R}$, but not on its direction.[11]

Eq. (6.14) is equivalent to the Schrödinger equation.

Eqs. (6.14) and (6.15) have been derived under the assumption that ψ_k of Eq. (6.7) satisfy Eq. (6.8). If instead of $\psi_k(\mathbf{r}; R)$ we use a (generally nonorthogonal) complete set $\left\{\bar{\psi}_k(\mathbf{r}; R)\right\}$ in (6.7), Eqs. (6.14) and (6.15) would change to

$$\left[-\frac{\hbar^2}{2\mu}\Delta_{\mathbf{R}} + \bar{E}_k(R) + H'_{kk}(R) - E\right] f_k = -\sum_{l(\neq k)}^{\mathcal{N}} \Theta_{kl}\, f_l, \tag{6.16}$$

for $k = 1, 2, \ldots \mathcal{N}$, with the nonadiabatic coupling operators

$$\Theta_{kl} = -\frac{\hbar^2}{\mu}\langle\bar{\psi}_k|\nabla_{\mathbf{R}}\bar{\psi}_l\rangle_e \nabla_{\mathbf{R}} + H'_{kl} + \langle\bar{\psi}_k|\bar{\psi}_l\rangle_e\left(-\frac{\hbar^2}{2\mu}\Delta_{\mathbf{R}}\right) \tag{6.17}$$

and $\bar{E}_k(R) \equiv \left\langle\bar{\psi}_k|\hat{H}_0\bar{\psi}_k\right\rangle_e$. Functions $\bar{\psi}_k(\mathbf{r}; R)$ may be chosen as the wave functions having some chemical significance.

6.3 Adiabatic approximation

If the curves $E_k^0(R)$ for different k are well separated in the energy scale, we may expect that the coupling between them is small, and therefore all Θ_{lk} for $k \neq l$ may be set equal to zero. This is called the adiabatic approximation. In this approximation we obtain the following from (6.14):

$$\left[-\frac{\hbar^2}{2\mu}\Delta_{\mathbf{R}} + E_k^0(R) + H'_{kk}(R)\right] f_k(\mathbf{R}) = E f_k(\mathbf{R}), \tag{6.18}$$

where the diagonal correction $H'_{kk}(R)$ is usually very small when compared to $E_k^0(R)$.

In the adiabatic approximation the wave function is approximated by a product:

$$\Psi \approx \psi_k(\mathbf{r}; R) f_k(\mathbf{R}). \tag{6.19}$$

---

[11] This follows from the fact that we have in $\hat{H}'$ (see Eq. (6.5)) the products of nablas, i.e., scalar products. The scalar products do not change upon rotation, because both vectors involved rotate in the same way and the angle between them does not change.

The function $f_k(\mathbf{R})$ depends explicitly not only on R, but also on the direction of vector $\mathbf{R}$, and therefore will describe future vibrations of the molecule (changes of R) as well as its rotations (changes in the direction of $\mathbf{R}$).

A simple analogy

Let us stop for a while to catch the sense of the adiabatic approximation.

To some extent the situation resembles an attempt to describe a tourist (an electron) and the Alps (nuclei). Not only the tourist moves, but also the Alps, as has been quite convincingly proved by geologists.[12] The probability of encountering the tourist may be described by a "wave function" computed for a fixed position of the mountains (shown by a map bought in a shop). This is a very good approximation, because when the tourist wanders over hundreds of miles, the Alps move a tiny distance, so that the map seems to be perfect all the time. On the other hand, the probability of having the Alps in a given configuration is described by the geologists' "wave function" f, saying for example, the probability that the distance between the Matterhorn and the Jungfrau is equal to R. When the tourist revisits the Alps after a period of time (say, a few million years), the mountains will be changed (the new map bought in the shop will reflect this fact). The probability of finding the tourist may again be computed from the new wave function *valid for the new configuration of the mountains (a parametric dependence)*. Therefore, the probability of finding the tourist in the spot indicated by the vector $\mathbf{r}$ at a given configuration of the mountains $\mathbf{R}$ can be *approximated by a product*[13] of the probability of finding the mountains at this configuration $|f_k(\mathbf{R})|^2 d\mathbf{R}$ and the probability $|\psi_k(\mathbf{r};\mathbf{R})|^2 d\mathbf{r}$ of finding the tourist in the position shown by the vector $\mathbf{r}$, when the mountains have this particular configuration $\mathbf{R}$. In the case of our molecule this means the adiabatic approximation (a product-like form) Eq. (6.19).

This parallel fails in one important point: the Alps do not move in the potential created by tourists; the dominant geological processes are tourist-independent. As we will soon see, nuclear motion is dictated by the electrons through the potential *averaged over the electronic motion*.

---

[12] The continental plates collide like billiard balls in a kind of quasiperiodic oscillation. During the current oscillation, the Indian tectonic plate, which moves at a record speed of about 20 cm a year, hits the Eurasian plate. This is why the Himalayan mountains are so beautiful. The collision is continuing and the Himalayas will be even more beautiful. Europe is hit from the South by a few plates moving at about 4 cm a year, and this is why we have much lower Alps. While visiting the Atlantic coast of Maine (USA), I noticed that the color of the rocks was very similar to those I remembered from Brittany (France). That was it! Once upon a time the two coasts made a common continent. Later we had to rediscover America. The Wegener theory of continental plate tectonics, when created in 1911, was viewed as absurd, although the mountain *ranges* suggested that some plates were colliding.

[13] This is an approximation, because in the nonadiabatic, i.e., fully correct, approach the total wave function is a superposition of many such products, corresponding to various electronic and rovibrational wave functions.

6.4 Born–Oppenheimer approximation

In the adiabatic approximation, $H'_{kk} = \int \psi_k^* H' \psi_k \mathrm{d}\tau_e$ represents a small correction to $E_k^0(R)$. *Neglecting this correction, which is nothing but assuming the masses of the nuclei equal to infinity, results in the all important Born–Oppenheimer approximation*

$$H'_{kk} = 0. \tag{6.20}$$

Note that in the Born–Oppenheimer approximation the potential energy for the motion of the nuclei $E_k^0(R)$ is independent of the masses of the nuclei, whereas in the adiabatic approximation the potential energy $E_k^0(R) + H'_{kk}(R)$ depends on the mass.

Julius Robert Oppenheimer (1904–1967), American physicist, professor at the University of California in Berkeley, the California Institute of Technology in Pasadena, and the Institute for Advanced Study in Princeton. In 1943–1945 Oppenheimer headed the Manhattan Project (development of the atomic bomb).
From John Slater's autobiography:
"… *Robert Oppenheimer was a very brilliant physics undergraduate at Harvard during the 1920s, the period when I was there on the faculty, and we all recognized that he was a person of very unusual attainments. Rather than going on for his graduate work at Harvard, he went to Germany, and worked with Born, developing what has been known as the Born–Oppenheimer approximation.*"

...and a "superiority of theory" over experiment

In experiments every chemist finds his molecule confined close to a minimum of the electronic energy hypersurface (most often of the ground state). A powerful theory might be able to predict the results of experiments even for the nuclear configurations that are far away from those accessible for current experiments. This is the case in quantum mechanics, which is able to describe in detail what would happen to the electronic structure[14] if the nuclear configuration were very strange, e.g., the internuclear distances were close to zero or even are exactly zero. Within the Born–Oppenheimer approximation the theoretician is free to put the nuclei wherever he/she wishes to do so. This means we are able to discuss and then just to test "what would be

---

[14] We just do not have any reason to doubt about it.

if," even if this "if" were crazy. For example, some small internuclear distances are achievable at extremely large pressures. At such pressures some additional difficult experiments have to be performed to tell us about the structure and processes. A theoretician just sets the small distances and makes a computer run. This is really exceptional: we may set some conditions which are out of reach of experiments, sometimes very expensive ones, and we are able to tell with confidence and at low cost what will be.

6.5 Vibrations of a rotating molecule

Our next step will be an attempt to separate rotations and oscillations within the adiabatic approximation. To this end the function $f_k(\mathbf{R}) = f_k(R, \theta, \phi)$ will be proposed as a *product* of a function Y which will account for rotations (depending on θ, ϕ) and a certain function $\frac{\chi_k(R)}{R}$ describing the oscillations, i.e., dependent on R,

$$f_k(\mathbf{R}) = Y(\theta, \phi)\frac{\chi_k(R)}{R}. \tag{6.21}$$

No additional approximation is introduced in this way. We say only that the isolated molecule vibrates independently of whether it is oriented towards the Capricorn or Taurus Constellations ("space is isotropic").[15] The function $\chi_k(R)$ is yet unknown, and we are going to search it; therefore, dividing by R in (6.21) is meaningless.[16]

Now, we will try to separate the variables θ, ϕ from the variable R in Eq. (6.18), i.e., to obtain two separate equations for them. First, let us define the quantity

$$U_k(R) = E_k^0(R) + H'_{kk}(R). \tag{6.22}$$

After inserting the Laplacian (in spherical coordinates, see Appendix H on p. 683) and the product (6.21) into (6.18) we obtain the following series of transformations:

$$\left[-\frac{\hbar^2}{2\mu}\left(\frac{1}{R^2}\frac{\partial}{\partial R}R^2\frac{\partial}{\partial R} + \frac{1}{R^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{R^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right) + U_k\left(R\right)\right]Y\frac{\chi_k}{R} = EY\frac{\chi_k}{R},$$

$$-\frac{\hbar^2}{2\mu}\left(\frac{Y}{R}\frac{\partial^2\chi_k}{\partial R^2} + \frac{\chi_k}{R}\frac{1}{R^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta} + \frac{\chi_k}{R}\frac{1}{R^2\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right) + YU_k\left(R\right)\frac{\chi_k}{R} = EY\frac{\chi_k}{R},$$

$$-\frac{\hbar^2}{2\mu}\left(\frac{1}{\chi_k}\frac{\partial^2\chi_k}{\partial R^2} + \frac{1}{Y}\left(\frac{1}{R^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta} + \frac{1}{R^2\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right)\right) + U_k\left(R\right) = E,$$

---

[15] It is an assumption about "the space," which is assumed not to be changed by the presence of Capricorn, Taurus, or another constellation.

[16] In the case of polyatomics the function $f_k(\mathbf{R})$ may be more complicated, because some *vibrations* (e.g., a rotation of the CH_3 group) may contribute to the total angular momentum, which has to be conserved (this is related to space isotropy, cf. p. 79).

$$-\left(\frac{R^2}{\chi_k}\frac{\partial^2\chi_k}{\partial R^2}\right)+\frac{2\mu}{\hbar^2}U_k(R)\,R^2-\frac{2\mu}{\hbar^2}ER^2=\frac{1}{Y}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right).$$

Look! The left-hand side depends on R only, and the right-hand side only on θ and ϕ. Both sides equal each other *independently of the values of the variables*. This can only happen if each side is equal to a *constant* (λ), the same for each. Therefore, we have

$$-\left(\frac{R^2}{\chi_k}\frac{\partial^2\chi_k}{\partial R^2}\right)+\frac{2\mu}{\hbar^2}U_k(R)\,R^2-\frac{2\mu}{\hbar^2}ER^2=\lambda, \tag{6.23}$$

$$\frac{1}{Y}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial Y}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2 Y}{\partial\phi^2}\right)=\lambda. \tag{6.24}$$

Now, we are amazed to see that (6.24) *is identical* (cf. p. 229) *to that which appeared as a result of the transformation of the Schrödinger equation for a rigid rotator Y* denoting the corresponding wave function. As we know from p. 230 this equation has a solution only if $\lambda=-J(J+1)$, where $J=0,1,2,\ldots$. *Since Y stands for the rigid rotator wave function, we now concentrate exclusively on the function χ_k*, which describes vibrations (changes in the length of $\mathbf{R}$).

After inserting the permitted values of λ into (6.23) we get

$$-\frac{\hbar^2}{2\mu}\left(\frac{\partial^2\chi_k}{\partial R^2}\right)+U_k(R)\,\chi_k-E\chi_k=-\frac{\hbar^2}{2\mu R^2}J(J+1)\chi_k.$$

Let us write this equation in the form of the eigenvalue problem for the unidimensional motion of a particle (we change the partial into the regular derivative) of mass μ, i.e.,

$$\left(-\frac{\hbar^2}{2\mu}\frac{d^2}{dR^2}+V_{kJ}\right)\chi_{kvJ}(R)=E_{kvJ}\chi_{kvJ}(R) \tag{6.25}$$

with potential energy (let us stress that $R>0$)

$$V_{kJ}(R)=U_k(R)+J(J+1)\frac{\hbar^2}{2\mu R^2}, \tag{6.26}$$

which takes the centrifugal force effect on the vibrational motion into account. The solution χ_k, as well as the total energy E, has been labeled by two additional indices: the rotational quantum number J (because the potential depends on the centrifugal effect now) and the numbering of the solutions $v = 0, 1, 2, \ldots$.

The solutions of Eq. (6.25) describe the vibrations of the rotating nuclei. The function $V_{kJ} = E_k^0(R) + H'_{kk}(R) + J(J+1)\hbar^2/(2\mu R^2)$ plays the role of the potential energy curve for the motion of the nuclei.

Therefore, also

the very notion of the *potential energy curve for the motion of the nuclei* appears only after the adiabatic (the product-like wave function, and H'_{kk} preserved) or the Born–Oppenheimer (the product-like wave function, but H'_{kk} removed) approximation is applied. Only in the Born–Oppenheimer approximation the potential energy V_{kJ} is mass-independent, e.g., the same for isotopomers H_2, HD, and D_2.

If $V_{kJ}(R)$ were a parabola (as it is for the harmonic oscillator), the system would never acquire the energy corresponding to the bottom of the parabola, because the harmonic oscillator energy levels (cf. p. 217) correspond to *higher* energy. The same pertains to V_{kJ} of a more complex shape.

6.5.1 One more analogy

The fact that the electronic energy $E_k^0(R)$ plays the role of the potential energy for vibrations represents not only the result of rather complex derivations, but is also natural and understandable. The nuclei keep together thanks to the electronic "glue" (we will come back to this in Chapter 8). Let us imagine two metallic balls (nuclei) in a block of transparent glue (electronic cloud, Fig. 6.3).

If we were interested in the motion of the *balls only*, we would have to take their potential energy as well as the kinetic energy into account. The potential energy would depend on the distance R between the balls, in the same way the glue's elastic energy depends on the stretching or squeezing of the glue to produce distance between the balls equal to R. Thus, the potential energy for the motion of the balls (nuclei) has to be the potential energy of the glue (electronic energy).[17]

---

[17] The adiabatic approximation is of more general importance than the separation of the electronic and nuclear motions. Its essence pertains to the problem of two coexisting time scales in some phenomena: fast and slow scales. The examples below indicate that we have to do with an important and general philosophical approach:

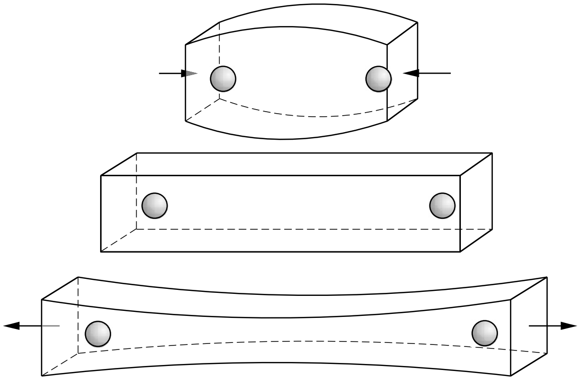


Fig. 6.3. Two metallic balls in a block of glue. How they will vibrate? This will be dictated by the elastic properties of the glue.

This situation corresponds to a nonrotating system. If we admit rotation, we would have to take the effect of centrifugal force on the potential energy of the gum into account. This effect is analogous to the second term in Eq. (6.26) for $V_{kJ}(R)$.

6.5.2 What vibrates, what rotates?

One may say that, as a result of averaging over electron coordinates, the electrons disappeared from the theory. The only effect of their presence are numbers: the potential energy term $U_k(R)$ of Eq. (6.22). Eq. (6.25) says that the vibrating and rotating objects are bare nuclei, which seems strange because they certainly move somehow with electrons. Our intuition says that what should vibrate and rotate are atoms, rather than nuclei. In our example with the glue it is evident that the iron balls should be a bit heavier since they pull the glue with them.

Where is this effect hidden? It has to be a part of the nonadiabatic effect, and it can be taken into account within the nonadiabatic procedure described on p. 315. It looks quite strange. Such an obvious effect[18] hidden in a theory that is hardly used in computational practice, because of its complexity?

- In Chapter V2-6 on chemical reactions, we will consider slow motion along a single coordinate and fast motions along other coordinates (in the configurational space of the nuclei). "Vibrationally adiabatic" approximation will also be introduced, and the slow motion will proceed in the potential energy averaged over fast motions and calculated at each fixed value of the slow coordinate.
- Similar reasoning was behind vibrational analysis in systems with hydrogen bonds (Y. Marechal, A. Witkowski, *Theor. Chim. Acta*, 9(1967)116). The authors selected a slow intermolecular motion proceeding in the potential energy averaged over fast intramolecular motions.

[18] The effect is certainly small, because the mass of the electrons that make the difference (move with the nucleus) is about 1836 times smaller than the mass of the nucleus alone.

If the excited electronic states are well separated from the ground electronic state $k = 0$, it turns out one may catch a good part of this effect for the ground state by using perturbation theory (see Chapter 5). It is possible to construct[19] a set of more and more advanced approximations for calculating the rovibrational levels. All of them stem from the following equation for the vibrational motion of the nuclei, a generalization of Eq. (6.25):

$$\left[-\frac{1}{R^2}\frac{\mathrm{d}}{\mathrm{d}R}\frac{R^2}{2\mu_{||}(R)}\frac{\mathrm{d}}{\mathrm{d}R}+W_{0J}(R)\right]\chi_{0vJ}(R)=E_{0vJ}\chi_{0vJ}(R), \tag{6.27}$$

where the operator on the left-hand side corresponds to the kinetic energy of vibration given in Eq. (6.25), but this time instead of the constant reduced mass μ of the nuclei we have the mass denoted as $\mu_{||}(R)$, which is R-dependent. The potential energy

$$W_{0J}(R)=E_0^0(R)+H'_{00}(R)+\frac{J(J+1)}{2\mu_\perp R^2}+\delta\mathcal{E}_{na}(R) \tag{6.28}$$

also resembles the potential energy of Eq. (6.26), but the reduced mass of the nuclei μ in the centrifugal energy, Eq. (6.26), is replaced now by a function of R denoted by $\mu_\perp(R)$. Visibly the nuclei are "dressed" by electrons, and this dressing is not only R-dependent, which is understandable, but also depends on what the nuclei are doing (vibration[20] or rotation)! There is also a nonadiabatic increment $\delta\mathcal{E}_{na}(R)$, which effectively takes into account the presence of higher states. We do not give here the formulae for $\delta\mathcal{E}_{na}(R)$, $\mu_{||}$, and $\mu_\perp$ (they have all been derived by Pachucki and Komasa[21]).

Therefore, the following sequence of approximations can be designed:

- the Born–Oppenheimer approximation: $\mu_{||}=\mu_\perp=\mu$; $W_{0J}(R)=E_0^0(R)+\frac{J(J+1)}{2\mu R^2}$;
- the adiabatic approximation: $\mu_{||}=\mu_\perp=\mu$; $W_{0J}(R)=E_0^0(R)+H'_{00}(R)+\frac{J(J+1)}{2\mu R^2}$;
- the effective nonadiabatic approximation: $\mu_{||},\mu_\perp$ taken as R-dependent; $W_{0J}(R)=E_0^0(R)+H'_{00}(R)+\frac{J(J+1)}{2\mu R^2}+\delta\mathcal{E}_{na}(R)$.

It is interesting to see what kind of object vibrates and rotates in the hydrogen molecule. As one can see from Fig. 6.4, $\mu_{||}(R=\infty)=\mu_\perp(R=\infty)=M_p+m$ while $\mu_{||}(R=0)=\mu_\perp(R=0)=M_p$, where M_p stands for the proton mass and m is the electron mass. Thus, for large R the hydrogen *atoms* vibrate and rotate, while for very small R only bare *protons*. For finite nonzero values of R the rotation-related effective atomic mass $m_\perp(R)$ changes monotonically, while the vibration-related effective atomic mass $m_{||}(R)$ undergoes quite peculiar changes, exhibiting a maximum mass at about 4 a.u. (a bit larger (!) than M_p+m) and addi-

---

[19] K. Pachucki, J. Komasa, *J. Chem. Phys.*, 129(2008)34102.

[20] The R-dependent μ has been introduced first by R.M. Herman and A. Asgharian, *J. Chem. Phys.*, 45(1966)2433.

[21] Eq. (6.28) may be treated as the most general definition of the potential energy curve for the motion of the nuclei.

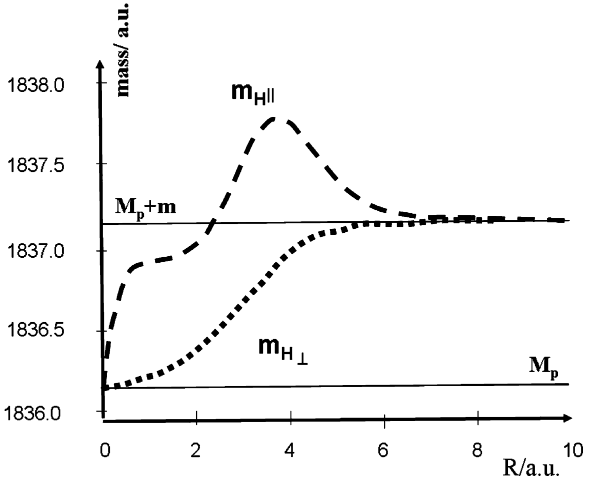


Fig. 6.4. What vibrates and what rotates in the hydrogen molecule? The variable reduced masses $\mu_{||}$ and $\mu_{\perp}$ correspond to the masses of the objects in the hydrogen molecule that vibrate ($m_{H||}$) and rotate ($m_{H\perp}$). We conclude that at large distances move atoms, at very short move the bare nuclei.

tionally an impressive plateau of about $M_p + \frac{3}{4}m$ is observed just before going to $m_{||} = M_p$ at $R = 0$. This is what equations give; we have problems with rationalizing such things though.

The approach reported is able to produce the nonadiabatic corrections to all rovibrational levels corresponding to the ground electronic state.[22]

6.5.3 The key message: the potential energy surface (PES) and molecular shape

In the case of a polyatomic molecule with N atoms ($N > 2$), V_{kJ} depends on $3N - 6$ variables determining the configuration of the nuclei. The potential energy function for the motion of the nuclei, $V_{kJ}(\mathbf{R})$, represents therefore a surface in $(3N - 5)$-dimensional space (a *hypersurface*) corresponding to the electronic state k and the rotational excitation J. This potential energy (hyper)surface $V_{kJ}(\mathbf{R})$ (equal to $E_k^0(\mathbf{R})$ in the case of $J = 0$ and the Born–Oppenheimer approximation, Fig. 6.5), or PES, for the motion of the nuclei, represents one of the most important ideas in chemistry. As shown in Fig. 6.5 a molecule has many electronic states and multiple stable geometries possible. In Fig. 6.5 the valleys of the $E_k^0(\mathbf{R})$ correspond to various patterns of chemical bonds. The electronic energy minima (of the $E_k^0(\mathbf{R})$) *define the molecular geometries – the subject of ubiquitous textbooks, lectures, and classroom drawings in the realm of chemistry*; an example can be found in Fig. 6.6.

---

[22] It is worth noting that, e.g., for H_2 and its lowest rovibrational level, making μ R-dependent (i.e., using $\mu_{||}$ and $\mu_{\perp}$ and neglecting $\delta\mathcal{E}_{na}(R)$) gives 84% of the total nonadiabatic effect, while neglecting this R dependence, i.e., putting $\mu_{||} = \mu_{\perp} = \mu$ and taking $\delta\mathcal{E}_{na}(R)$ into account, gives 15%; these two effects seem to be quite independent.

Fig. 6.5 highlights, however, that the internal molecular vibrations move the molecular geometry from its minimum-energy position and the molecule remains electronically stable only if it vibrates in a single valley.

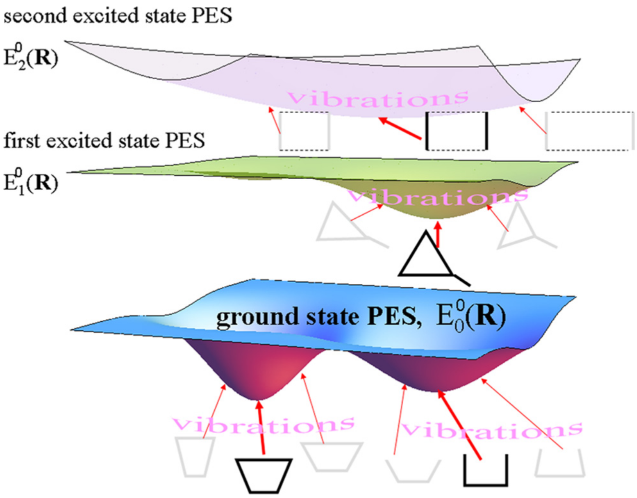


Fig. 6.5. The crucial concept of chemistry: the molecular 3D structure. Within the Born–Oppenheimer approximation and no rotation assumed ($J = 0$, see Eqs. (6.22), (6.20), and (6.26)) the electronic energy $E_k^0(\mathbf{R})$ as a function of $\mathbf{R}$ becomes the PES for the motion of the nuclei (their configuration denoted as $\mathbf{R}$, the corresponding axes not shown). The figure shows an example of three (fictitious) PESs: for the ground electronic state ($k = 0$) and for the two excited electronic states ($k = 1, 2$). The electronic energy minima, indicated by the thick arrows, correspond to the electronically most stable three-dimensional structures, each having some fixed chemical bond pattern (this idea will be introduced in Chapter 8). Around these stable positions vibrations are possible (more detail in Chapter 7) and the resulting off-minimum configurations $\mathbf{R}$ (indicated by the label "vibrations" and the thin arrows) differ from those corresponding to the minima. Some of the vibrations may lead to breaking of the chemical bond, as shown at the bottom of the figure. Almost all chemical phenomena, except those of photochemistry, depend on the electronic ground-state PES, $E_0^0(\mathbf{R})$.

This concept makes possible contact with what chemists call the spatial "structure" of a molecule, identified with its nuclear configuration corresponding to the minimum of the PES for the electronic ground state. It is only because of the adiabatic approximation that we may imagine the three-dimensional shape of a molecule as a configuration of its nuclei bound by the chemical bonds (Fig. 6.6). This object moves and rotates in space, and in addition, the nuclei vibrate about their equilibrium positions with respect to other nuclei (which may be visualized as a rotation-like motion close to the minimum of an energy valley).

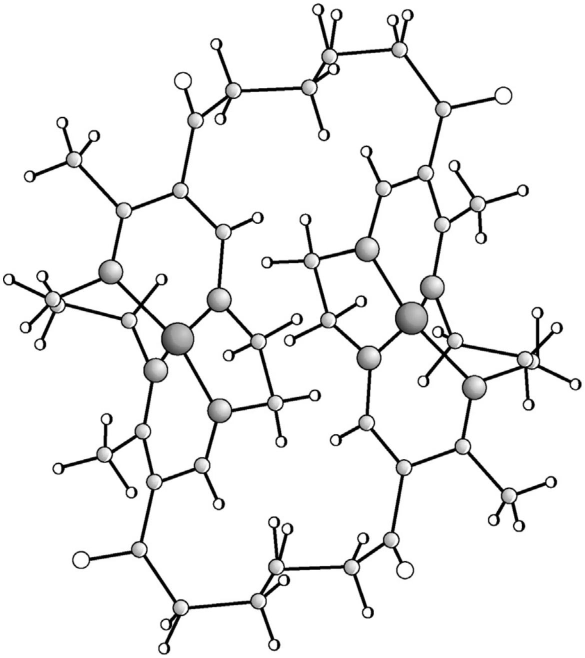


Fig. 6.6. A three-dimensional *model* (called its 3D "structure") of a molecule allows us to focus our attention on spatial and temporal relations that are *similar* to those we know from the macroscopic world. Although the concept of "spatial structure" may occasionally fail, in virtually all cases in chemistry and physics we use a three-dimensional molecular model which resembles what is shown in the figure for a particular molecule (using a two-dimensional projection of the three-dimensional model). There are "balls" and "connecting sticks." The balls represent atoms (of various sizes, the size characterizes the corresponding element), the sticks of different length are supposed to represent what are called "chemical bonds." What should be taken seriously, what not? First, the scale. The real molecule is about 100 000 000 times smaller than the picture in the figure. Second, the motion. This static model shows a kind of averaging over all the snapshots of the real vibrating atoms. In Chapter 8 we will see that indeed the atoms of which the molecule is composed stay together because of a pattern (which characterizes the electronic state of the molecule) of interatomic chemical bonds that *to some extent resemble* sticks. An atom in a molecule is never spherically symmetric, but can be *approximated* by its spherical core ("ball"). The particular molecule in the figure has two tetraazaanulene macrocycles that coordinate two Ni^{2+} ions (the largest spheres). The macrocycles are held together by two $-(CH_2)_4-$ molecular links. Note that any atom of a given type binds a certain number of its neighbors. The most important message is: *if such structural information offered by the three-dimensional molecular model were not available, it would not be possible to design and carry out the complex synthesis of the molecule.* Courtesy of Professor B. Korybut-Daszkiewicz.

Fig. 6.7 shows the sequence of the approximations, which lead to the concept of the three-dimensional structure. First, Fig. 6.7a shows the Born–Oppenheimer PES resulting from the assumption of infinite nuclear masses. Fig. 6.7b demonstrates that the Born–Oppenheimer and adiabatic PESs differ slightly as a result of the correction ($H'_{kk} \neq 0$) for the finite mass of the nuclei. Fig. 6.7c shows that rotation of the molecule ($J \neq 0$) destabilizes the molecule, while the concept of the PES is still valid for a given J.

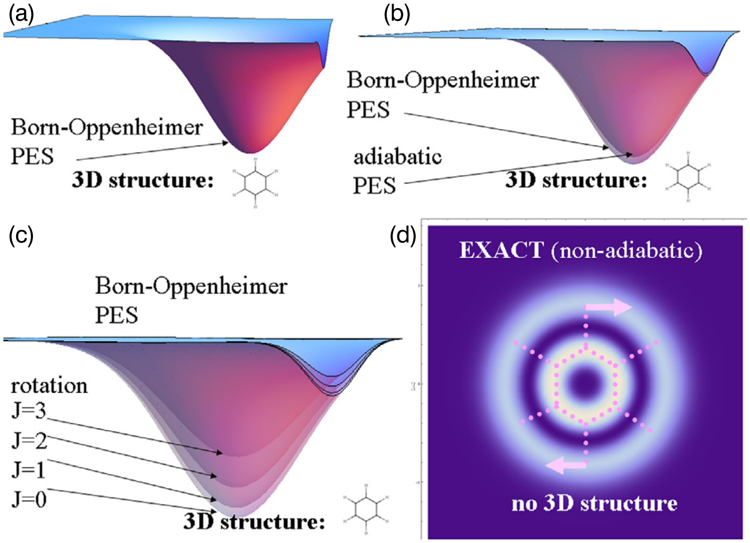


Fig. 6.7. A sequence of approximations versus the concept of the PES – a schematic presentation. (a) The Born–Oppenheimer approximation (infinite mass of the nuclei assumed) leads us to the PES, identical for all isotopomers, as well as to the concept of a three-dimensional molecular structure (in this case of the benzene molecule [C_6H_6]). (b) The adiabatic and the Born–Oppenheimer approximations lead to a bit different PESs, as a result of the mass of the nuclei assumed to be finite, e.g., for different isotopomers. (c) The PESs differ also within the Born–Oppenheimer approximation for different rotational states ($J = 0, 1, ...$). (d) Beyond the adiabatic approximation, in an exact, or nonadiabatic, theory any molecule has a spherical symmetry (due to the space isotropy) with respect to the center of mass. In our case the carbon atoms and the hydrogen atoms become smeared out (which is suggested in the figure by the arrows) within the two fuzzy rings of probability density of finding a particle (shown as a section $z = 0$ of a spherical distribution). The ideas of the three-dimensional structure and of the PES are now irrelevant, while the three-dimensional spatial relationships among the particles' positions become a complicated matter of the interparticle correlations hidden in the total wave function.

Without the adiabatic approximation (Fig. 6.7d), questions about the molecular three-dimensional structure of the benzene molecule could only be answered in a very enigmatic way, e.g.,

- any molecule must exhibit spherical symmetry (any orientation is equally probable), because its ground rotational state does not depend on orientation in space ($Y_{J=0}^{M=0} = \text{const}$);
- the molecule does not have any particular three-dimensional shape;
- the motion of the electrons and nuclei is very complicated;
- correlations of motion of all the particles exist (electron–electron, nucleus–nucleus, electron–nucleus);
- these correlations are in general very difficult to elucidate.

Identical answers would be given if we were to ask about the structure of the DNA molecule. Obviously, something is going wrong; we should expect more help from theory.

For the benzene molecule (Fig. 6.7d), we could answer questions like "What is the mean value of the carbon–carbon, carbon–proton, proton–proton, electron–electron, electron–proton, and electron–carbon distances in the benzene molecule in its ground and excited states?" Note that because all identical particles are indistinguishable, for example, the carbon–proton distance pertains to any carbon and any proton, and so on. To discover that the benzene molecule is essentially a planar hexagonal object (Fig. 6.7a,b,c) would be quite difficult. What could we say about a protein? A pile of paper with such numbers would give us *the* true (nonrelativistic though) picture of the protein molecule, but it would be useless, just like a map of the world with 1:1 scale would be useless for a tourist. It is just too exact. If we relied on this, progress in the investigation of the molecular world would more or less stop. A radical approach in science, even if more rigorous, is very often less fruitful or fertile. Science needs models, simpler than reality but capturing the essence of it, which direct human thought towards more fertile regions.

The adiabatic approximation offers a *simple three-dimensional model* of a molecule – *an extremely useful concept* with great interpretative potential. In the chapters to come, this model will gradually be enriched by introducing the notion of chemical bonds between *some* atoms, angles between consecutive chemical bonds, electronic lone pairs, electronic pairs that form chemical bonds, etc. Such a model inspires our imagination.[23] This is the foundation of all chemistry, all organic syntheses, conformational analysis, most of spectroscopy, etc. Without this beautiful model progress in chemistry would be extremely difficult.

Chemists' life between a sesame entrance and an abyss

What is the Born–Oppenheimer structure for C_6H_6? Every chemist thinks about benzene – a familiar hexagonal planar structure, which is very stable (see p. 197) and preserved in many

---

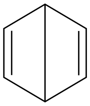
[23] Sometimes too much. We always have to remember that the useful model represents nothing more than a kind of better or worse pictorial representation of a more complex and unknown reality.

reactions. This is, however, a safe side of our story. If one considers possible displacements of the atoms, one finds some other Born–Oppenheimer structures possible, in the spirit of Fig. 6.5. How many? Well, chemists have calculated that, and 217 basic isomeric skeletal structures have been found[24]! If one takes into account their possible diastereomers and enantiomers, the total number of isomers is 328. They are acyclic (17), monocyclic (71), bicyclic (134), tricyclic (87), or tetracyclic (19).

Although all structures formally satisfy the valency requirements of carbon (4) and hydrogen (1), many of the isomers are strained, which means high electronic energy E_k^0. Some are strained to such an extent that we have to do even with "inverted" carbon atoms, having four bonds pointing to one hemisphere. This intramolecular stress reduces considerably the chances to see such chemical structures. Still, however, among all the isomers, as many as 80 C_6H_6 Born–Oppenheimer structures (with possible substituents) look as prospective synthesis targets. And now a surprise: *40 of these Born–Oppenheimer C_6H_6 structures have already been synthesized!*

Why do we omit them in chemical practice? Well, not always. It is true that usually the exotic Born–Oppenheimer structures correspond to high electronic energy and therefore we do not see them in experiment. However, there are cases, when the low-energy Born–Oppenheimer structures are defavorized by a specific chemical neighborhood like sterical hindrance. For example, if one substitutes the six hydrogen atoms in benzene with six bulky substituents, there is simply no space to keep them all in plane. As a result, the system's ground state does not look as

but instead it acquires what theoreticians know as the Dewar resonance structure:



Note that C_6H_6 is a relatively simple object. What awaits us when considering larger systems? This discussion will be continued on p. 376 and illustrated in Fig. 6.25. The humans are only at the beginning of chemistry.

---

[24] After Gopalpur Nagendrappa, *Resonance*, 6(2001)74.

6.6 Basic principles of electronic, vibrational and rotational spectroscopy

6.6.1 Electronic and vibrational structure

Eq. (6.25) represents the basis of molecular spectroscopy, because its solutions produce the molecular electronic, vibrational, or rotational states of a diatomic molecule. Fig. 6.8a shows an example how the curves $U_k(R)$ (also $E_k^0(R)$) may look for three electronic states $k = 0, 1, 2$ of a diatomic molecule. Two of these curves ($k = 0, 2$) have a typical "hook-like" shape for bonding states, the third ($k = 1$) is also typical, but for repulsive electronic states.

It was assumed in Fig. 6.8a that $J = 0$ and therefore $V_{kJ}(R) = U_k(R)$. Next, Eq. (6.25) was solved at a given $U_0(R)$ and a series of solutions $\chi_{kvJ} = \chi_{0v0}$ was found: $\chi_{000}, \chi_{010}, \chi_{020}, \ldots$ with energies $E_{000}, E_{010}, E_{020}, \ldots$, respectively. Then, in a similar way, for $k = 2$, the series of solutions $\chi_{200}, \chi_{210}, \chi_{220}, \ldots$ with the corresponding energies $E_{200}, E_{210}, E_{220}, \ldots$ is obtained. This means that these two electronic levels ($k = 0, 2$) have *a vibrational structure* ($v = 0, 1, \ldots$); the corresponding vibrational levels are shown in Fig. 6.8b. Any attempt to find the vibrational levels for the electronic state $k = 1$ would fail.

The pattern of the vibrational levels looks similar to those for the Morse oscillator (p. 224). The low levels are nearly equidistant, reminding us of the results for the harmonic oscillator. The corresponding wave functions also resemble those for the harmonic oscillator. Higher-energy vibrational levels are getting closer and closer, as for the Morse potential. This is a consequence of the anharmonicity of the potential – we are just approaching the dissociation limit where the $U_k(R)$ curves differ qualitatively from the harmonic potential.

6.6.2 Rotational structure

What would happen if we took $J = 1$ instead of $J = 0$? This corresponds to the potential energy curves $V_{kJ}(R) = U_k(R) + J(J+1)\hbar^2/(2\mu R^2)$, in our case $V_{k1}(R) = U_k(R) + 1(1+1)\hbar^2/(2\mu R^2) = U_k(R) + \hbar^2/(\mu R^2)$ for $k = 0, 1, 2$. The new curves (Fig. 6.9) therefore represent the old curves plus the term $\hbar^2/(\mu R^2)$, which is the same for all the curves. This corresponds to a *small* modification of the curves for large R and a *larger* modification for small R. The potential energy curves just go up a little bit on the left.[25] Of course, this is why the solution of Eq. (6.25) for these new curves will be similar to that which we had before, but this tiny shift upwards will result in a tiny shift upwards of all the computed vibrational levels. Therefore the levels E_{kv1} for $v = 0, 1, 2, \ldots$ will be a little higher than the corresponding E_{kv0} for $v = 0, 1, 2, \ldots$ (this pertains to $k = 0, 2$, there will be no vibrational states for $k = 1$). This means that each vibrational level v will have its own *rotational structure* corresponding to $J = 0, 1, 2, \ldots$.

---

[25] With an accompanying small shift to the right of the position of the minimum.

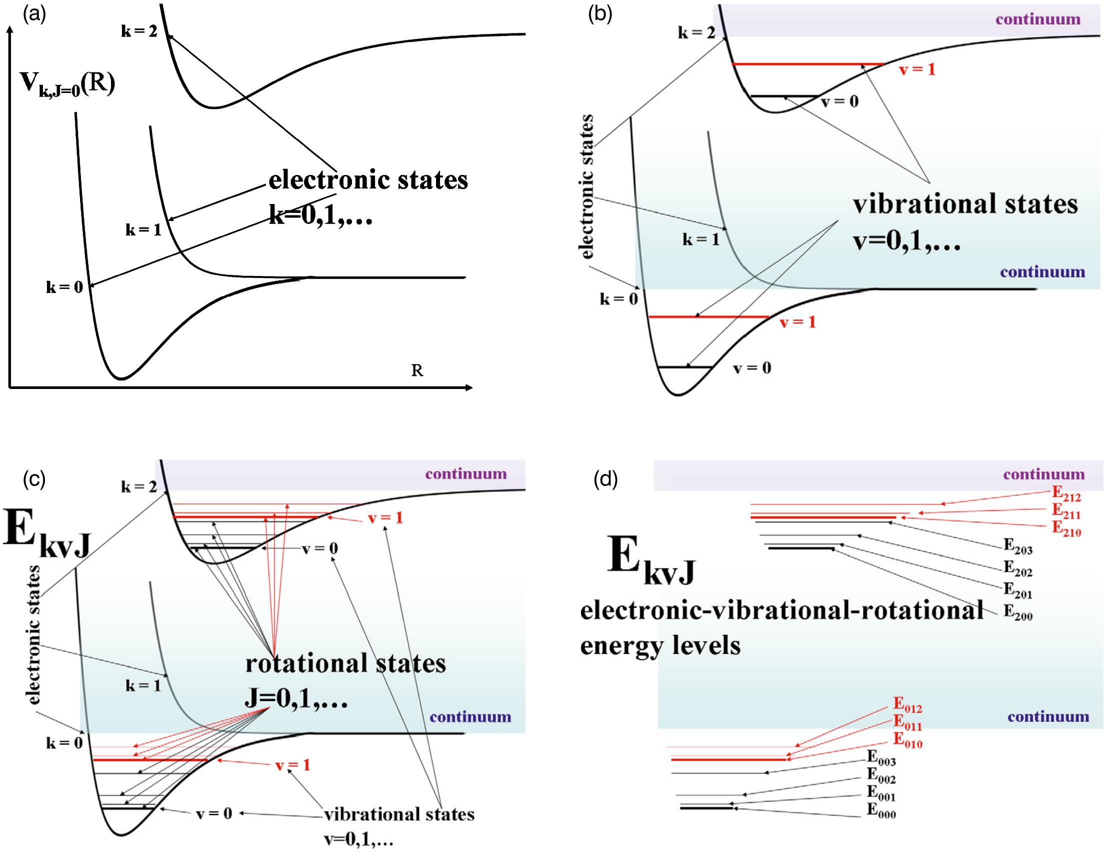


Fig. 6.8. The physical basis of molecular spectroscopy in the UV-VIS, infrared, and microwave regions, i.e., how the electronic-vibrational-rotational energy levels are emerging step by step from the theory (example of a diatomic, R is the interatomic distance). (a) Solutions ($k = 0, 1, 2$) of the Schrödinger equation with the electronic Hamiltonian (Eq. (6.4) on p. 314): electronic energies as functions of R. (b) These functions serve as the potential energies for describing vibration and rotation of the molecule (Eqs. (6.25)–(6.26) on p. 322). In the figure two vibrational levels $v = 0, 1$ are shown (the horizontal sections) for $k = 0, 2$ and no such a state is found for $k = 1$. This result has been obtained for the rotational quantum number $J = 0$ (no rotation assumed). Two energy continua are visible as well: they correspond to possible dissociations of the corresponding electronic states. (c) If the same analysis is repeated for a given $J \neq 0$, one obtains a similar set of the states except that all of them are shifted upwards in the energy scale. This produces for each vibrational level its rotational progression for $J = 0, 1, 2, \ldots$. (d) All this produces a set of the electronic-vibrational-rotational energy levels allowed for the diatomic. Even if this set looks a bit mysterious at the beginning, its origin is now clear.

Increasing J means that the potential energy curve becomes shallower.[26] At some lower J's the molecule may accommodate all or part of the vibrational levels that exist for $J = 0$. It

---

[26] The curve $V_{kJ}(R)$ becomes more shallow, the system gets less stable, but for small J the force constant paradoxically (the second derivative at minimum, if any) *increases, i.e., the system becomes more stiff due to rotation.*

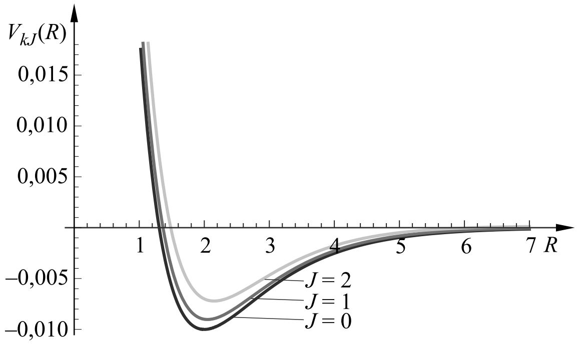


Fig. 6.9. The potential energy curve, $V_{kJ}(R)$, for the motion of the nuclei for a diatomic, where k means the quantum number for the electronic state and $J = 0, 1, 2$ are the rotational quantum numbers. One can see a weakening of the binding energy by the centrifugal force.

may happen, however, that after a high-energy rotational excitation (to a large J) the potential energy curve will be so shallow that no vibrational energy level will be possible. This means that the molecule will undergo dissociation due to the excessive centrifugal force.

Separation between energy levels

For molecules other than hydrides, the separation between rotational levels $(E_{kvJ+1} - E_{kvJ})$ is smaller by two to three orders of magnitude than the separation between vibrational levels $(E_{k,v+1,J} - E_{kvJ})$, and the latter is smaller by one or two orders of magnitude when compared to the separation of the electronic levels $(E_{k+1,v,J} - E_{kvJ})$.

> This is why electronic excitation corresponds to absorption of UV or visible light, vibrational excitation to absorption of infrared radiation, and rotational excitation to absorption of microwave radiation.

This is what we use in a microwave oven. Some chicken on a ceramic plate is irradiated by microwaves. This causes rotational excitation of the water molecules[27] always present in food. The “rotating” water molecules cause a transfer of kinetic energy to protein, similarly as would happen in traditional cooking. After removing the food from the microwave the chicken is hot, but the plate is cool (nothing to rotate in it).

Indeed, the second derivative of the rotational energy is equal to $J(J+1)\frac{3\hbar^2}{\mu R^4} > 0$ and, if the position of the minimum of the new curve shifted only a bit (J not too large) with respect to the position of the minimum of $U_k(R)$, the force constant would increase due to the rotational excitation!

[27] Such rotation is somewhat hindered in the solid phase.

In practice we always have to do with the absorption or emission spectra of a specimen from which we are trying to deduce the relative positions of the energy levels of the molecules involved. We may conclude that, in theoretical spectra computed in the center-of-mass system, there will be allowed and forbidden energy intervals.[28] There are no energy levels in the forbidden intervals.[29] In the allowed intervals, any region corresponds to an electronic state, whose levels exhibit a pattern, i.e., clustering into vibrational series: one cluster corresponding to $v = 0$, the second to $v = 1$, etc. Within any cluster we have rotational levels corresponding to $J = 0, 1, 2, \ldots$. This follows from the fact that the distances between the levels with different k are large, with different v are smaller, and with different J are even smaller.

6.7 Approximate separation of rotations and vibrations

Vibrations cannot be exactly separated from rotations for a very simple reason: during vibrations the length R of the molecule changes, and this changes the moment of inertia $I = \mu R^2$ and influences the rotation of the molecule[30] according to Eq. (6.26), p. 322.

The separation is feasible only when making an approximation, e.g., when assuming the mean value of the moment of inertia instead of the moment itself. Such a mean value is close to $I = \mu R_e^2$, where R_e stands for the position of the minimum of the potential energy V_{k0}. So, we may decide to accept the potential (6.26) for the vibrations in the approximate form[31]

$$V_{kJ}(R) \approx U_k(R) + J(J+1)\frac{\hbar^2}{2\mu R_e^2}.$$

Since the last term is a constant, this immediately gives the separation of the rotations from the vibrational equation (6.25)

$$\left(-\frac{\hbar^2}{2\mu}\frac{d^2}{dR^2} + U_k(R)\right)\chi_{kvJ}(R) = E'\chi_{kvJ}(R), \tag{6.29}$$

where the constant

$$E' = E_{kvJ} - E_{rot},$$

---

[28] In an SFCS (see p. 691) we always have to do with a continuum of states (due to translations, see p. 77).

[29] Corresponding to bound states. The nonbound states densely fill the total energy scale above the dissociation limit of the ground state.

[30] Let us recall the energetic pirouette of a charming dancer. Her graceful movements, stretching arms out or aligning them along the body, immediately translate into slow or fast rotational motion.

[31] Which looks reasonable for *small* amplitude vibrations only. This amplitude becomes however larger under rotational excitations. Thus, in principle R_e should increase if J increases, and the rotational energy is therefore lower than shown by the formula.

$$E_{rot}(J) = J(J+1)\frac{\hbar^2}{2\mu R_e^2}. \tag{6.30}$$

Now, we may always write the potential $U_k(R)$ as a *number* $U_k(R_e)$ plus the "rest" labeled by $V_{vibr}(R)$, i.e.,

$$U_k(R) = U_k(R_e) + V_{vibr}(R). \tag{6.31}$$

Then, it is appropriate to call $U_k(R_e)$ the *electronic energy* E_{el} (corresponding to the equilibrium internuclear distance in electronic state k), while the function $V_{vibr}(R)$ stands, therefore, for the vibrational potential satisfying $V_{vibr}(R_e) = 0$. After introducing this into Eq. (6.29) we obtain the equation for vibrations (in general anharmonic)

$$\left(-\frac{\hbar^2}{2\mu}\frac{d^2}{dR^2} + V_{vibr}(R)\right)\chi_{kvJ}(R) = E_{vibr}\chi_{kvJ}(R),$$

where the constant $E_{vibr} = E' - E_{el}$, hence (after adding the translational energy – recalling that we have separated the center-of-mass motion) we have the final approximation

$$E_{kvJ} \approx E_{trans} + E_{el}(k) + E_{vibr}(v) + E_{rot}(J), \tag{6.32}$$

where the corresponding quantum numbers are given in parentheses: the electronic (k), the vibrational (v), and the rotational (J).

6.8 Understanding the IR spectrum of a diatomic: HCl

Assume we have a diluted gas[32] of HCl and we are testing its optical absorption in the microwave region. It is worth to note that H atoms and the Cl atoms by themselves have zero absorption in this range of spectrum. The spectrum of the HCl molecules represents however a strange sequence of double peaks in a peculiar quasiperiodic order. This means the absorption is a direct result of making a molecule from these atoms. We will have to do there with some relative motion of the two interacting atoms, which will be described by molecular vibrational and rotational states and optical transitions between them (the electronic state staying the same).

---

[32] No intermolecular interaction will be assumed.

6.8.1 Selection rules are consequences of conservation laws

Not all transitions are allowed.

All selection rules stem ultimately from conservation laws.

> The conservation of energy law makes that only a photon of energy $\hbar\omega$ that fits the difference of energy levels can be absorbed.

This fitting is however not enough. There must be also a coupling (oscillating with frequency ω) between the electromagnetic field and the system. From the theory of interaction of matter and electromagnetic field, we know that the most important coupling term is equal to $-\hat{\boldsymbol{\mu}} \cdot \boldsymbol{\mathcal{E}} = -(\hat{\mu}_x \mathcal{E}_x + \hat{\mu}_y \mathcal{E}_y + \hat{\mu}_z \mathcal{E}_z)$ (cf. p. 114), where $\boldsymbol{\mathcal{E}}$ is the oscillating electric field vector of the electromagnetic field and $\hat{\boldsymbol{\mu}}$ is the dipole moment operator. We will assume that the electromagnetic wave propagates along the z axis; therefore $\mathcal{E}_z = 0$ and only $\hat{\mu}_x$ and $\hat{\mu}_y$ will count. The quantity $\boldsymbol{\mathcal{E}}$ provides the necessary oscillations in time, while the absorption is measured by $|c|^2$ with $c = \left\langle \Psi_k | \hat{\boldsymbol{\mu}}(\mathbf{r}, \mathbf{R}) \Psi_{k'} \right\rangle_{e,n}$ the coupling between the initial electronic rovibrational state $\Psi_k(\mathbf{r}, \mathbf{R}) = \psi_k\,(\mathbf{r}; \mathbf{R})\, f_k\,(\mathbf{R}) = \psi_{k=0}\,(\mathbf{r}; \mathbf{R})\, \chi_{v=0}(R) Y_J^M(\theta, \phi)$ and the final electronic rovibrational state $\Psi_{k'}(\mathbf{r}, \mathbf{R}) = \psi_{k'=0}\,(\mathbf{r}; \mathbf{R})\, \chi_{v'}(R) Y_{J'}^{M'}(\theta, \phi)$, where we decided to be within the ground electronic state ($k = k' = 0$) and start from the ground vibrational state ($v = 0$) and the symbol $\langle | \rangle_{e,n}$ denotes the integration over the coordinates of all electrons ("e") and nuclei ("n").[33] We will integrate first within $c = \left\langle \Psi_k | \hat{\boldsymbol{\mu}} \Psi_{k'} \right\rangle_{e,n}$ over electronic coordinates, $\left\langle \psi_0\,(\mathbf{r}; \mathbf{R})\, | \hat{\boldsymbol{\mu}}(\mathbf{r}, \mathbf{R}) \psi_0\,(\mathbf{r}; \mathbf{R}) \right\rangle_e \equiv \boldsymbol{\mu}_{00}(\mathbf{R})$, and we obtain $c = \left\langle \chi_{v=0}(R) Y_J^M(\theta, \phi) | \boldsymbol{\mu}_{00}(\mathbf{R}) \chi_{v'}(R) Y_{J'}^{M'}(\theta, \phi) \right\rangle_n$. The quantity $\boldsymbol{\mu}_{00}(\mathbf{R})$ is the dipole moment of the molecule in the ground electronic state and oriented in space along $\mathbf{R}$.

> In case $\boldsymbol{\mu}_{00}(\mathbf{R}) = 0$ (also in case $\boldsymbol{\mu}_{00}(\mathbf{R}) \cdot \boldsymbol{\mathcal{E}} = 0$ for any $\mathbf{R}$) there will be no absorption. Thus, to get a nonzero absorption in rotational and vibrational (microwave or IR) spectra, one has to do with polar molecules at least for certain $\mathbf{R}$. Therefore, all homonuclear diatomics, although having a rich structure of rovibrational levels, are unable to absorb electromagnetic radiation in the microwave as well as in the IR range.

A vector in three-dimensional space may be defined in a Cartesian coordinate system by giving the x, y, z components, but also in the spherical coordinate system by giving R, θ, ϕ polar

---

[33] When describing the electronic function we have put explicitly the position in space of the nuclei ($\mathbf{R}$) instead of the usual notation with R (which does not tell us how the nuclear axis is oriented in space).

coordinates. Now, let us write the dipole moment $\boldsymbol{\mu}_{00}(\mathbf{R})$ in spherical coordinates[34]: $\boldsymbol{\mu}_{00}(\mathbf{R}) = \sqrt{\frac{8\pi}{3}}\mu_{00}(R)Y_1^m(\theta,\phi)$ (see p. V2-615) with $m = \pm 1$, while $m = 0$ is automatically excluded since it represents $\boldsymbol{\mu}_{00,z} = 0$, which is irrelevant in view of $\mathcal{E}_z = 0$ (no coupling in such a case). Thus,

$$c = \left\langle \chi_{v=0}(R)Y_J^M(\theta,\phi)|\boldsymbol{\mu}_{00}(\mathbf{R})\chi_{v'}(R)Y_{J'}^{M'}(\theta,\phi)\right\rangle_n =$$
$$\sqrt{\frac{8\pi}{3}}\,\langle\chi_0(R)|\mu_{00}(R)\chi_{v'}(R)\rangle_R \left\langle Y_J^M|Y_1^m Y_{J'}^{M'}\right\rangle_{\theta,\phi},$$

where at each integral we have indicated the coordinates to integrate over. Now, introducing the equilibrium internuclear distance R_e (the position of the minimum of the potential energy curve) and the displacement $Q = R - R_e$, as well as expanding $\mu_{00}(R)$ in the Taylor series, $\mu_{00}(R) = \mu_{00}(R_e) + \left(\frac{\partial\mu_{00}}{\partial R}\right)_{R=R_e} Q + ...$, and neglecting the higher terms denoted as $+...$, one obtains

$$c = \sqrt{\frac{8\pi}{3}}\left[\mu_{00}(R_e)\,\langle\chi_0(R)|\chi_{v'}(R)\rangle_R + \left(\frac{\partial\mu_{00}}{\partial R}\right)_{R=R_e}\langle\chi_0(R)|Q\chi_{v'}(R)\rangle_R\right]\left\langle Y_J^M|Y_1^m Y_{J'}^{M'}\right\rangle_{\theta,\phi} = 0 +$$
$$\sqrt{\frac{8\pi}{3}}\left(\frac{\partial\mu_{00}}{\partial R}\right)_{R=R_e}\langle\chi_0(R)|Q\chi_{v'}(R)\rangle_R \left\langle Y_J^M|Y_1^{\pm1} Y_{J'}^{M'}\right\rangle_{\theta,\phi}.$$

There is only one such wave function $\chi_{v'}$ of the harmonic oscillator, for which $\langle\chi_0(R)|Q\chi_{v'}(R)\rangle_R \neq 0$: it happens only for[35] $v' = 1$.

We obtain the selection rule for IR spectroscopy: it is necessary that during the vibration the dipole moment changes. The main effect of the IR absorption from the $v = 0$ state is that the vibrational quantum number has to change from 0 to 1.

The integral $\left\langle Y_J^M|Y_1^{\pm1} Y_{J'}^{M'}\right\rangle_{\theta,\phi}$ is nonzero only if[36] $M' = M - m$, $J' = J \pm 1$. This integral has to do with conservation of the total angular momentum and with the conservation of the parity of the system. Any photon has the spin quantum number[37] $s = 1$ (cf. p. 29), which means that

---

[34] $\mu^{(10)} = qR\sqrt{\frac{4\pi}{3}}Y_1^0 = \mu(R)\sqrt{\frac{4\pi}{3}}Y_1^0$,
$\mu^{(1,\pm1)} = \mu(R)\sqrt{\frac{8\pi}{3}}Y_1^{\pm1}$.

[35] Simply $Q\chi_0(R)$ is proportional to the Hermite polynomial H_1, i.e., is proportional to $\chi_1(R)$. Due to the orthonormal character of all χ_v, this gives $v = 1$ as the only possibility.

[36] The rule $M' = M - m$ follows from $\int_0^{2\pi}\exp[i(-M + m + M')\phi]d\phi = 2\pi\delta_{M',M-m}$.

[37] With two polarizations: $m_s = 1$ or $m_s = -1$; the polarization $m_s = 0$ is excluded due to the zero mass.

besides its energy it carries the angular momentum $\hbar$ or $-\hbar$ (right or left circular polarizations of the photon, the electric field $\mathcal{E}$ rotating within the xy plane). After absorption the photon disappears, but *it does not matter*: *the total angular momentum has to be conserved whatever happens*. Therefore, the total system, molecule + photon, before as well as after absorption, *has to have* the total angular momentum with the quantum number equal to[38] $|J-s|, J, J+s$, i.e., $J-1, J, J+1$. The second possibility (with J) would mean that in IR spectroscopy violation of parity occurs.[39] Indeed, the parity of Y_J^M is equal[40] to $(-1)^J$. Therefore, the case $J'=J$ in view of parity of $Y_1^{\pm 1}$ would mean an odd function to integrate, which would give the integral equal zero.[41] Thus,

> the selection rule for IR and microwave spectroscopy reads as follows: no photon absorption can happen unless $\Delta J = \pm 1$.

6.8.2 Microwave spectrum gives the internuclear distance

The lowest energy needed to excite the system would be to change J only; the related frequencies (for transitions that are allowed by selection rules[42]: $kvJ = 00J \to 00(J+1)$, $J = 0, 1, 2, \ldots$) are in the range of microwaves. From Eq. (6.30) we get the theoretical estimation of the transition energy $h\nu = hc\bar{\nu} = (J+1)(J+2)\frac{\hbar^2}{2\mu R_e^2} - J(J+1)\frac{\hbar^2}{2\mu R_e^2} = 2(J+1)\frac{\hbar^2}{2\mu R_e^2} = (J+1)2B$. Using the recorded microwave spectrum we may estimate from this formula the equilibrium interatomic distance for HCl. For the consecutive J we get 1.29 Å, independently of J (not too large though[43]). Thus from the microwave spectrum of HCl we can read the "interatomic distance." We may compare this distance with, say, the position of the minimum of the computed potential energy curve $U_0(R_e)$ of Eq. (6.31).

6.8.3 IR spectrum and isotopic effect

What about the IR region? Fig. 6.10a gives the recorded absorption.

---

[38] We will describe this problem of quantum mechanical adding of two angular momenta in a more general way on p. 405.

[39] The conservation of parity is violated in Nature, but this effect is much too small to be seen in the analyzed spectrum.

[40] Please recall the s,p,d,... orbitals of the hydrogen atom. They correspond to Y_l^m, $l = 0, 1, 2, \ldots$, respectively, and they are of even ($l = 0, 2$) or odd parity ($l = 1$).

[41] This is why we do not have the peak ("missing"): $v = 0, J = 0 \to v = 1, J = 0$.

[42] See Appendix C.

[43] Because the minimum of the potential energy curve V_{0J} shifts for large J (see Fig. 6.16 on p. 354).

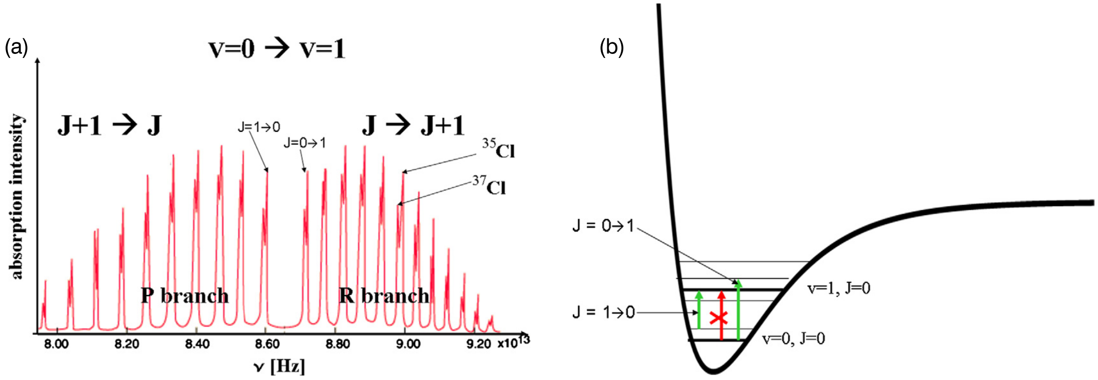


Fig. 6.10. What can we learn about the HCl molecule from its IR spectrum? (a) The IR spectrum (each doublet results from two chlorine isotopes: ^{35}Cl and ^{37}Cl present in the specimen). (b) The central position in the spectrum seems to be missing, because the transition $v = 0, J = 0 \rightarrow v = 1, J = 0$ is forbidden by the selection rules (see the text); its hypothetical position can be determined with high precision as the mean value of the two transitions shown: $J = 0 \rightarrow J = 1$ and $J = 1 \rightarrow J = 0$. This allows us to compute the force constant of the HCl bond. The energy difference of the same two quanta allows us to estimate the moment of inertia, and therefore the H–Cl distance. Note that the rotational levels corresponding to the vibrational state $v = 1$ are closer to each other than those for $v = 0$ due to the wider and wider well and longer and longer equilibrium distance corresponding to the rotationally corrected potential for the motion of the nuclei.

First, where do these strange doublets in the IR spectrum for HCl come from (Fig. 6.10)? Well, the reason is quite trivial: two natural chlorine isotopes[44]: ^{35}Cl and ^{37}Cl, always present in the natural specimen (with proportion 3 : 1). The H^{35}Cl molecule rotates (and oscillates) differently than the H^{37}Cl, because of the reduced mass difference (see Eq. (6.30)). This difference in μ is very small, since what decides μ is the small mass of the proton. Thus we understand this superposition of the "two sets of spectra," shifted a bit with respect to one another, the heavier isotope towards smaller frequencies.

Fig. 6.10a can be understood with the help of Eq. (6.29), which plays for us the role of a model of the phenomena taking place. At room temperature most of the molecules (Boltzmann law) are in their ground electronic and vibrational state ($k = 0$, $v = 0$). IR quanta are unable to change the quantum number k; they have however sufficient energy to change quantum numbers v and J. Fig. 6.10b shows what in fact has been recorded. From the transition selection rules (see above) we have $\Delta v = 1 - 0 = 1$ and either the transitions of the kind $\Delta J = (J + 1) - J =$

---

[44] The presence of a deuterium-substituted molecule has very serious consequences (a larger shift of the spectrum), because what first of all counts for the reduced mass is the light atom. And the reduced mass decides about rotation and vibration.

$+1$ (what is known as the R branch, right-hand side of the spectrum) or of the kind $\Delta J = J - (J+1) = -1$ (the P branch, left-hand side).

6.8.4 Internuclear distance

The remarkable regularity of the spectrum comes from the fact that the transition energy *difference (of the nearest neighbor peak positions)* in a given branch is:

- for the R branch: $E_{excit,J} = h\nu = h\nu_0 + (J+1)(J+2)\frac{\hbar^2}{2\mu R_e^2} - J(J+1)\frac{\hbar^2}{2\mu R_e^2} = (J+1)2\frac{\hbar^2}{2\mu R_e^2} = (J+1)2B$ and $E_{excit,J+1} - E_{excit,J} = (J+2)2B - (J+1)2B = 2B$,
- for the P branch: $E_{excit,J} = h\nu = h\nu_0 + J(J+1)\frac{\hbar^2}{2\mu R_e^2} - (J-1)\,J\frac{\hbar^2}{2\mu R_e^2} = J2\frac{\hbar^2}{2\mu R_e^2} = J2B$ and $E_{excit,J} - E_{excit,J-1} = J2B - (J-1)2B = 2B$.

From the known distance $2B$ we can compute the estimation for the equilibrium distance R_e. We see they are indeed quite equidistant in the spectrum for the R branch and for the P branch separately, but there is a small difference in the B's for these branches. Is the theory described wrong? No, only our oversimplified theory fails a little. The B for the P branch is a bit larger, because the mean interatomic distance gets larger for larger J (due to the centrifugal force).

6.8.5 Why we have a spectrum "envelope"

What about the overall shape of the peaks' intensity ("the envelope") of the R and P branches? It looks quite strange: as if the transition from the levels with $v = 0$, $J = 2$ and $v = 0$, $J = 3$ had the largest intensity... Why? Look, please, at the rotational levels; they are so close that they are significantly populated at a given temperature. At thermal equilibrium the population of the levels by HCl molecules is proportional to the degeneracy of the level number J times the Boltzmann factor, i.e., to $p(J;T) = (2J+1)\exp[-\frac{J(J+1)B}{k_B T}]$. Let us find for which J the probability[45] $p(J;T)$ attains a maximum:

$$\frac{\mathrm{d}p}{\mathrm{d}J} = 0 = 2\exp[-\frac{J(J+1)B}{k_B T}] - (2J+1)\frac{(2J+1)\,B}{k_B T}\exp[-\frac{J(J+1)B}{k_B T}],$$

which gives for J_{opt} the equation $2 - (2J_{opt}+1)^2\frac{B}{k_B T} = 0$, or $(2J_{opt}+1)^2 = \frac{2k_B T}{B}$. For $T = 300$ K this gives $J_{opt} = 2.7$, i.e., between $J = 2$ and $J = 3$. This is just what we see! We may say, therefore, that the spectrum shown has been recorded close to room temperature.

---

[45] Not normalized to unity, but it does not matter here.

6.8.6 Intensity of isotopomers' peaks

There is a problem left. Since the isotopes ^{35}Cl and ^{37}Cl occur at a ratio of 3 : 1, we might expect a similar intensity ratio of the two spectra. Why then do we have the ratio (Fig. 6.10a) looking more like 4 : 3 (for low J)? There are two possible explanations: heavier rotator and heavier oscillator have lower energies and their levels are more populated at nonzero temperature, and/or this spectrum is of too low resolution and we are comparing the maxima, while we should compare the integral intensity of the peaks.[46] It turns out that in a higher resolution spectrum the second effect prevails, and for the integral intensities we indeed see the ratio 3 : 1.

Thus, we may say we understand the spectrum of HCl given in Fig. 6.10a.

6.9 A quasiharmonic approximation

The detailed form of $V_{vibr}(R)$ is obtained from $U_k(R)$ of Eq. (6.31) and therefore from the solution of the Schrödinger equation (6.25) with the clamped nuclei Hamiltonian. In principle there is no other way but to solve Eq. (6.29) numerically. It is tempting, however, to get an idea of what would happen if a harmonic approximation were applied, i.e., when a harmonic spring was installed between both vibrating atoms. Such a model is very popular when discussing molecular vibrations. There is an unexpected complication though: such a spring cannot exist even in principle. Indeed, even if we constructed a spring that elongates according to Hooke's law, *one cannot ensure the same for shrinking*. It is true that at the beginning, the spring may fulfill the harmonic law for shrinking too, but when $R \to 0_+$ the two nuclei just bump into each other and the energy goes to infinity instead of being parabolic. For the spring to be strictly harmonic, we have to admit $R < 0$, which is simply forbidden, because R is a distance. Fig. 6.11 shows the difference between the harmonic potential and the quasiharmonic approximation for Eq. (6.29).

What do we do? Well, sticking to principles is always the best choice.[47] Yet, even in the case of the potential wall shown in Fig. 6.11c we have an analytical solution.[48] The solution is quite complex, but it gets much simpler assuming $\frac{V_0}{h\nu} \equiv \alpha \gg v$, where $v = 0, 1, 2, \ldots$ stands for the vibrational quantum number we are going to consider and $V_0 \equiv V_{vibr}(0)$. This means that

---

[46] That is, the area under the signal recorded.

[47] Let me stress once more that the problem appears when making the quasiharmonic approximation, not in the real system we have.

[48] E. Merzbacher, "*Quantum mechanics*," Wiley, New York, 2nd edition, 1970. The solution we are talking about has to be extracted from a more general problem in the reference above. The potential energy used in the reference also has its symmetric counterpart for $R < 0$. Hence, the solution needed here corresponds to the antisymmetric solutions in the more general case (only for such solutions where the wave function is equal to zero for $R = 0$).

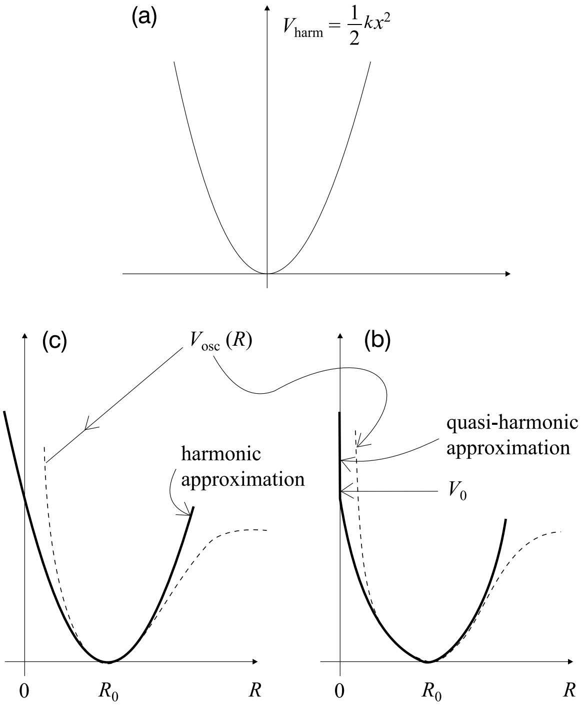


Fig. 6.11. The difference between harmonic and quasiharmonic approximations for a diatomic molecule. (a) The potential energy for the harmonic oscillator. (b) The harmonic approximation to the oscillator potential $V_{vibr}(R)$ for a diatomic molecule is nonrealistic, since at $R = 0$ (and at $R < 0$) the energy is finite, whereas it should go asymptotically to infinity when R tends to 0. (c) A more realistic (quasiharmonic) approximation: the potential is harmonic up to $R = 0$, and for negative R it goes to infinity. The difference between the harmonic and quasiharmonic approximations pertains to such high energies (high oscillation amplitudes) that practically it is of negligible importance. In cases b and c, there is a range of small amplitudes where the harmonic approximation is applicable.

we limit ourselves to those vibrational states that are far below V_0. This is quite satisfactory, because the hypothetical bump of the two nuclei would occur at vast (even unrealistic) V_0. In such a case the vibrational energy is equal to $E_v = h\nu\left(v' + \frac{1}{2}\right)$, where the modified "quantum number" $v' = v + \varepsilon_v$ with a tiny modification

$$\varepsilon_v = \frac{1}{\sqrt{2\pi}}\frac{1}{v!}(4\alpha)^{v+\frac{1}{2}}\exp(-2\alpha).$$

The corresponding wave functions very much resemble those of the harmonic oscillator, except that for $R \leq 0$ they are equal to zero. The strictly harmonic approximation results in $\varepsilon_v = 0$, and therefore, $E_v = h\nu\left(v + \frac{1}{2}\right)$ (see Chapter 4).

We conclude that the quasiharmonic approximation means almost the same as the (less realistic) harmonic one.

6.10 Polyatomic molecules

6.10.1 Kinetic energy expression

A similar procedure can be carried out for polyatomic molecules.

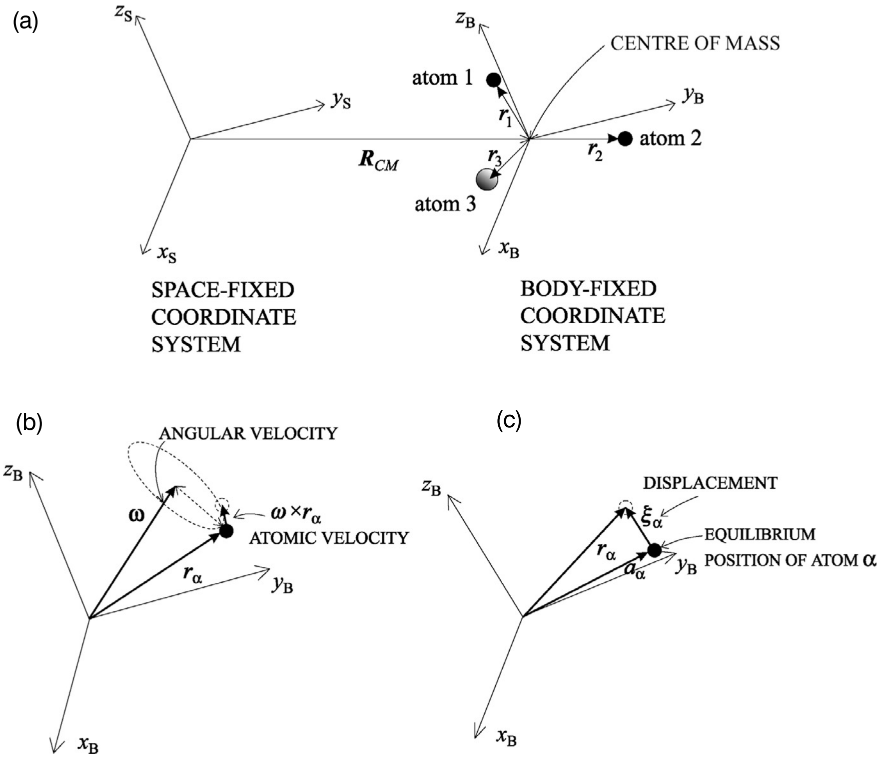


Fig. 6.12. Space- and body-fixed coordinate systems (SFCS and BFCS) for a polyatomic molecule. (a) SFCS is a Cartesian coordinate system arbitrarily chosen in space (left). The origin of the BFCS is located in the center of mass of the molecule (right). The center of mass is shown by the vector $\mathbf{R}_{CM}$ from the SFCS. The nuclei of the atoms are indicated by vectors $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \ldots$ from the BFCS. (b) The velocity of atom α when the system is rotating with an angular velocity given as vector $\boldsymbol{\omega}$. In such a case the atom acquires additional velocity $\boldsymbol{\omega} \times \mathbf{r}_\alpha$. (c) If the molecule vibrates, then atomic positions $\mathbf{r}_\alpha$ differ from the equilibrium positions $\mathbf{a}_\alpha$ by the displacements $\boldsymbol{\xi}_\alpha$.

Let us consider a Cartesian SFCS (see Appendix J on p. 691) and vector $\mathbf{R}_{CM}$ indicating the center of mass of a molecule composed of M atoms (Fig. 6.12). Let us construct a Cartesian

BFCS with the origin in the center of mass and the axes parallel to those of the SFCS (the third possibility in Appendix J).

In the BFCS an atom α of mass[49] M_α is indicated by the vector $\mathbf{r}_\alpha$, and its equilibrium position[50] by $\mathbf{a}_\alpha$. The vector of displacement is $\boldsymbol{\xi}_\alpha = \mathbf{r}_\alpha - \mathbf{a}_\alpha$. If the molecule were *rigid and did not rotate* in the SFCS, then the velocity of the atom α would be equal to $\mathbf{V}_\alpha = \frac{d}{dt}(\mathbf{R}_{CM} + \mathbf{r}_\alpha) = \dot{\mathbf{R}}_{CM}$ (dots mean time derivatives), because the vector $\mathbf{r}_\alpha$, indicating the atom from the BFCS, would not change at all. If, in addition, the molecule, still preserving its rigidity, *rotated* about its center of mass with angular velocity $\boldsymbol{\omega}$ (the vector having the direction of the rotation axis, right-handed screw orientation, and length equal to the angular velocity in radians per second), then the velocity of the atom α would equal[51] $\mathbf{V}_\alpha = \dot{\mathbf{R}}_{CM} + (\boldsymbol{\omega} \times \mathbf{r}_\alpha)$. However, our molecule *is not rigid*; everything moves inside it (let us call these motions "vibrations"[52]). Note that *no restriction was made with respect to the displacements* $\boldsymbol{\xi}_\alpha$ – they could be some giant internal motions. Then the velocity of the atom α with respect to the SFCS is

$$\mathbf{V}_\alpha = \dot{\mathbf{R}}_{CM} + (\boldsymbol{\omega} \times \mathbf{r}_\alpha) + \dot{\boldsymbol{\xi}}_\alpha. \tag{6.33}$$

When these velocities $\mathbf{V}_\alpha$ are inserted into the kinetic energy T of the molecule calculated in the SFCS, we get

$$T = \frac{1}{2}\sum_\alpha M_\alpha (\mathbf{V}_\alpha)^2 = \frac{1}{2}\left(\dot{\mathbf{R}}_{CM}\right)^2 \sum_\alpha M_\alpha + \frac{1}{2}\sum_\alpha M_\alpha (\boldsymbol{\omega} \times \mathbf{r}_\alpha)^2 + \frac{1}{2}\sum_\alpha M_\alpha \left(\dot{\boldsymbol{\xi}}_\alpha\right)^2 +$$
$$\dot{\mathbf{R}}_{CM} \cdot \left[\boldsymbol{\omega} \times \left(\sum_\alpha M_\alpha \mathbf{r}_\alpha\right)\right] + \dot{\mathbf{R}}_{CM} \cdot \sum_\alpha M_\alpha \dot{\boldsymbol{\xi}}_\alpha + \sum_\alpha M_\alpha (\boldsymbol{\omega} \times \mathbf{r}_\alpha) \cdot \dot{\boldsymbol{\xi}}_\alpha.$$

The first three ("diagonal") terms have a clear interpretation, i.e., the kinetic energy of the center of mass, the kinetic energy of rotation, and the kinetic energy of vibrations. The last three terms ("nondiagonal") denote the *roto-translational, vibro-translational, and vibro-rotational couplings*.

---

[49] What this mass means was discussed earlier in this chapter.

[50] We assume that such a position exists. If there are several equilibrium positions, we just choose one of them.

[51] $|\boldsymbol{\omega} \times \mathbf{r}_\alpha| = \omega r_\alpha \sin\theta$, where θ stands for the angle axis/vector $\mathbf{r}_\alpha$. If the atom α is on the rotation axis, this term vanishes ($\theta = 0$ or π). In other cases the rotation radius is equal to $r_\alpha \sin\theta$.

[52] Such a "vibration" may mean a vibration of the OH bond, but also a rotation of the –CH_3 group or a large displacement of a molecular fragment.

6.10.2 Quasirigid model – simplifying by Eckart conditions

There is a little problem with the expression for the kinetic energy: we have a redundancy in the coordinates. Indeed, we have three coordinates for defining translation ($\mathbf{R}_{CM}$), three that determine rotation ($\boldsymbol{\omega}$), and on top of that M vectors $\mathbf{r}_\alpha$. Too many. Six are redundant. Using such coordinates would be very annoying, because we would not be sure whether they are consistent.

We may impose six relations among the coordinates and in this way (if they are correct) get rid of the redundancy. The first three relations are evident, because the origin of the BFCS is simply the center of mass. Therefore,

$$\sum_\alpha M_\alpha \mathbf{r}_\alpha = \mathbf{0}, \tag{6.34}$$

also true when the atoms occupy equilibrium positions

$$\sum_\alpha M_\alpha \mathbf{a}_\alpha = \mathbf{0}.$$

Hence, we obtain a useful relation,

$$\sum_\alpha M_\alpha \left(\mathbf{r}_\alpha - \mathbf{a}_\alpha\right) = \mathbf{0},$$

$$\sum_\alpha M_\alpha \boldsymbol{\xi}_\alpha = \mathbf{0},$$

which, after differentiation with respect to time, becomes

$$\sum_\alpha M_\alpha \dot{\boldsymbol{\xi}}_\alpha = \mathbf{0}. \tag{6.35}$$

If there were several sets of $\mathbf{a}_\alpha$'s, i.e., several minima of the potential energy, we would have a problem. This is one of the reasons we need the assumption of the quasirigid molecule.

Inserting (6.34) *and* (6.35) *into the kinetic energy expression makes the roto-translational and vibro-translational couplings vanish.* Thus, we have

$$T = \frac{1}{2}\left(\dot{\mathbf{R}}_{CM}\right)^2 \sum_\alpha M_\alpha + \frac{1}{2}\sum_\alpha M_\alpha \left(\boldsymbol{\omega} \times \mathbf{r}_\alpha\right)^2 + \frac{1}{2}\sum_\alpha M_\alpha \left(\dot{\boldsymbol{\xi}}_\alpha\right)^2 + \\ + \sum_\alpha M_\alpha \left(\boldsymbol{\omega} \times \mathbf{r}_\alpha\right) \cdot \dot{\boldsymbol{\xi}}_\alpha.$$

Noting that $\mathbf{r}_\alpha = \mathbf{a}_\alpha + \boldsymbol{\xi}_\alpha$ and using the relation[53] $(\mathbf{A} \times \mathbf{B}) \cdot \mathbf{C} = \mathbf{A} \cdot (\mathbf{B} \times \mathbf{C})$, we obtain immediately

$$T = \frac{1}{2}\left(\dot{\mathbf{R}}_{CM}\right)^2 \sum_\alpha M_\alpha + \frac{1}{2}\sum_\alpha M_\alpha \left(\boldsymbol{\omega} \times \mathbf{r}_\alpha\right)^2 + \frac{1}{2}\sum_\alpha M_\alpha \left(\dot{\boldsymbol{\xi}}_\alpha\right)^2 +$$
$$+\boldsymbol{\omega} \cdot \sum_\alpha M_\alpha \left(\mathbf{a}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha\right) + \boldsymbol{\omega} \cdot \sum_\alpha M_\alpha \left(\boldsymbol{\xi}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha\right).$$

We completely get rid of the redundancy if the second *Eckart condition*[54] is introduced (equivalent to three conditions for the coordinates):

$$\sum_\alpha M_\alpha \left(\mathbf{a}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha\right) = \mathbf{0}. \tag{6.36}$$

The condition means that *we do not want the internal motion to generate any angular momentum.*[55] This completes our final expression for the kinetic energy T of a polyatomic quasirigid molecule,

$$T = T_{trans} + T_{rot} + T_{vibr} + T_{Coriolis}. \tag{6.37}$$

The kinetic energy in an SFCS is composed of:

- the kinetic energy of the center of mass (translational energy), $T_{trans} = \frac{1}{2}\left(\dot{\mathbf{R}}_{CM}\right)^2 \sum_\alpha M_\alpha$,
- the rotational energy of the whole molecule, $T_{rot} = \frac{1}{2}\sum_\alpha M_\alpha \left(\boldsymbol{\omega} \times \mathbf{r}_\alpha\right)^2$,
- the kinetic energy of the internal motions ("vibrations"), $T_{vibr} = \frac{1}{2}\sum_\alpha M_\alpha \left(\dot{\boldsymbol{\xi}}_\alpha\right)^2$,
- the last term, usually very small, is known as the Coriolis energy,[56] $T_{Coriolis} = \boldsymbol{\omega} \cdot \sum_\alpha M_\alpha \left(\boldsymbol{\xi}_\alpha \times \dot{\boldsymbol{\xi}}_\alpha\right)$; it couples the internal motions ("vibrations") within the molecule with its rotation.

After the Eckart conditions are introduced, all the coordinates, i.e., the components of the vectors $\mathbf{R}_{CM}$, $\boldsymbol{\omega}$, and all $\boldsymbol{\xi}_\alpha$, *can be treated as independent.*

---

[53] These are two ways of calculating the volume of the parallelepiped according to the formula: surface of the base times the height.

[54] Carl Eckart, professor at the California Institute of Technology, contributed to the birth of quantum mechanics (e.g., C. Eckart, *Phys. Rev.*, 28(1926)711).

[55] The problem is whether indeed we do not generate any momentum by displacing the nuclei from their equilibrium positions. A flexible molecule may have quite a number of different equilibrium positions (see Chapter 7). We cannot expect all of them to satisfy Eq. (6.36), where one of these equilibrium positions is treated as a reference. *Assuming* Eq. (6.36) means that we restrict the molecular vibrations to have only small amplitudes about a single equilibrium position (quasirigid model).

[56] Gaspard Gustav de Coriolis (1792–1843), a French engineer and mathematician, director of the École Polytechnique in Paris. In 1829 Coriolis introduced the notion of work, the equivalence of work and energy, and also a coupling of rotation and vibration.

6.10.3 Approximation: decoupling of rotations and vibrations

Since the Coriolis term is small, in the first approximation we may decide to neglect it. Also, when assuming *small vibrational amplitudes* $\boldsymbol{\xi}_\alpha$, which is a reasonable approximation in most cases, we may replace $\mathbf{r}_\alpha$ by the corresponding equilibrium positions $\mathbf{a}_\alpha$ in the rotational term of Eq. (6.36): $\sum_\alpha M_\alpha (\boldsymbol{\omega} \times \mathbf{r}_\alpha)^2 \approx \sum_\alpha M_\alpha (\boldsymbol{\omega} \times \mathbf{a}_\alpha)^2$, in full analogy with Eq. (6.30). After these two approximations have been made the kinetic energy represents the sum of the three *independent* terms (i.e., each depending on different variables)

$$T \approx T_{trans} + T_{rot} + T_{vibr}, \tag{6.38}$$

with $T_{rot} \approx \frac{1}{2} \sum_\alpha M_\alpha (\boldsymbol{\omega} \times \mathbf{a}_\alpha)^2$.

6.10.4 Spherical, symmetric, and asymmetric tops

Eq. (6.38) may serve to construct the corresponding kinetic energy operator for a polyatomic molecule. There is no problem (see Chapter 1) with the translational term: $-\frac{\hbar^2}{2\sum_\alpha M_\alpha} \Delta_{\mathbf{R}_{CM}}$, the vibrational term will be treated in Chapter 7, p. 416.

There is a problem with the rotational term. A rigid body (the equilibrium atomic positions $\mathbf{a}_\alpha$ are used), e.g., the benzene molecule, rotates, but due to symmetry it may have some special axes characterizing the *moments of inertia*. The moment of inertia represents a tensor of rank 3 with the following components:

$$\left\{ \begin{array}{lll} \sum_\alpha M_\alpha (a_{y,\alpha}^2 + a_{z,\alpha}^2) & \sum_\alpha M_\alpha a_{x,\alpha} a_{y,\alpha} & \sum_\alpha M_\alpha a_{x,\alpha} a_{z,\alpha} \\ \sum_\alpha M_\alpha a_{x,\alpha} a_{y,\alpha} & \sum_\alpha M_\alpha (a_{x,\alpha}^2 + a_{z,\alpha}^2) & \sum_\alpha M_\alpha a_{y,\alpha} a_{z,\alpha} \\ \sum_\alpha M_\alpha a_{x,\alpha} a_{z,\alpha} & \sum_\alpha M_\alpha a_{y,\alpha} a_{z,\alpha} & \sum_\alpha M_\alpha (a_{x,\alpha}^2 + a_{y,\alpha}^2) \end{array} \right\},$$

to be computed in the BFCS (see Appendix J on p. 691). The diagonalization of the matrix (Appendix L on p. 703) corresponds to a certain rotation of the BFCS to a coordinate system *rotating with the molecule* (RMCS), and gives as the eigenvalues I_{xx}, I_{yy}, I_{zz}.

When $I_{xx} = I_{yy} = I_{zz}$, the rotating body is called a spherical rotator or a spherical top (example: methane molecule); when $I_{xx} = I_{yy} \neq I_{zz}$, it is called a symmetric top (examples: benzene, ammonia molecules); when $I_{xx} \neq I_{yy} \neq I_{zz}$, then the top is asymmetric (example: water molecule).

Fig. 6.13 gives four classes of rotators (tops).

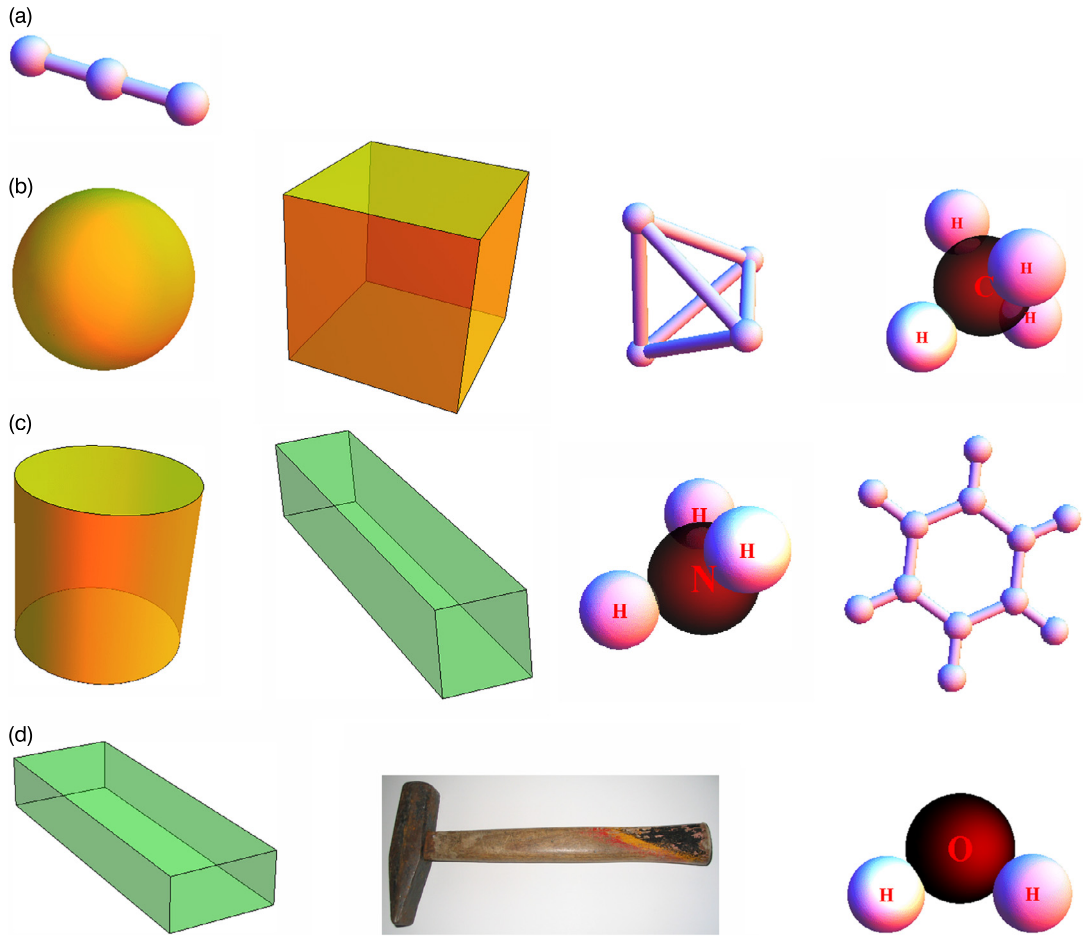


Fig. 6.13. Four classes of tops (rotators) – examples. The numbers I_{xx}, I_{yy}, I_{zz} represent the eigenvalues of the tensor of inertia computed in a BFCS. There are four possibilities: (a) a linear rotator ($I_{xx} = I_{yy} = 0, I_{zz} \neq 0$), e.g., a diatomic or CO_2 molecule; (b) a spherical rotator ($I_{xx} = I_{yy} = I_{zz}$), e.g., a sphere, a cube, a regular tetrahedron, or a methane molecule; (c) a symmetric rotator ($I_{xx} = I_{yy} \neq I_{zz}$), e.g., a cylinder, a rectangular parallelepiped with square base, benzene, or an ammonia molecule; (d) an asymmetric rotator ($I_{xx} \neq I_{yy} \neq I_{zz}$), e.g., a general rectangular parallelepiped, a hammer, or a water molecule.

Then, the classical expression for the kinetic energy of rotation takes the form[57]

$$\frac{1}{2}\sum_{\alpha} M_{\alpha}\left(\boldsymbol{\omega}\times\mathbf{a}_{\alpha}\right)^{2} = \frac{1}{2}\left(I_{xx}\omega_x^2 + I_{yy}\omega_y^2 + I_{zz}\omega_z^2\right) = \frac{J_x^2}{2I_{xx}} + \frac{J_y^2}{2I_{yy}} + \frac{J_z^2}{2I_{zz}}, \tag{6.39}$$

---

[57] H. Goldstein, "*Classical Mechanics*," 2nd edition, Addison-Wesley, 1980.

where $\omega_x, \omega_y, \omega_z$ stand for the components of $\boldsymbol{\omega}$ in the RMCS and J_x, J_y, J_z represent the components of angular momentum also computed in the RMCS.[58]

It is not straightforward to write down the corresponding kinetic energy *operator*. The reason is that in the above expression, we have curvilinear coordinates (because of the rotation from the BFCS to the RMCS[59]), whereas the quantum mechanical operators were introduced (Chapter 1) only for the Cartesian coordinates (p. 23). How do we write an operator expressed in some curvilinear coordinates q_i and the corresponding momenta p_i? Boris Podolsky solved this problem.[60] The result is

$$\hat{T} = \frac{1}{2} g^{-\frac{1}{2}} \hat{\mathbf{p}}^T g^{\frac{1}{2}} \mathbf{G}^{-1} \hat{\mathbf{p}}, \tag{6.40}$$

where $\hat{p}_i = -i\hbar \frac{\partial}{\partial q_i}$ and $\mathbf{G}$ represents a symmetric matrix (*metric tensor*) of the elements g_{rs}, defined by the square of the length element $ds^2 \equiv \sum_r \sum_s g_{rs} dq_r dq_s$, with $g = \det \mathbf{G}$ and g_{rs} (g and all g_{rs} being in general some functions of q_r).

6.10.5 Separation of translational, rotational, and vibrational motions

Eq. (6.40) represents the kinetic energy operator. To obtain the corresponding Hamiltonian we have to add to this energy the potential energy for the motion of the nuclei, U_k, where k labels the electronic state. The last energy depends uniquely on the variables $\boldsymbol{\xi}_\alpha$ that describe atomic vibrations and corresponds to the electronic energy $U_k(R)$ of Eq. (6.31), except that instead of the variable R, which pertains to the oscillation, we have the components of the vectors $\boldsymbol{\xi}_\alpha$. Then, in full analogy with (6.31), we may write

$$U_k(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ...\boldsymbol{\xi}_M) = U_k(\mathbf{0}, \mathbf{0}, ...\mathbf{0}) + V_{k,vibr}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ...\boldsymbol{\xi}_M),$$

where the *number* $U_k(\mathbf{0}, \mathbf{0}, ...\mathbf{0}) = E_{el}$ may be called the *electronic energy in state* k, and $V_{k,vibr}(\mathbf{0}, \mathbf{0}, ...\mathbf{0}) = 0$.

Since (after the approximations made) the translational, rotational, and "vibrational" ("internal motion") operators depend on their own variables, after separation the total wave function represents a product of three eigenfunctions (translational, rotational, and vibrational) and the total

---

[58] We recall from classical mechanics that an expression for rotational motion results from the corresponding one for translational motion by replacing mass by moment of inertia, momentum by angular momentum, and velocity by angular velocity. Therefore, the middle part of the above formula for kinetic energy represents an analog of $\frac{mv^2}{2}$ and the last part is an analog of $\frac{p^2}{2m}$.

[59] The rotation is carried out by performing three successive rotations by what is known as Euler angles. For details see R.N. Zare, "*Angular Momentum*," Wiley, New York, 1988, p. 78.

[60] B. Podolsky, *Phys. Rev.*, 32(1928)812.

energy is the sum of the translational, rotational, and vibrational energies (fully analogous with Eq. (6.32))

$$E \approx E_{trans} + E_{el}(k) + E_{rot}(J) + E_{vibr}(v_1, v_2, ...v_{3M-6}), \tag{6.41}$$

where k denotes the electronic state, J is the rotational quantum number, and v_i are the vibrational quantum numbers that describe the vibrational excitations (in Chapter 7 we will see a harmonic approximation for these oscillations).

6.11 Types of states

6.11.1 Repulsive potential

If we try to solve Eq. (6.29) for vibrations with a repulsive potential, we would not find any solution of class Q. Among continuous, but not square integrable, functions we would find an infinite number of the eigenfunctions and the corresponding eigenvalues would form a continuum. These eigenvalues usually reflect the simple fact that the system has dissociated and its dissociation products may have any kinetic energy larger than the dissociation limit (i.e., having dissociated fragments with no kinetic energy), all energies measured in the SFCS. Any collision of two fragments (that correspond to the repulsive electronic state) will finally result in the fragments flying off. Imagine that the two fragments are located at a distance R_0, with corresponding total energy E, and that the system is allowed to relax according to the potential energy shown in Fig. 6.14a. The system slides down the potential energy curve (the potential energy lowers) and, since the total energy is conserved, its kinetic energy increases accordingly. Finally, the potential energy curve flattens, attaining $E_A + E_B$, where E_A denotes the internal energy of fragment A (similarly for B). The final kinetic energy is equal to $E - (E_A + E_B)$ in the SFCS.

6.11.2 "Hook-like" curves

Another typical potential energy curve is shown in Fig. 6.14b. It has the shape of a hook. Solving (6.29) for such a curve usually[61] gives a series of bound states, i.e., with their wave functions (Fig. 6.15) concentrated in a finite region of space and exponentially vanishing on leaving it. The figure shows the three discrete energy levels found and the continuum of states above the dissociation limit, similar to the curve in Fig. 6.14a. The continuum has, in principle, the same origin as before (any kinetic energy of the fragments).

---

[61] For a sufficiently deep and wide potential energy well.

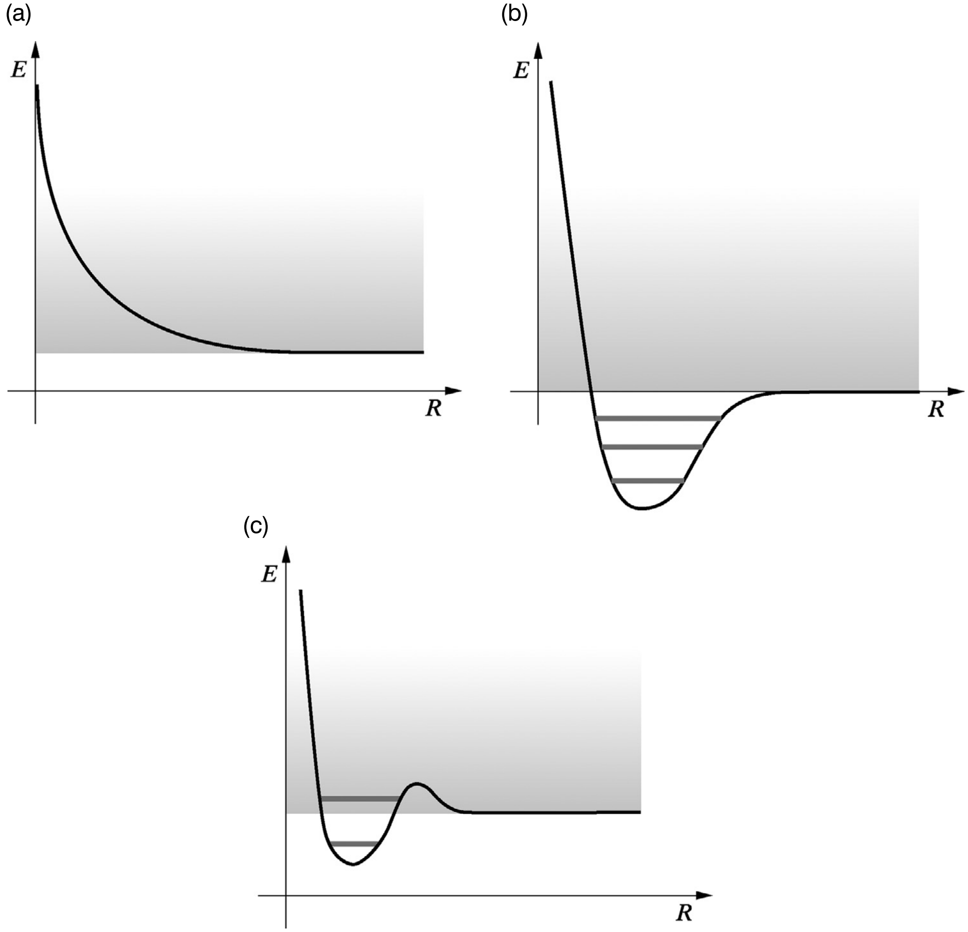


Fig. 6.14. An example of three different electronic states. (a) Repulsive state (no vibrational states, a ball representing the nuclear configuration will slide down, resulting in dissociation). (b) Three bound (vibrational) states (the ball will oscillate within the well, differently in each of the vibrational states). (c) One bound vibrational state (the ball oscillates) and one metastable vibrational state (the ball oscillates for some time and then goes to infinity, which means dissociation). A continuum of allowed states (shadowed area) with nonzero kinetic energy of the dissociation products is above the dissociation limit.

Thus, the overall picture is that a system may have some bound states, but above the dissociation limit it can also acquire any energy and the corresponding wave functions are nonnormalizable (not square integrable).

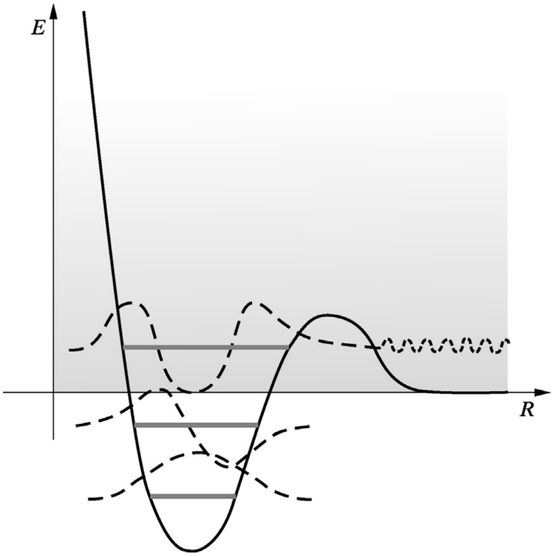


Fig. 6.15. The bound, continuum, and resonance (metastable) states of an anharmonic oscillator. Two discrete bound states are shown (energy levels and wave functions) in the lower part of the figure. The continuum (shaded area) extends above the dissociation limit, i.e., the system may have any of the energies above the limit. There is one resonance state in the continuum, which corresponds to the third level in the potential energy well of the oscillator. Within the well, the wave function is very similar to the third state of the harmonic oscillator, but there are differences. One is that the function has some low-amplitude oscillations on the right-hand side. They indicate that the function is nonnormalizable and that the system will sooner or later dissociate.

6.11.3 Continuum

The continuum may have a quite complex structure. First of all, the number of states per energy unit depends, in general, on the position on the energy scale where this energy unit is located. Thus the continuum may be characterized by the *density of states* (the number of states per unit energy) as a function of energy. This may cause some confusion, because the number of continuum states in any energy section is infinite. The problem is, however, that the infinities differ; some are "more infinite than others." The continuum does not mean a banality of the states involved (Fig. 6.14c). The continuum extends upward the dissociation limit irrespective what kind of potential energy curve one has for finite values of R. In cases similar to that of Fig. 6.14c the continuum will exist independently of how wide and high the barrier is. But the barrier may be so wide that the system will have no idea about any "extra-barrier life," and therefore will have its "quasidiscrete" states with the energy higher than the dissociation limit. Yet, these states despite the similarity to bound states belong to the continuum (are nonnormalizable). Such states are metastable and are called *resonances* (cf. p. 211) or encounter complexes. The system in a metastable state will sooner or later dissociate, but before this happens it may have a quite successful long life. Fig. 6.15 shows how the metastable and stationary states differ: the metastable ones do not vanish in infinity.

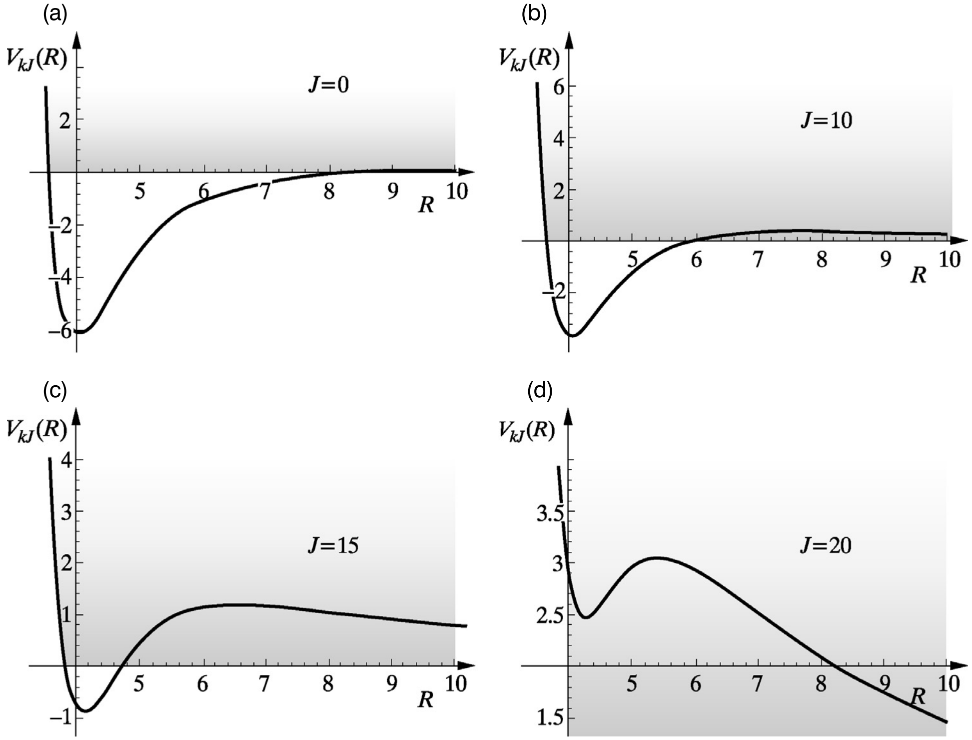


Fig. 6.16. A rotational excitation may lead to creating the resonance states. As an illustration a potential energy curve $V_{kJ}(R)$ has been chosen that resembles what we would have for two water molecules bound by the hydrogen bond. Its first component $U_k(R)$ is taken in the form of the so-called Lennard-Jones potential $U_k(R) = \varepsilon_k[\left(\frac{R_{0k}}{R}\right)^{12} - 2\left(\frac{R_{0k}}{R}\right)^{6}]$ with the following parameters for the electronic ground state ($k = 0$): $\varepsilon_0 = 6$ kcal/mol and $R_{00} = 4$ a.u. and the corresponding reduced mass $\mu = 16560$ a.u. For $J = 0$ (a) the parameter ε_0 stands for the well depth and R_{00} denotes the position of the well minimum. (a)–(d) correspond to $V_{kJ}(R) = U_k(R) + J(J+1)\hbar^2/(2\mu R^2)$ with $J = 0, 10, 15, 20$, respectively. The larger J, the shallower the well: the rotation *weakens* the bond, but in a peculiar way. Due to the centrifugal force a possibility of existence of the metastable resonance states appears. These are the “normal” vibrational states pushed up by the centrifugal energy beyond the energy of the dissociation limit. For $J = 20$ already *all* states (including the potential resonances) belong to the continuum. Note the narrowing of the well after excitations.

As shown in Fig. 6.16, rotational excitations may lead to a qualitative change of the potential energy curve for the motion of the nuclei. Rotational excitations destabilize the molecule, but in a quite specific way. First, they always introduce a barrier for dissociation (*centrifugal barrier*), but in spite of that the dissociation becomes easier due to a large "pushing up" of the well region. Second, by increasing the energy for small distances the rotational excitations either make some vibrational levels disappear or may also change the character of the levels from the stationary ones to metastable vibrational states (vibrational resonances in the continuum). Third, as one can see from Fig. 6.16, there is also another kind of the pushing effect: not only the curves go up, but they also become more narrow (rotational increasing of the force constant, see footnote on p. 333), and therefore the vibrational levels go even further up. Fourth, the equilibrium distance increases upon rotational excitations.

> Besides the typical continuum states, which result from the fact that the dissociation products fly slower or faster, one may also have the continuum metastable or resonance states, which resemble the bound states.

Our human mind wants to translate such situations onto simple pictures, which help to "understand" what happens. Fig. 6.17 shows an analogy associated to astronomy: the Earth and the Moon are in a bound state, the Earth and an asteroid are in a "primitive" continuum-like state, but if it happens that an asteroid went around the Earth several times and then flew away into space, then one has to do with an analog of a metastable or resonance state (characterized by a finite and nonzero lifetime).

The Schrödinger equation $H\psi = E\psi$ is time-independent and, therefore, its solutions do not inform us about the sequence of events, but only of all the possible events with their probability amplitudes.[62] This is why the wave function for the metastable state of Fig. 6.15 exhibits oscillations at large x; they inform us about a possibility of dissociation.

6.11.4 Wave function "measurement"

Could we know the vibrational wave function in a given electronic and rotational state from an experiment? It seemed that such a question could only be answered by quantum mechanical calculations. It turned out,[63] however, that the answer can also come from experiment. In this experiment three states are involved: the electronic ground state (S), an electronic excited state M, in particular its vibrational state (*this state will be measured*), and the third electronic state of a repulsive character (REP) that helps as a detector (see Fig. 6.18).

---

[62] As Einstein said: "*the only reason for time is so that everything does not happen at once*." The time-independent Schrödinger equation behaves as if "*everything would happen at once*."

[63] W. Koot, P.H.P. Post, W.J. van der Zande, J. Los, *Zeit. Physik D*, 10(1988)233. The experimental data pertain to the hydrogen molecule.

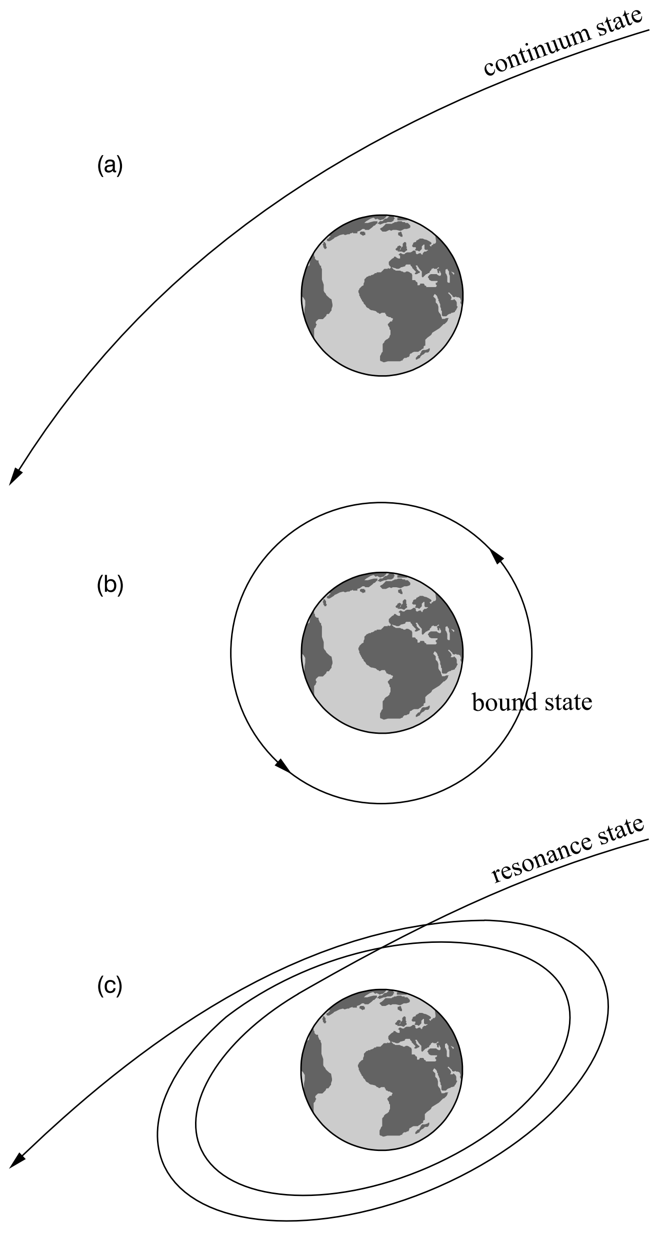


Fig. 6.17. Continuum, bound, and resonance states – an analogy to the "states" of the Earth and an interacting body. (a) A "primitive" continuum state: an asteroid flies by the Earth and changes a little bit its trajectory. (b) A bound state: the Moon is orbiting around the Earth. (c) A resonance state: the asteroid was orbiting several times about the Earth and then flew away.

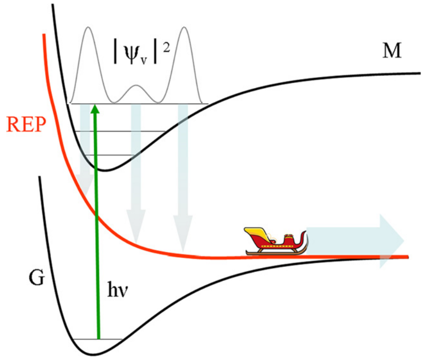


Fig. 6.18. A "measurement" of the wave function ψ_v, or, more exactly, of the corresponding probability density $|\psi_v|^2$. A molecule is excited from its electronic ground state G to a certain vibrational state ψ_v in the electronic excited state M. From M the molecule undergoes a fluorescence transition to the REP state. Since the REP state is of repulsive character the potential energy transforms into the kinetic energy (the total energy being preserved). By measuring the kinetic energy of the dissociation products (that slide from the hill) one is able to calculate what their starting potential energy was, i.e., how high they were on the REP curve. This enables us to calculate $|\psi_v|^2$.

James Franck (1882–1964), German physicist, professor at the Kaiser Wilhelm Institut für Physikalische Chemie in Berlin, then at the University of Göttingen, from 1935 at the John Hopkins University in the USA, and then at the University of Chicago. Frank also participated in the Manhattan Project. As a freshman at the Department of Law at the University of Heidelberg he made the acquaintance of the student Max Born. Born persuaded him to resign from his planned career as a lawyer and choose chemistry, geology, and then physics. In 1914 Franck and his colleague Gustav Hertz used electrons to bombard mercury atoms. The young researchers noted that electrons lose 4.9 eV of their kinetic energy after colliding with mercury atoms. This excess energy is then released by emitting a UV photon. This was the first experimental demonstration that atoms have the electronic energy levels foreseen by Niels Bohr. Both scientists obtained the Nobel Prize in 1925. The fact that, during the First World War, Franck was twice decorated with the Iron Cross was the reason that Franck was one of the few Jews whom the Nazis tolerated in academia. Franck, a citizen of the Third Reich, illegally deposited his Nobel Prize Medal (with his engraved name) in the Niels Bohr Institute in Copenhagen, Denmark. When in April 1940 the attacking German troops marched through the streets of the capital, George de Hevesy (a future Nobel laureate, 1943) hid the golden medal in a strange and very chemical way: just dissolving it in aqua regia. The bottle safely stayed on the shelf the whole occupation period under the nose of the Germans. After the war the Nobel Committee exchanged the bottle for a new medal for Franck.

Edward Condon (1902–1974), American physicist, one of the pioneers of quantum theory in the USA. In 1928 Condon and Gurney discovered the tunneling effect. More widely known is his second great achievement – the Franck–Condon rule.

We excite the molecule from the ground vibrational state of S to a certain vibrational state ψ_v of M using a laser. Then the molecule undergoes a spontaneous *fluorescence transition* to REP. The electronic state changes so fast that the nuclei have no time to move (*Franck–Condon rule*). Whatever falls (vertically, because of the Franck–Condon rule) on the REP state as a result of fluorescence dissociates, because this state is repulsive. The kinetic energy of the dissociation products depends on the internuclear distance R when the fluorescence took place, i.e., on the length of the slide the system had down the REP. How often such an R occurs depends on $|\psi_v(R)|^2$. Therefore, investigating the kinetic energy of the dissociation products gives $|\psi_v|^2$.

6.12 Adiabatic, diabatic, and nonadiabatic approaches

Let us summarize the diabatic, adiabatic, and nonadiabatic concepts (Fig. 6.19).

Adiabatic case. Suppose we have a Hamiltonian $\hat{\mathcal{H}}(\mathbf{r};\mathbf{R})$ that depends on the electronic coordinates $\mathbf{r}$ and parametrically depends on the configuration of the nuclei $\mathbf{R}$; in practical applications, most often $\hat{\mathcal{H}}(\mathbf{r};\mathbf{R}) \equiv \hat{H}_0(\mathbf{r};\mathbf{R})$, the electronic clamped nuclei Hamiltonian corresponding to Eq. (6.8) and generalized to polyatomic molecules. The eigenfunctions $\psi(\mathbf{r};\mathbf{R})$ and the eigenvalues $E_i(\mathbf{R})$ of the Hamiltonian $\hat{\mathcal{H}}(\mathbf{r};\mathbf{R})$ are called *adiabatic* (Fig. 6.19). If we take $\hat{\mathcal{H}} = \hat{H}_0(\mathbf{r};\mathbf{R})$, then in the adiabatic approximation (p. 318) the total wave function is represented by the product

$$\Psi(\mathbf{r},\mathbf{R}) = \psi(\mathbf{r};\mathbf{R})\, f(\mathbf{R}), \tag{6.42}$$

where $f(\mathbf{R})$ is a rovibrational wave function that describes the rotations and vibrations of the system.

Diabatic case. Imagine now a basis set $\bar{\psi}_i(\mathbf{r};\mathbf{R})$, $i = 1, 2, 3, \ldots M$, of some particular electronic wave functions (we will call them *diabatic*) that also depend parametrically on $\mathbf{R}$. There are two reasons for considering such a basis set. The first is that we are going to solve the Schrödinger equation $\hat{\mathcal{H}}\Psi_i = E_i\Psi_i$ by using the Ritz method (Chapter 5) and we need a basis set of the expansion functions

$$\psi(\mathbf{r};\mathbf{R}) \approx \sum_i^M c_i \bar{\psi}_i(\mathbf{r};\mathbf{R}). \tag{6.43}$$

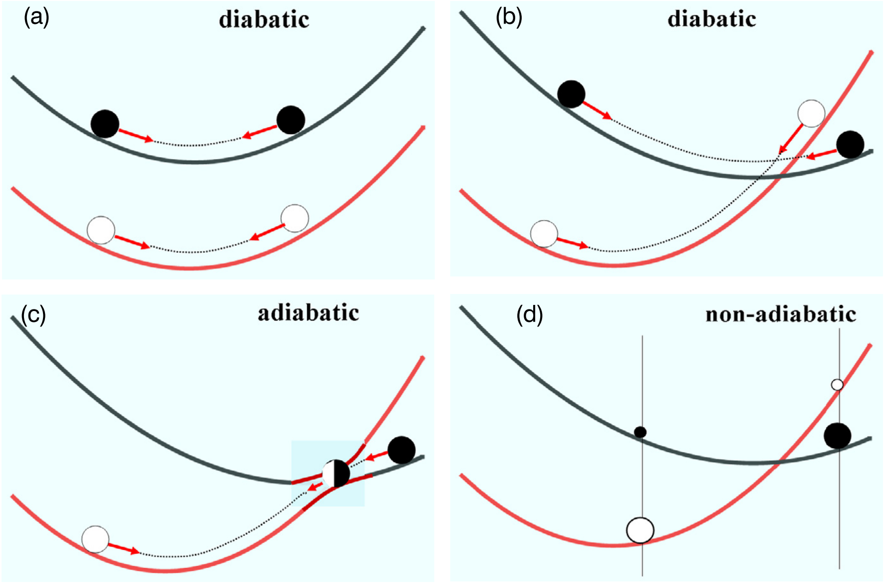


Fig. 6.19. The diabatic, adiabatic, and nonadiabatic approaches to the motion of nuclei (a schematic view). (a) A state that preserves the chemical structure for any molecular geometry is called diabatic (e.g., is always ionic, or always covalent). The energies of these states are calculated as the mean values of the clamped nuclei Hamiltonian. In the lower-energy state, the system is represented by a white ball (say, in the ionic state), in the second the system is represented by the black ball (say, covalent structure). These balls vibrate all the time in the corresponding wells, preserving the chemical structure. (b) It may happen that two diabatic states cross. If the nuclear motion is fast, the electrons are unable to adjust and the nuclear motion may take place on the diabatic curves (i.e., the bond pattern does not change during this motion). (c) The adiabatic approach, where the diabatic states mix (mainly at a crossing region). Each of the adiabatic states is an eigenfunction of the clamped nuclei Hamiltonian. If the nuclear motion is slow, the electrons are able to adjust to it instantaneously and the system follows the lower adiabatic curve. The bond pattern changes qualitatively during this motion (black ball changes to white ball, e.g., the system undergoes a transition from covalent to ionic). The total wave function is a product of the adiabatic electronic state and a rovibrational wave function. (d) The nonadiabatic approach. In this particular case, two diabatic curves come into play. The total wave function is the sum of two functions (their contributions are geometry-dependent, a larger ball means a larger contribution), each function is a product of a diabatic electronic state times a rovibrational wave function. The system is shown at two geometries (vertical lines). Changing the nuclear geometry, it is as if the system has moved on two diabatic surfaces at the same time. This motion is accompanied by changing the proportions (visualized by the size of the balls) of the electronic diabatic states, i.e., the chemical character of the system changes during the motion of the nuclei.

The second reason pertains to chemical interpretation: usually any of the diabatic wave functions are chosen as corresponding to a particular electronic distribution (chemical bond pattern) in the system,[64] and from (6.43) we may recognize what kind of chemical structure dominates Ψ. Thus, using the diabatic basis, there is a chance of obtaining insight into the chemistry going on in the system.[65]

The wave functions $\bar{\psi}_i$ are in general nonorthogonal (we assume them normalized). For each of them we may compute the mean value of the energy (the integration is over the electronic coordinates)

$$\bar{E}_i(\mathbf{R}) = \langle \bar{\psi}_i | \hat{\mathcal{H}}(\mathbf{R}) \bar{\psi}_i \rangle, \tag{6.44}$$

and we will call it the diabatic energy.

The key point is that we may compare the eigenvalues and eigenfunctions of $\hat{\mathcal{H}}(\mathbf{R})$, i.e., the adiabatic states with $\bar{E}_i$ and $\bar{\psi}_i$, respectively. If the diabatic states are chosen in a realistic way, they are supposed to be close to the adiabatic states for most configurations $\mathbf{R}$ (Fig. 6.19a,b,c). These relations will be discussed in a minute.

Nonadiabatic case. The diabatic states or the adiabatic states may be used to construct the basis set for the motion of the electrons *and* nuclei in the nonadiabatic approach. Such a basis function is taken as a product of the electronic (diabatic or adiabatic) wave function and of a rovibrational wave function that depends on $\mathbf{R}$. In a nonadiabatic approach the total wave function is a superposition of these product-like contributions, i.e.,

$$\Psi(\mathbf{r}; \mathbf{R}) = \sum_k \bar{\psi}_k(\mathbf{r}; \mathbf{R}) f_k(\mathbf{R}). \tag{6.45}$$

This sum means that in the nonadiabatic approach the motion of the system involves many PESs at the same time (Fig. 6.19d).

The diabatic and the adiabatic electronic states are simply two possible basis sets in nonadiabatic calculations. If the sets were complete, the results would be identical. The first choice

---

[64] Let us take the example of the NaCl molecule: $\bar{\psi}_1$ may describe the ionic Na^+Cl^- distribution, while $\bar{\psi}_2$ may correspond to the covalent bond Na–Cl. The adiabatic wave function $\psi(\mathbf{r}; \mathbf{R})$ of the NaCl molecule may be taken as a superposition of $\bar{\psi}_1$ and $\bar{\psi}_2$. The valence bond (VB) wave functions (VB structures) described in Chapter V2-2 may be viewed as diabatic states.

[65] Very important for chemical reactions, in which a chemical structure undergoes an abrupt change. In chemical reactions large changes of nuclear configuration are accompanied by motions of electrons, i.e., large changes in the chemical bond pattern (a qualitative change of c_i of Eq. (6.43)). Such a definition leaves liberty in the choice of diabatic states. This liberty can be substantially reduced by the following. Let us take two adiabatic states that dissociate to different products, well separated on the energy scale. However, for some reason the two adiabatic energies are getting closer for some finite values of R. For each value of R we define a space spanned by the two adiabatic functions for that R. Let us find in this space two normalized functions that maximize the absolute value of the overlap integral with the two dissociation states. These two (usually nonorthogonal) states may be called diabatic.

underlines the importance of the chemical bond pattern and the interplay among such patterns. The second basis set highlights the order of the eigenvalues of $\hat{\mathcal{H}}(\mathbf{R})$ (the lower/higher-energy adiabatic state).[66]

6.13 Crossing of potential energy curves for diatomics

6.13.1 The noncrossing rule

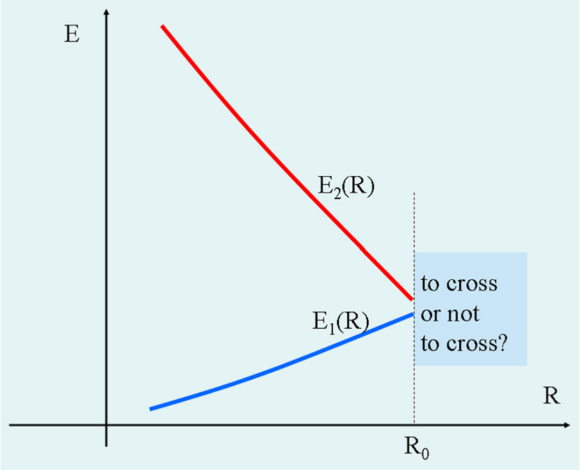


Fig. 6.20. Will the two curves cross?

Can the adiabatic curves $E_k^0(R)$ cross when R changes? (See Fig. 6.20.)

To solve this problem in detail let us limit ourselves to the simplest situation: the two-state model (Appendix D). Let us consider a diatomic molecule and such an internuclear distance R_0 that the two electronic adiabatic states[67] $\psi_1(\mathbf{r}; R_0)$ and $\psi_2(\mathbf{r}; R_0))$ correspond to the nondegenerate (but close in the energy scale) eigenvalues of the clamped nuclei Hamiltonian $\hat{H}_0(R_0)$. Then we have

$$\hat{H}_0(R_0)\psi_i(\mathbf{r}; R_0) = E_i(R_0)\psi_i(\mathbf{r}; R_0), \ i = 1, 2.$$

---

[66] In polyatomic systems there is a serious problem with the adiabatic basis (this is why the diabatic functions are preferred). As we will see later, the adiabatic electronic wave function is multi-valued, and the corresponding rovibrational wave function, having to compensate for this (because the total wave function must be single-valued), also has to be multi-valued.

[67] These states are adiabatic only for $R = R_0$, but when considering $R \neq R_0$ they may be viewed as diabatic (because they are not the eigenfunctions for that R).

Since $\hat{H}_0$ is Hermitian and $E_1 \neq E_2$, we have the orthogonality of $\psi_1(\mathbf{r}; R_0)$ and $\psi_2(\mathbf{r}; R_0)$: $\langle\psi_1|\psi_2\rangle = 0$.

Now, we are interested in solving

$$\hat{H}_0(R)\psi(\mathbf{r}; R) = E\psi(\mathbf{r}; R)$$

for R in the vicinity of R_0 and examine whether is it possible for the energy eigenvalues to cross.

The eigenfunctions of $\hat{H}_0$ will be sought as linear combinations of ψ_1 and ψ_2, i.e.,

$$\psi(\mathbf{r}; R) = c_1(R)\psi_1(\mathbf{r}; R_0) + c_2(R)\psi_2(\mathbf{r}; R_0). \tag{6.46}$$

Note that for this distance R

$$\hat{H}_0(R) = \hat{H}_0(R_0) + V(R), \tag{6.47}$$

and $V(R)$ is certainly small, because R is close to R_0 and $V(R_0) = 0$. Using the Ritz method (Chapter 5, Appendix D, case III), we arrive at two adiabatic solutions, and the corresponding energies read as

$$E_{\pm}(R) = \frac{\bar{E}_1 + \bar{E}_2}{2} \pm \sqrt{\left(\frac{\bar{E}_1 - \bar{E}_2}{2}\right)^2 + |V_{12}|^2}, \tag{6.48}$$

where $V_{ij}(R) \equiv \left\langle\psi_i|\hat{V}(R)\psi_j\right\rangle$ and

$$\bar{E}_i(R) = \left\langle\psi_i(\mathbf{r}; R_0)|\hat{H}_0(R)\psi_i(\mathbf{r}; R_0)\right\rangle = E_i(R) + V_{ii}(R). \tag{6.49}$$

The crossing of the energy curves at a given R means that $E_+ = E_-$, and from this it follows that the expression under the square root symbol has to equal zero. Since, however, the expression is the sum of two squares, the crossing needs *two* conditions to be satisfied simultaneously, i.e.,

$$\bar{E}_1 - \bar{E}_2 = 0, \tag{6.50}$$

$$|V_{12}| = 0. \tag{6.51}$$

Two conditions, and a *single* parameter R to change. If we adjust the parameter to fulfill the first condition, the second one is violated, and *vice versa*. The crossing $E_+ = E_-$ may occur only when, for some reason, e.g., because of the symmetry, the *coupling constant* is automatically equal to zero, $|V_{12}| = 0$, for all R. Then we have only a *single* condition to be fulfilled, and it can be satisfied by changing the parameter R, i.e., *crossing can occur*. The condition $|V_{12}| = 0$ is equivalent to $|H_{12}| \equiv \left\langle \psi_1 | \hat{H}_0(R) \psi_2 \right\rangle = 0$, because $\hat{H}_0(R) = \hat{H}_0(R_0) + \hat{V}$, and $\left\langle \psi_1 | \hat{H}_0(R_0) \psi_2 \right\rangle = 0$ due to the orthogonality of both eigenfunctions of $\hat{H}_0(R_0)$.

Now we will refer to group theory (see Appendix C, p. 605). The Hamiltonian represents a fully symmetric object, whereas the wave functions ψ_1 and ψ_2 are not necessarily fully symmetric, because they may belong to other irreducible representations of the symmetry group. Therefore, in order to make the integral $|H_{12}| = |V_{12}| = 0$, it is sufficient that ψ_1 and ψ_2 transform according to *different* irreducible representations (have different symmetries).[68] Thus,

> the adiabatic curves cannot cross if the corresponding wave functions have the same symmetry.

What will happen if such curves are heading for something that looks like an inevitable crossing? Such cases are quite characteristic and look like an avoided crossing. The two curves look as if they repel each other and avoid the crossing.

> If two states of a diatomic molecule correspond to different symmetry, then the corresponding potential energy curves can cross.

6.13.2 Simulating the harpooning effect in the NaCl molecule

Our goal now is to show, in an example, *what happens to adiabatic states (eigenstates of* $\hat{\mathcal{H}}(R)$*) if two diabatic energy curves (mean values of the Hamiltonian with the diabatic functions) do cross*. Although we are not aiming at an accurate description of the NaCl molecule (we prefer simplicity and generality), we will try to construct a toy (a model) that mimics this particular system.

The sodium atom has 11 electrons (the electronic configuration[69] is $1s^2 2s^2 2p^6 3s^1$), and the chlorine atom contains 17 electrons ($1s^2 2s^2 2p^6 3s^2 3p^5$). The solution of the Schrödinger equation for 28 electrons is difficult. But, we are not looking for trouble. Note that in the NaCl

---

[68] H_{12} transforms according to the representation being the direct product of three irreducible representations: that of ψ_1, that of ψ_2, and that of $\hat{H}$ (the last is, however, fully symmetric and, therefore, does not count in this direct product). In order to have $H_{12} \neq 0$ this direct product, after decomposition into irreducible representations, *has to contain* a fully symmetric irreducible representation. This, however, is possible only when ψ_1 and ψ_2 transform according to *the same* irreducible representation.

[69] What these configurations really mean is explained in Chapter 8.

story the real movie star is a single electron that goes from the sodium to the chlorine atom, making Na^+ and Cl^- ions. The ions attract each other by the Coulombic force and form the familiar ionic bond. But wait a minute! There is a problem. What is of lower energy: the two noninteracting *atoms* Na and Cl or the two noninteracting *ions* Na^+ and Cl^-? The ionization energy of sodium is $I = 495.8$ kJ/mol $= 0.1888$ a.u, whereas the electron affinity of chlorine is only $A = 349$ kJ/mol $= 0.1329$ a.u. This means that the NaCl molecule in its ground state dissociates into *atoms*, not ions.

To keep the story simple, let us *limit ourselves to the single electron* mentioned above.[70] First, let us define the two diabatic states (the basis set) of the system: only the $3s$ orbital of Na (when the electron resides on Na, we have atoms) denoted by $|3s\rangle$ and the $3p$ orbital of Cl (when the electron is on Cl, we have ions) $|3p\rangle$. Now, what about the Hamiltonian $\hat{\mathcal{H}}$? Well, a reasonable model Hamiltonian may be taken as[71]

$$\hat{\mathcal{H}}(\mathbf{r}; R) = -I\ |3s\rangle\langle 3s| - A\,|3p\rangle\langle 3p| - \frac{1}{R}|3p\rangle\langle 3p| + \exp(-R)\,.$$

Indeed, the mean values of $\hat{\mathcal{H}}$ in the $|3s\rangle$ and $|3p\rangle$ states are equal to

$$\bar{E}_1(R) \equiv \mathcal{H}_{11} = \left\langle 3s|\hat{\mathcal{H}}(3s)\right\rangle = -I - AS^2 - \frac{1}{R}S^2 + \exp(-R)\,,$$
$$\bar{E}_2(R) \equiv \mathcal{H}_{22} = \left\langle 3p|\hat{\mathcal{H}}(3p)\right\rangle = -IS^2 - A - \frac{1}{R} + \exp(-R)\,,$$

where (assuming the diabatic functions to be real) the overlap integral $S \equiv \langle 3s|3p\rangle = \langle 3p|3s\rangle$. First of all, this Hamiltonian gives the correct energy limits $\bar{E}_1(R) = -I$ and $\bar{E}_2(R) = -A$, when $R \to \infty$ (the electron binding energy by the sodium and by the chlorine for dissociation into atoms and ions, respectively), which is already very important. The term $\exp(-R)$ mimics the repulsion of the inner shells of both atoms[72] and guarantees that the energies go up (they should do) at $R \to 0$. Note also that the $\bar{E}_1(R)$ and $\bar{E}_2(R)$ curves indeed mimic the approaching Na and Cl and Na^+ and Cl^-, respectively, because in $\bar{E}_2(R)$ there is a Coulomb term $-\frac{1}{R}$, while in $\bar{E}_1(R)$ such an interaction practically disappears for large R. All this gives us a certain confidence that our Hamiltonian $\hat{\mathcal{H}}$ grasps the most important physical effects for the NaCl molecule. The resulting nondiagonal element of the Hamiltonian reads as

---

[70] The other electrons in our approach will only influence the numerical values of the interaction parameters.

[71] $\mathbf{r}$ stands for the coordinates of the electron, for the diatomic molecule R replaces $\mathbf{R}$.

[72] It prevents the two cores collapsing (cf. Chapter V2-5).

$$\left\langle 3s|\hat{\mathcal{H}}(3p)\right\rangle \equiv \mathcal{H}_{12} = S[-I - A - \frac{1}{R} + \exp(-R)].$$

As to S, we could in principle calculate it by taking some approximate atomic orbitals, but our goal is less ambitious than that. Let us simply set $S = R\exp(-R/2)$. Why? Since $S = \langle 3s|3p\rangle = 0$, if $R \to \infty$ or if $R \to 0$, and $S > 0$ for other values of R, at least our formula takes care of this. In addition, Fig. 6.21a,b shows that such a formula for S also gives a quite reasonable set of diabatic curves $\bar{E}_1(R)$ and $\bar{E}_2(R)$: both curves have a single minimum; the minimum for the ionic curve is found at about 5.23 a.u., close to the experimental value of 5.33 a.u., and the binding energy is 0.11 a.u. (0.13 for the adiabatic case, see below), also quite close to the experimental value of 0.15 a.u. Thus, our model to a reasonable extent resembles the real NaCl molecule.

Our goal is to compute the adiabatic energies using the diabatic basis chosen, Eq. (6.43). Appendix D (general case) gives the eigenvalues ($E_+(R)$ and $E_-(R)$) and the eigenfunctions (ψ_+ and ψ_-). Fig. 6.21c,d shows the adiabatic compared to the diabatic curves. The avoided crossing at about 17.9 a.u. is the most important. If the two *atoms* begin to approach (Fig. 6.21a, light gray) the energy does not change too much (flat energy curve), but if the *ions* do the same the energy goes down, because of the long-range Coulombic attraction (dark gray). Thus, the two adiabatic curves (which nearly coincide with the two diabatic curves, especially for large R) are going to cross each other (Figs. 6.21a,b), but the two states have the same symmetry with respect to the molecular axis (as witnessed by $S \neq 0$). Therefore, the crossing cannot occur (Fig. 6.21d). As a result, the two curves avoid the crossing and, as shown in Fig. 6.21c–f, the "atomic" curve switches to the "ionic" curve, and *vice versa*. This switching means an electron jump from Na to Cl and, therefore, formation of the ions Na^+ and Cl^- (then the ions approach fast – this is the *harpooning effect*, introduced to chemistry by Michael Polanyi). The jump occurs at long distances, of the order of 9 Å.

Is this jump inevitable in experiments? It depends on the systems involved and the velocity of the nuclei.

> If the electron is able to adapt instantaneously to the position of the nuclei (slow nuclear motion), the system follows the adiabatic curve. If the nuclear motion is very fast, the system follows the diabatic curve and no electron transfer takes place. The electron transfer is more probable if the gap $2|\mathcal{H}_{12}|$ between $E_+(R)$ and $E_-(R)$ is large.

For large distances the adiabatic states are practically identical with the diabatic states, except in the avoided crossing region (Fig. 6.21c,d).

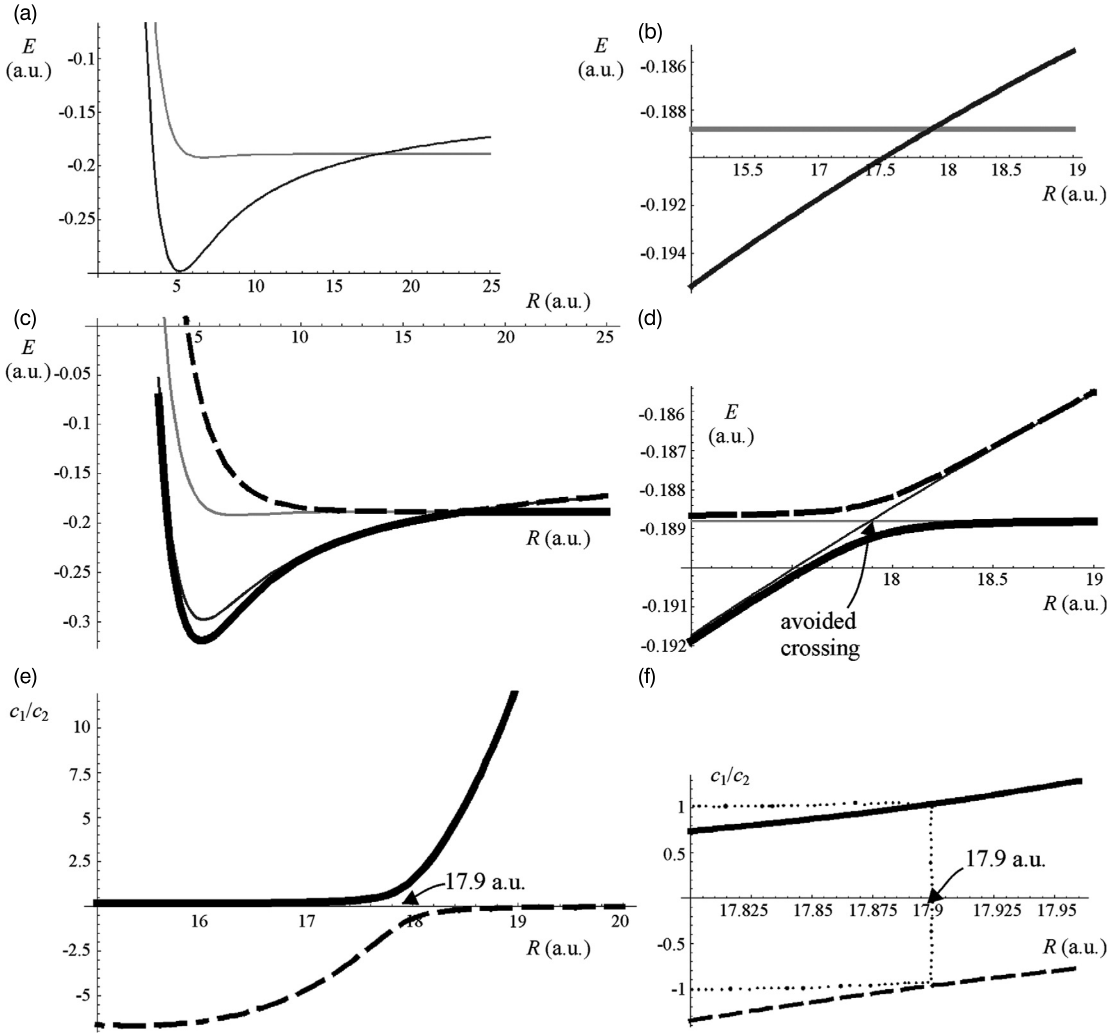


Fig. 6.21. A simple one-electron model of electron transfer in the NaCl molecule. (a) The mean values of the Hamiltonian with two *diabatic* states: one (light gray) being the $3s$ atomic orbital of the sodium atom (atomic curve), the second (dark gray) being the $2p$ atomic orbital of the chlorine atom (ionic curve). The two diabatic curves intersect. (b) A closer view of the intersection. (c) The two diabatic curves (gray, as in (a,b)) and the two *adiabatic* curves (black), the lower-energy (solid) and the higher-energy (dashed). (d) Although the drawing looks like intersection, in fact the adiabatic curves "repel" each other. (e) Each of the adiabatic states is a linear combination of two diabatic states (atomic and ionic). The ratio c_1/c_2 of the coefficients for the lower-energy (solid line) and higher-energy states (dashed line) is shown, where c_1 is the contribution of the atomic function and c_2 of the ionic function. As we can see, the lower-energy (higher-energy) adiabatic state is definitely atomic (ionic) for $R > 17.9$ a.u. and definitely ionic (atomic) for smaller R in the vicinity of the avoided crossing. (f) The ratio c_1/c_2 very close to the avoided crossing point. As we can see, at this point one of the adiabatic states is the sum and the other the difference of the two diabatic states.

6.14 Polyatomic molecules and the conical intersection

Crossing for polyatomics

The noncrossing rule for a diatomic molecule has been based on Eq. (6.48). To achieve the crossing we had to make vanish two independent terms with only one parameter (the internuclear distance R) to vary. It is important to note that in the case of a polyatomic molecule *the formula would be the same*, but the number of parameters would be larger: $3M - 6$ in a molecule with M nuclei. For $M = 3$ one has already, therefore, three such parameters. No doubt even for a three-atom molecule we would be able to make the two terms equal to zero and, therefore, achieve $E_+ = E_-$, i.e., crossing of the two diabatic hypersurfaces would occur.

Let us investigate this possibility, which, for reasons that will become clear later, is called conical intersection. We will approach this concept by a few steps.

Cartesian system of $3M$ coordinates (O_{3M})

All the quantities in Eq. (6.48) depend on $n = 3M - 6$ coordinates of the nuclei. These coordinates may be chosen in many different ways; the only thing we should bother ourselves with is that they have to determine the positions of M point objects. Just to begin, let us construct a Cartesian system of $3M$ coordinates ($\mathbf{O}_{3M}$). Let us locate (Fig. 6.22) nucleus 1 at the origin (in this way we eliminate three degrees of freedom connected with the translation of the system); nucleus 2 will occupy the point x_2 on the x axis, i.e., $y_2 = z_2 = 0$. In this way we have eliminated two rotations of the system. The total system may still be rotated about the x axis. This last possibility can be eliminated when we decide to locate nucleus 3 on the xy plane (i.e., the coordinate $z_3 = 0$).

Thus, six degrees of freedom have been eliminated from the $3M$ coordinates. The other nuclei may be indicated by vectors (x_i, y_i, z_i) for $i = 4, 5, \ldots M$. As we can see there has been a lot of arbitrariness in these choices.[73]

Cartesian system of $3M - 6$ coordinates (O_{3M-6})

This choice of coordinate system may be viewed a little differently. We may construct a Cartesian coordinate system with $3M - 6 = n$ axes, which may be labeled (in the sequence given above) in a uniform way: $\bar{x}_i$, $i = 1, 2, ...n$. A single point $\mathbf{R} = (\bar{x}_1, \bar{x}_2, ...\bar{x}_{3M-6})$ in this n-dimensional space determines the positions of all M nuclei of the system.

---

[73] By the way, if the molecule were diatomic, the third rotation would not need to be determined and the number of variables would be equal to $n = 3 \times 2 - 5 = 1$.

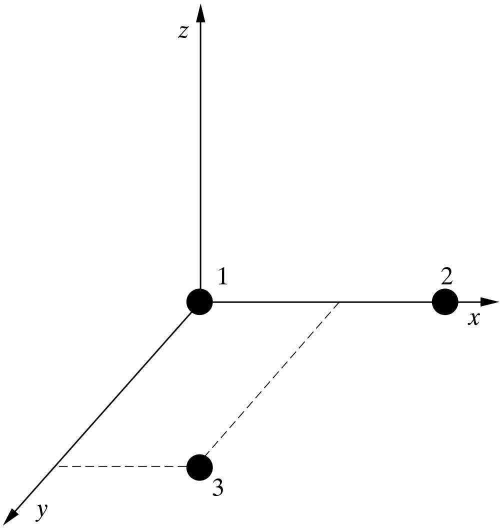


Fig. 6.22. The Cartesian coordinate system $\mathbf{O}_{3M}$ and atoms 1, 2, 3 with their fixed positions.

Two special vectors in the O_{3M-6} space

Let us consider two functions $\bar{E}_1 - \bar{E}_2$ and V_{12} of the configuration of the nuclei $\mathbf{R} = (\bar{x}_1, \bar{x}_2, ...\bar{x}_{3M-6})$, i.e., with domain being the $\mathbf{O}_{3M-6}$ space. Now, let us construct two vectors (gradients) in $\mathbf{O}_{3M-6}$, i.e.,

$$\nabla(\bar{E}_1 - \bar{E}_2) = \sum_{i=1}^{3M-6} \mathbf{i}_i \left(\frac{\partial(\bar{E}_1 - \bar{E}_2)}{\partial \bar{x}_i} \right)_0,$$

$$\nabla V_{12} = \sum_{i=1}^{3M-6} \mathbf{i}_i \left(\frac{\partial V_{12}}{\partial \bar{x}_i} \right)_0,$$

where $\mathbf{i}_i$ stands for the unit vector along axis $\bar{x}_i$, while the derivatives are calculated in a point of the configurational space for which

$$\sqrt{\left(\frac{\bar{E}_1 - \bar{E}_2}{2}\right)^2 + |V_{12}|^2} = 0$$

i.e., where according to Eq. (6.48) one has the intersection of the adiabatic hypersurfaces.

6.14.1 Branching space and seam space

We may introduce any coordinate system. We are free to do this because our object (molecule) stays immobile, but our way of determining the nuclear coordinates changes. We will change the coordinate system in n-dimensional space once more. This new coordinate system is formed from the old one ($\mathbf{O}_{3M-6}$) by rotation.

> The rotation will be done in such a way as to make the plane determined by the two first axes ($\bar{x}_1$ and $\bar{x}_2$) of the old coordinate system coincide with the plane determined by the two vectors $\nabla(\bar{E}_1 - \bar{E}_2)$ and $\nabla(V_{12})$.

Let us denote the coordinates in the rotated coordinate system by $\xi_i, i = 1, 2, ..., n$. The new coordinates can, of course, be expressed as some linear combinations of the old ones, but these details need not bother us. The most important thing is that we have the axes of the coordinates ξ_1 and ξ_2, which determine the same plane as the vectors $\nabla\left(\bar{E}_1 - \bar{E}_2\right)$ and ∇V_{12}. *This plane is known as the branching space (plane). The space of all vectors* $(0.0, \xi_3 ... \xi_{3M-6})$ *is called the seam space.* The directions $\nabla\left(\bar{E}_1 - \bar{E}_2\right)$ and ∇V_{12} need not be orthogonal, although they look this way in figures shown in the literature.[74]

Now we are all set to define the conical intersection.

6.14.2 Conical intersection

Why has this slightly weird coordinate system been chosen? We see from Eq. (6.48) for E_+ and E_- that ξ_1 and ξ_2 correspond to the fastest change of the first term and the second term under the square root sign, respectively.[75]

> Any change of other coordinates (being orthogonal to the plane $\xi_1\xi_2$) does not influence the value of the square root, i.e., does not change the difference between E_+ and E_- (although it changes the values of E_+ and E_-).

Therefore, the hypersurface E_+ intersects with the hypersurface E_-, and their common part, i.e., the intersection set, is all those vectors of the n-dimensional space that fulfill the conditions $\xi_1 = 0$ and $\xi_2 = 0$. The intersection represents an $(n - 2)$-dimensional subspace of

---

[74] F. Bernardi, M. Olivucci, M.A. Robb, *Chem. Soc. Rev.*, (1996)321. These experts in the domain of conical intersections confirmed to me that the angle is often quite small.

[75] Let us take a scalar field V and calculate its value at the point $\mathbf{r}_0 + \mathbf{r}$, where we assume $|\mathbf{r}| \ll 1$. From the Taylor expansion we have with good accuracy $V(\mathbf{r}_0 + \mathbf{r}) = V(\mathbf{r}_0) + (\nabla V)_{\mathbf{r}=\mathbf{r}_0} \cdot \mathbf{r} = V(\mathbf{r}_0) + \left|(\nabla V)_{\mathbf{r}=\mathbf{r}_0}\right| \cdot r \cos\theta$. We obtain the largest absolute value of the increment of V for $\theta = 0$ and $\theta = 180°$, i.e., along the vector $(\nabla V)_{\mathbf{r}=\mathbf{r}_0}$.

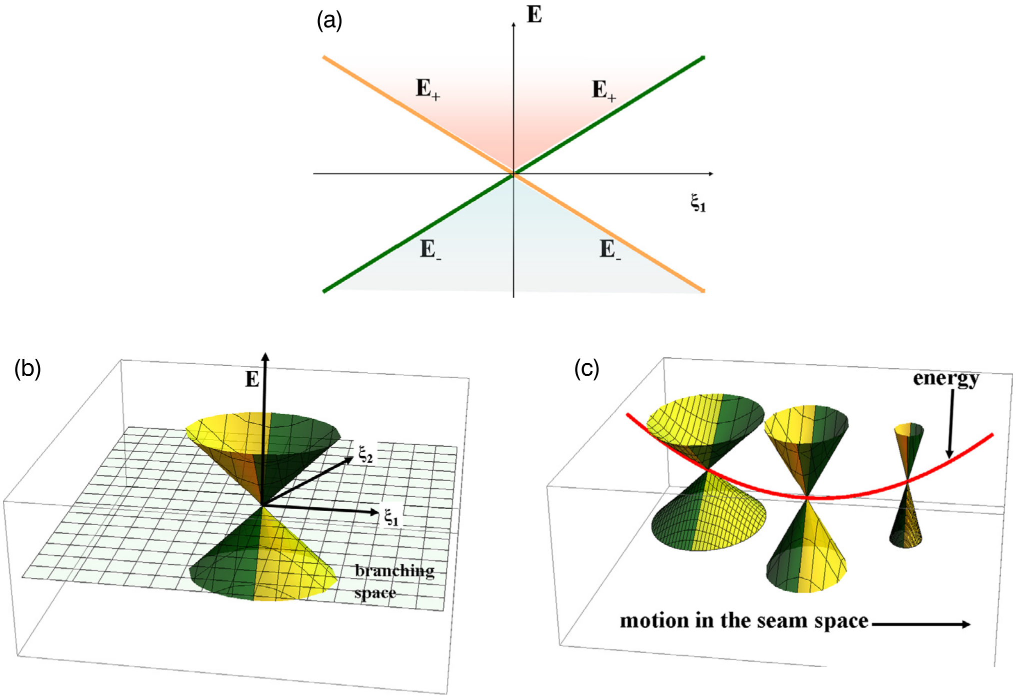


Fig. 6.23. Conical intersection (scheme). E represents the electronic energy as a function of coordinates of the nuclei: $\xi_1, \xi_2, \xi_3, \xi_4, \ldots\xi_{3M-6}$. The figure shows only the coordinates ξ_1 and ξ_2; they define what is known as the branching space $\xi_1\xi_2$, while the space of all vectors $(0, 0, \xi_3, \xi_4, \ldots\xi_{3M-6})$ known as the seam space is not shown in (a) and (b). (a) Section of the cones along ξ_1 at a given point of the seam space; the equality $E_+ = E_-$ holds for the conical intersection point (origin). (b) The vectors $\nabla(\bar{E}_1 - \bar{E}_2)$ and ∇V_{12} span the branching space (the horizontal plane, both vectors are calculated at the conical intersection). The upper cone E_+ and the lower cone E_- correspond to Eq. (6.48); each consists of two diabatic surfaces (differing by color). (c) Staying at the branching point $(0, 0)$, but moving in the seam space, one remains all the time in the conical intersection, but the cones look differently (different cone opening) and the energy $E_+ = E_-$ changes (solid line).

the n-dimensional space of the nuclear configurations.[76] When we withdraw from the point $(0, 0, \xi_3, \xi_4, \ldots \xi_{3M-6})$ by changing the coordinates ξ_1 and/or ξ_2, a difference between E_+ and E_- appears. For small increments $d\xi_1$ the changes in the energies E_+ and E_- are proportional to $d\xi_1$ and E_+ and E_- differ in sign. This means that the hypersurfaces E_+ and E_- as functions of ξ_1 (at $\xi_2 = 0$ and fixed other coordinates) have the shapes shown in Fig. 6.23a. For ξ_2 the situation is similar, but the cone may differ by its angle. From this it follows that

---

[76] If the axes ξ_1 and ξ_2 were chosen in another way *on the plane* determined by the vectors $\nabla\left(\bar{E}_1 - \bar{E}_2\right)$ and ∇V_{12}, the conical intersection would be described in a similar simple way. If, however, the axes were chosen outside the plane, it might happen that moving along *more than just two axes* they would split into E_+ and E_-. Our choice stresses that the intersection of E_+ and E_- represents an $(n-2)$-dimensional subspace (seam space).

> the ground and excited state hypersurfaces intersect with each other (at any seam space point) and split when we move within the branching space according to the *cone rule*, i.e., E_+ and E_- change *linearly* when moving in the $\xi_1\xi_2$ plane from the point $(0, 0)$.

This point is called the *conical intersection* (Fig. 6.23b). The cone opening angle is in general different for different points of the seam space (see Fig. 6.23c).

The conical intersection plays a fundamental role in the theory of chemical reactions (Chapter V2-6). The lower (ground-state) as well as the higher (excited-state) hypersurfaces are composed of two diabatic parts (Fig. 6.23b), which in polyatomics correspond to *different patterns of chemical bonds*. This means that the system (represented by a point) when moving on the ground-state adiabatic hypersurface towards the join of the two parts passes near the conical intersection point, over the energy barrier, and goes to the products. This is the essence of a chemical reaction.

6.14.3 Berry phase

We will focus on the adiabatic wave functions close to the conical intersection. Our goal will be to show something strange, i.e., that

> when going around the conical intersection point in the configurational space, the electronic wave function changes its phase, and after coming back to the starting point this change results in the opposite sign of the function.

First let us prepare an itinerary in the configuration space around the conical intersection. We need a parameter, which will be an angle α, and will define our position during our trip around the point. Let us introduce some abbreviations in Eq. (6.48), i.e., $\Delta \equiv \frac{\bar{E}_1 - \bar{E}_2}{2}$, $h \equiv V_{12}$, and define α in the following way:

$$\sin\alpha = \Delta/\rho,$$
$$\cos\alpha = h/\rho,$$

where $\rho = \sqrt{\Delta^2 + h^2}$.

We will move around the conical intersection within the plane given by the vectors $\nabla\Delta$ and ∇h. The conical intersection point is defined by $|\nabla\Delta| = |\nabla h| = 0$. Changing α from 0 to 2π we have to go, at a distance $\rho(\alpha)$, once through a maximum of h (say, in the direction of the maximum gradient ∇h) and once through its minimum $-h$ (the opposite direction). This is ensured by $\cos\alpha = h/\rho$. Similarly, we have a single maximum and a single minimum of $\nabla\Delta$ (as must be when going around), when assuming $\sin\alpha = \Delta/\rho$. We do not need more information about our itinerary because what we are interested in is how the wave function changes after making a complete trip (i.e., 360°) around the conical intersection and returning to the starting point.

The adiabatic energies are given in Eq. (6.48) and the corresponding coefficients of the diabatic states are reported in Appendix D (the first, most general case):

$$\left(\frac{c_1}{c_2}\right)_{\pm} = \frac{1}{h}\left[\Delta \pm \sqrt{\Delta^2 + h^2}\right] = \tan\alpha \pm \frac{1}{\cos\alpha}.$$

Thus,

$$\frac{c_{1,+}}{c_{2,+}} = \frac{\sin\alpha + 1}{\cos\alpha} = \frac{\left(\sin\frac{\alpha}{2} + \cos\frac{\alpha}{2}\right)^2}{\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}} = \frac{\left(\sin\frac{\alpha}{2} + \cos\frac{\alpha}{2}\right)}{\left(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\right)},$$

$$\frac{c_{1,-}}{c_{2,-}} = \frac{\sin\alpha - 1}{\cos\alpha} = \frac{-\left(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\right)^2}{\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}} = -\frac{\left(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\right)}{\left(\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\right)}.$$

To specify the coefficients in $\psi_+ = c_{1,+}\psi_1 + c_{2,+}\psi_2$ and $\psi_- = c_{1,-}\psi_1 + c_{2,-}\psi_2$ with ψ_1 and ψ_2 denoting the diabatic states, we have to take the two normalization conditions into account, i.e., $c_{1,+}^2 + c_{2,+}^2 = 1$, $c_{1,-}^2 + c_{2,-}^2 = 1$, and the orthogonality of ψ_+ and $\psi_- : c_{1,+}c_{1,-} + c_{2,+}c_{2,-} = 0$. After a little algebra we get

$$c_{1,+} = \frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\right),$$

$$c_{2,+} = \frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\right),$$

$$c_{1,-} = -\frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2}\right),$$

$$c_{2,-} = \frac{1}{\sqrt{2}}\left(\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\right).$$

Now, let us make this journey by considering the wave functions ψ_+ and ψ_- at the angle α and at the angle $\alpha + 2\pi$. Note that $\cos\frac{\alpha+2\pi}{2} = \cos\left(\frac{\alpha}{2} + \pi\right) = -\cos\frac{\alpha}{2}$ and $\sin\frac{\alpha+2\pi}{2} = \sin\left(\frac{\alpha}{2} + \pi\right) = -\sin\frac{\alpha}{2}$. Therefore, both the electronic functions ψ_+ and ψ_- have to change their signs after the journey ("geometric" phase or Berry phase), i.e.,

$$\psi_+(\alpha + 2\pi) = -\psi_+(\alpha)$$

and

$$\psi_{-}(\alpha + 2\pi) = -\psi_{-}(\alpha).$$

This is how the conical intersection is usually detected.

> Since the total wave function has to be single-valued, this means the function that describes the motion of the nuclei (and multiplies the electronic function) has to compensate for that change and has to undergo a change of sign.

6.14.4 The role of the conical intersection – nonradiative transitions and photochemical reactions

The conical intersection was underestimated for a long time. However, photochemistry demonstrated that it happens much more frequently than expected (Fig. 6.24).

Let us assume that the nuclei in the electronic ground state have their optimal positions – the Born–Oppenheimer configuration, characterized by point 1 in the configurational space (they vibrate in its neighborhood, but let us ignore the quantum nature of these vibrations[77]). This point corresponds to a particular chemical bond pattern (common for the whole dark surface in Fig. 6.24), which we call here “reactants.” Right upwards, in the excited electronic state, we have the same configuration of the nuclei, but the chemical bond pattern is different (characteristic of the whole white surface in the figure) and they are called “products.” Thus, in point 2 we have a structure that has the geometry of reactants, but its electronic structure resembles products; we call this unstable configuration (note its slope) “products’ image.” The situation is inverted on the right-hand side of Fig. 6.24. Over there the electronic ground state has a character of the products, with the corresponding energy minimum (another ground-state Born–Oppeneheimer structure). Right upwards, on the adiabatic excited-state surface, the electronic structure is called “reactants’ image,” because it has the nuclear configuration of the products, but its electronic structure resembles reactants.

Point 2 (FC) in the figure shows the very essence of the Franck–Condon rule – a vertical transition due to the laser excitation.

---

[77] Electronic energy hypersurfaces represent the PES for the motion of the nuclei. In the quantum mechanical picture only some energies will be allowed: we will have the vibrational and rotational energy levels, as for diatomics. The same energy levels corresponding to E_{+} may be close in the energy scale to those of E_{-}. Moreover, it may happen that the vibrational wave functions of two such levels may overlap significantly in space, which means that there is a significant probability that the system will undergo a transition from one to the other vibrational state. In short, in the quantum mechanical picture, the motion of the system is not necessarily bound to a single PES, but the two PESs are quite penetrable.

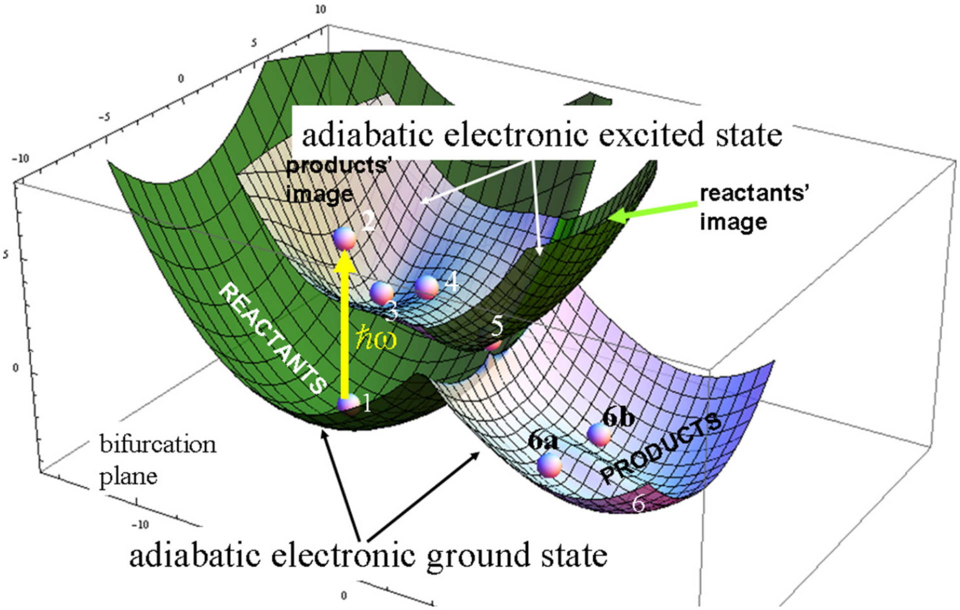


Fig. 6.24. Nonradiative transitions explained by the photochemical funnel effect (related to the conical intersection). The figure shows the electronic energy as a function of the coordinates ξ_1 and ξ_2 within the branching space. There are two adiabatic surfaces: the lower one corresponds to the ground electronic state (E_-), the upper one pertains to the excited electronic state (E_+). Each of the surfaces is composed of two parts corresponding originally to the diabatic states: the darker corresponds to the electronic structure of the reactants, the lighter to that of the products. The spheres indicate some particular configuration of the nuclei (in an electronic state). Sphere 1 indicates the reactants' configuration (R), the arrow symbolizes a photoexcitation by absorption of a photon with the appropriate energy $h\nu = \hbar\omega$. The excitation takes place instantaneously at a fixed reactants' configuration R (Franck–Condon rule). Note, that the electronic state corresponds already to the products ("products' image") and the forces acting on the nuclei correspond to the excited surface slope at the point labeled as FC (sphere 2). The forces make the system move towards the minimum (sphere 3). If the kinetic energy acquired is large enough to overcome the barrier (sphere 4), the system enters the funnel, inevitably reaches the conical intersection point (sphere 5), and in a radiationless process begins moving on the ground-state adiabatic hypersurface. The system may end up at different products (spheres 6a and 6b) or may go back to the configuration of the reactants (sphere 1). Note that right above in the excited adiabatic state we have the electronic structure, which we call the "reactants' image."

The change of electronic state takes place so fast that the nuclei do not have enough time to move. Thus the positions of the nuclei in the excited state are identical to those in the ground state (Franck–Condon rule).

The corresponding nuclear configuration may differ quite significantly from the nearest potential energy minimum (point 3) in the excited-state PES. In a few or a few tens of femtoseconds, the system slides down from FC to the neighborhood of point 3, transforming its potential energy into kinetic energy. Usually point 3 is separated from the conical intersection configuration C by a barrier with the corresponding potential energy saddle point 4 ("transition state"). Behind the saddle point there is usually an energy valley[78] with a deep funnel ending in the conical intersection configuration (point 5). As soon as the system overcomes the barrier at the transition state (4), by going over it or by tunneling, it will be sucked in by the conical intersection attractor with almost 100% probability.

> The system goes through the "funnel" to the electronic ground-state hypersurface with probability 1.

Then the system will continue its path in the ground-state PES, E_-, going either towards products 6*a* or 6*b*, or going back to point 1 (nonreactive path).

Of course, the total energy has to be conserved. The nonradiative process described will take place if the system finds a way to dissipate its energy, i.e., to transfer an excess of electronic energy into the vibrational, rotational, and translational degrees of freedom of its own or neighboring molecules (e.g., of the solvent).[79]

6.14.5 What is the number of conical intersections?

We may ask whether we will find some other conical intersections in the ground-state PES. In general the answer is positive. There are at least two reasons for this.

> In the simplest case the conical intersection represents the dilemma of an atom C (approaching molecule AB): to attach either to A or B?

Thus any encounter of three atoms causes a conical intersection, and of course such conical intersections differ usually by energy. In *each* case the important thing is a configuration of nuclei, where a small variation may lead to *distinct sets of chemical bonds*. Similar "pivot points" may happen for four, five, six, etc., atoms as well. Thus, we will encounter not only the minima, maxima, and saddle points, but also the conical intersection points when traveling in the ground-state PES (Fig. 6.25).

---

[78] On the excited-state PES.

[79] The energy is usually distributed among the degrees of freedom in an unequal way.

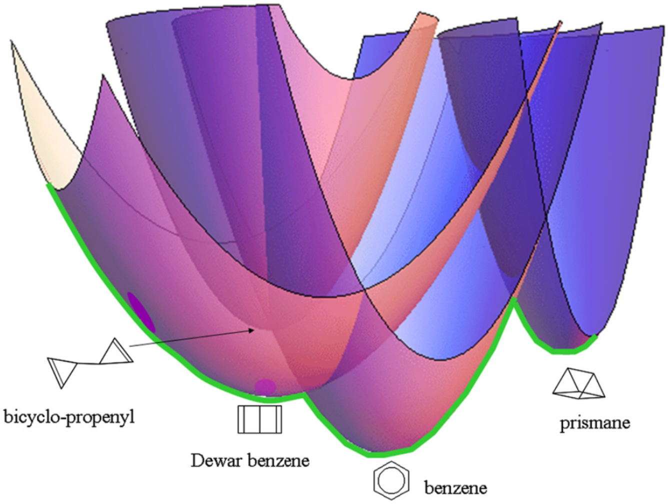


Fig. 6.25. An idea of what the Born–Oppenheimer structures represent, an example of C_6H_6. In this schematic presentation only four idealized diabatic PESs are shown (in the form of paraboloids); from left to right: bicyclopropenyl, one of the three Dewar benzene structures, benzene, and prismane. These diabatic PESs intersect and produce several adiabatic electronic states. The lowest-energy adiabatic PES (highlighted by the thick bottom line) represents the electronic ground-state PES. The ground state is essentially the familiar benzene structure (the lowest minimum). Note, however, that for sufficiently displaced nuclei the electronic ground state for C_6H_6 may resemble other diabatic structures like the Dewar one, etc. Higher in the energy scale are the excited-state PESs.

The purpose of the figure is uniquely to pass the main idea of PESs for any system. Fig. 6.25 mimics four diabatic surfaces for C_6H_6 (depicted as paraboloids), each preserving its chemical (i.e., electronic) character, when the nuclear coordinates deviate from the minima of the electronic energy. These minima define the corresponding Born–Oppenheimer structures and the figure shows only four of these isomers: bicyclo-propenyl, one of the three Dewar benzene structures, benzene (the result of the two familiar Kekulé structures[80]), and prismane. These diabatic PESs exhibit numerous conical intersections (which means mixing of the diabatic states); the latter lead to numerous adiabatic PESs: the ground electronic state and several excited states. The figure stresses the dominance of the Kekulé structure in the electronic ground state (the central lower part of the figure). Note, however, that when allowing some displacements of the nuclei, the electronic structure – still being in the ground electronic state – may resemble either of the other structures (i.e., the Dewar one, the bicyclo-propenyl, the prismane, or other isomers). This is why in some chemical neighborhoods these “exotic benzenes” may

---

[80] Despite a lot of effort made, until now any experimental and theoretical attempts to separate the two Kekulé structures as two distinct Born–Oppenheimer minima failed.

be seen. Also, Fig. 6.25 indicates why some excited states may also be similar to one of those isomeric forms, e.g., to the Dewar form in the figure.

The second reason for having many conical intersections is the permutational symmetry. Very often the system contains the same kinds of nuclei. Any exchange of positions of such nuclei moves the point representing the system in configuration space to some distant regions, while the energy does not change at all. Therefore, any PES has to exhibit the corresponding permutational symmetry. All the details of PES will repeat $M!$ times for a system with M identical nuclei. This will multiply the number of conical intersections.

More about conical intersection will be given in Chapter V2-6, when we will be equipped with the theoretical tools to describe how the electronic structure changes during chemical reactions.

6.15 A travel beyond the adiabatic approximation

6.15.1 Vibronic coupling

In polyatomic molecules a diabatic state represents a product of an electronic wave function $\psi_i^{(\Gamma_1)}$ and a rovibrational function[81] $f_v^{(\Gamma_2)}$, i.e., a *rovibronic state*

$$\psi_i^{(\Gamma_1)}(\mathbf{r};\mathbf{R})\, f_v^{(\Gamma_2)}(\mathbf{R}), \tag{6.52}$$

where the upper indices are related to the irreducible representations of the symmetry group of the clamped nuclei Hamiltonian the functions belong to (i.e., according to which the corresponding functions transform, cf. Appendix C, p. 605). If one considers the electronic and the vibrational states only[82] the product (6.52) denotes a *vibronic state*. The product function transforms according to the direct product representation $\Gamma_1 \times \Gamma_2$.

If one is interested in those solutions of the Schrödinger equation which belong to the irreducible representation Γ, the function $\psi_i^{(\Gamma_1)}(\mathbf{r};\mathbf{R})\, f_v^{(\Gamma_2)}(\mathbf{R})$ is useful as a basis function only if $\Gamma_1 \times \Gamma_2$ contains Γ. For the same reason another basis function may be useful, $\psi_{i'}^{(\Gamma_3)}(\mathbf{r};\mathbf{R})\, f_{v'}^{(\Gamma_4)}(\mathbf{R})$, as well as other similar functions:

$$\psi^{(\Gamma)} = c_{i1}\psi_i^{(\Gamma_1)}(\mathbf{r};\mathbf{R})\, f_1^{(\Gamma_2)}(\mathbf{R}) + c_{i'2}\psi_{i'}^{(\Gamma_3)}(\mathbf{r};\mathbf{R})\, f_2^{(\Gamma_4)}(\mathbf{R}) + \ldots. \tag{6.53}$$

If, say, coefficients c_{i1} and $c_{i'2}$ are large, an effective superposition of two vibronic states is taking place, which means what is known as vibronic coupling.

---

[81] That is, describing rotations and vibrations of the molecule.

[82] We skip here for the sake of simplicity the rotational wave function.

We are, therefore, beyond the adiabatic approximation (which requires a single vibronic state, a product function) and the very notion of the single potential energy hypersurface for the motion of the nuclei becomes irrelevant. In the adiabatic approximation the electronic wave function is computed from Eq. (6.8) with the clamped nuclei Hamiltonian, i.e., the electronic wave function does not depend on what the nuclei are *doing*, but only on where they *are*. In other words the electronic structure is determined (by finding a suitable $\psi_i^{(\Gamma_1)}(\mathbf{r};\mathbf{R})$ through solution of the Schrödinger equation) at a fixed position $\mathbf{R}$ of the nuclei. This implies that in this approximation the electrons have *always enough time* to adjust themselves to any instantaneous position of the nuclei. One may say that in a sense the electrons and the nuclei are perfectly correlated in their motion: electrons follow the nuclei. Therefore,

> a nonadiabatic behavior (or vibronic coupling) means a weakening of this perfect correlation, which is equivalent to saying that it may happen that the electrons do not have enough time to follow a (too fast) motion of the nuclei.

This weakening is usually allowed by taking a linear combination (6.53), which may be thought of as a kind of frustration for electrons *which* vibration ("type of motion") of the nuclei to follow. If $\Gamma_1 \neq \Gamma_3$ and $\Gamma_2 \neq \Gamma_4$ we have to do with such an electronic state which resembles $\psi_i^{(\Gamma_1)}$ when the molecule participates in a vibration of symmetry Γ_2, and resembles $\psi_{i'}^{(\Gamma_3)}$ when the molecule vibrates according to Γ_4.

This idea may be illustrated by the following examples.

Example 1 (Dipole-bound electron). Imagine a molecular dipole; one may think of it as having a + and a − pole. We are interested in its + pole, because now we consider an extra electron, which will be bound with the dipole by the + pole–electron attraction. Obviously, such an attraction should depend on the dipole moment of the dipolar molecule. How strong must a point dipole be to be able to bind an electron? This question has been asked before; the answer[83] is that this happens for the point-like dipole moment larger than[84] 1.625 D. If the dipole itself represents an electronic closed shell molecule the extra electron is usually very far (Fig. 6.26a), even at distances of the order of 50 Å!

Now imagine the dipole starts to rotate (Fig. 6.26b). At small angular momentum the electron supposedly does not have any problem with following the motion of the positive pole. For larger angular momenta the electron speeds up, its distance to the dipole increases due to the

---

[83] E. Fermi, E. Teller, *Phys. Rev.*, 47(1947)399.

[84] For nonpoint-like dipoles one may expect this limiting value to be less important, since the essence of the problem is binding an electron by a positive charge. This, however, happens even for marginally small positive charge (see the hydrogen-like atom).

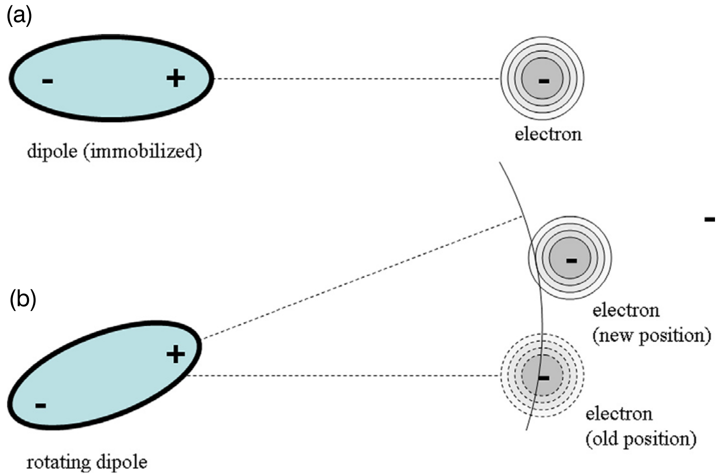


Fig. 6.26. A strange situation: an electron is unable to follow the motion of the nuclei (we are beyond the adiabatic approximation, a nonadiabatic case). (a) Some molecular dipoles with sufficiently large dipole moment may bind an extra electron (a cloud on the right), which in such a case is far away from the dipole and is attracted by its "+" pole. The positive pole plays the role of a pseudonucleus for the extra electron. (b) When the dipole starts to rotate (a state with nonzero angular momentum) the electron follows the motion of the "+" pole. This is, however, difficult for high angular momenta (the electron has not enough time to adapt its position right towards the "+" pole) and even harder because the centrifugal force pushes the extra electron further off.

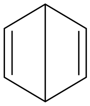
centrifugal force, and, when this happens, it gets harder and harder to follow the motion of the positive pole. The electron does not have enough time. This means a larger and larger nonadiabatic correction.

In terms of Eq. (6.53) one has to use more than one rovibronic state $\psi_i^{(\Gamma_1)}(\mathbf{r};\mathbf{R})\, f_v^{(\Gamma_2)}(\mathbf{R})$, those ones that couple the rotational states $f_v^{(\Gamma_2)}(\mathbf{R})$ of the molecular dipole with the electronic states. They should be designed with not only the extra electron sitting at a certain distance right face-to-face to the positive pole, but also including those with other distances and with different orientations with respect to the dipole.

Example 2 (Hydrogen molecule). Let us form two diabatic states, i.e., $\psi_i^{(\Gamma_1)}(\mathbf{r};R)$, corresponding to the double occupation of the bonding orbital $1s_a + 1s_b$, and $\psi_{i'}^{(\Gamma_3)}(\mathbf{r};R)$, corresponding to the double occupation of the bonding excited orbital $2s_a + 2s_b$. In this case we will take $\Gamma_3 = \Gamma_1$ (it does not mean the f functions are the same). The rovibrational function will be taken as (we assume the vibrational and rotational ground state) $f_1^{(\Gamma_2)}(\mathbf{R}) = \chi_1(R)\, Y_0^0(\theta,\phi) = \chi_1(R)$, but $f_2^{(\Gamma_4)}(\mathbf{R}) = \chi_2(R)\, Y_0^0(\theta,\phi) = \chi_2(R)$. The vibrational ground state $\chi_1(R)$ has its maximum at $R = R_{1s}$, where the minimum of the potential energy curve $E_{1s}(R)$ is, while $\chi_2(R)$ has its maximum at $R = R_{2s} > R_{1s}$, where the potential energy curve $E_{2s}(R)$ exhibits the minimum. The mixing coefficients c_{i1} and c_{i2} will obviously depend on R. For $R = R_{1s}$ we will have $c_{i1} \gg c_{i2}$, because the ground-state bonding orbital will describe well the electronic charge distribution and for this R the $2s_a + 2s_b$ orbital will have a very high energy (the size of the $2s$ orbitals does not fit the distance). However, when R increases the energy corresponding to $1s_a + 1s_b$ will increase, while the energy corresponding $2s_a + 2s_b$ will decrease (because of better fitting). This will result in a more important value of $|c_{i2}|$ and a slightly smaller value of $|c_{i1}|$ than was the case for $R = R_{1s}$. There is therefore a coupling of vibration with the electronic state – a vibronic coupling.

Example 3 (Harpooning effect). The harpooning effect from p. 365 represents also an example of a vibronic coupling, if the two diabatic states, the ionic one $\psi_i^{(\Gamma_1)}(\mathbf{r}; R)$ and the neutral one $\psi_{i'}^{(\Gamma_3)}(\mathbf{r}; R)$, are considered with their corresponding vibrational states.

Example 4 (Benzene). Let us consider a benzene molecule. Already a long time ago chemists realized that all CC bonds in this molecule are equivalent (some quantum chemical arguments for this view are presented on p. 193). The benzene molecule does not represent a static hexagonal object. The molecule undergoes $3M - 6 = 30$ vibrations (normal modes, discussed in Chapter 7). One of these modes, say, described by the vibrational wave function $f_v^{(\Gamma_2)}(\mathbf{R})$, resembles a kind of ring pulsing ("breathing") and during these vibrations the electronic wave function $\psi_i^{(\Gamma_1)}(\mathbf{r}; \mathbf{R})$ describes the six equivalent CC bonds. There is also another vibrational mode corresponding to the vibrational function $f_{v'}^{(\Gamma_4)}(\mathbf{R})$ that, in its certain phase, corresponds to shortening of the two opposite CC bonds and lengthening of the four other CC bonds. During such a motion the electronic structure changes and will correspond to what is known as the *Dewar structure*



(the shortened bonds will resemble the double bonds, the others single bonds), corresponding to the electronic wave function $\psi_{i'}^{(\Gamma_3)}(\mathbf{r}; \mathbf{R})$. There will be many more such possibilities, which is symbolized in Eq. (6.53) by "...".

The rovibrational functions $f_v^{(\Gamma_2)}(\mathbf{R})$ and $f_{v'}^{(\Gamma_4)}(\mathbf{R})$ must exhibit a strong asymmetry with respect to the equilibrium point (*vibronic anharmonicity*). Indeed, it is natural that the above-mentioned Dewar structure is energetically favored for the vibrational deviations that shorten the to-be-double CC bond and becomes unfavorable for the opposite deviations.

6.15.2 Consequences for the quest of superconductors

Superconductivity, discovered by the Dutch scholar Heike Kamerlingh Onnes in 1911, is certainly one of most fascinating phenomena. In some substances (like originally in mercury, tin, and lead) measurement of the electric conductivity as a function of lowering the temperature ended up by an abrupt decrease (below a critical temperature) of the electric resistance to zero! Such a property would be great for operating technical devices or sending electric energy over large distances. The problem is that the critical temperature turned out to be extremely low; until 1987 it was always lower than about 23 K. The situation changed after the discovery of what is now known as high-temperature superconductors (HTSs) by J. George Bednorz and K. Alex Müller in 1987. Nowadays, after discovering hundreds of new HTSs, the highest critical temperature found is equal to 250 K, i.e. as high as -23 degrees of Celsius! The discovery[85] was made in LaH_{10} under a huge pressure of 170 gigapascals. However, in virtually all cases it turned out that the HTSs have a characteristic atomic layer structure with alternating copper and oxygen atoms.

The "magic" Cu–O distance

No current theory explains properly the phenomena exhibited by the cuprate HTSs. There are several theoretical concepts, but their striking weakness is that they provide no indication as to the class of promising materials one should explore questing after HTSs. After decades of research an intriguing conclusion has been found, however, i.e., that the closer the Cu–O distance to the "magic value" $R_{Cu-O} = 1.922$ Å is, the higher is the corresponding critical temperature.[86] This remarkable correlation was virtually unnoticed by the solid state physics community for a long time.

Why on earth might such a precise criterion be supposed to work? Well, this strongly suggests that something important happens at distance $R_{Cu-O} = 1.922$ Å, but for some reason it does not happen when away from this value.

Primum non nocere...[87]

Why does a bulb emit light? Well, because the motion of the electrons in a thin wire inside the bulb meets a resistance of chaotic vibrations of the nuclei. The kinetic energy of the electrons

---

[85] M.I. Eremets et al., *Nature*, 569(2019)528.

[86] C.N.R. Rao, A.K. Ganguli, *Chem. Soc. Review*, 24(1995)1.

[87] "*First, do not harm*," a phrase attributed to the Greek physician Hippocrates (460–370 BC) as a suggested minimum for medical doctors bringing relief to a patient.

(resulting from the costly generation of electric energy) is lost partially due to collisions with the nuclei. These collisions lead to high-energy electronic states, which emit light when relaxing. In principle this is why we pay our electricity bills.

And what if nuclear motion instead of interfering would help electrons to move? Well, then the resistance would drop down, similarly as it does in superconductivity! Maybe there is something in it.

In 1993 Jeremy K. Burdett postulated some possible reason for superconductivity.[88] His hypothesis is related to a crossing of potential energy curves, the subject of our interest before a while. According to Burdett, the "magic Cu–O distance" possibly corresponds to a crossing of two close-in-energy electronic diabatic states: a diabatic state characterized by the electron holes mainly on the copper atoms and another one with the electron holes mainly on the oxygen atoms. Thus, these two states differ by the electronic charge distribution,[89] similarly as is the case for NaCl (see p. 365). However, unlike the case for NaCl, the minima, according to Burdett, do not differ much (if at all) in energy. Another important difference is that for NaCl the crossing takes place for the Na–Cl distances that are several times as large as the nearest neighbor Na–Cl distance in the crystal of the rock salt, while for the HTSs the Cu–O distance in crystals is close to the corresponding crossing point. This means that atomic vibrations may cause oscillation about the crossing point. As usually, from crossing of the diabatic curves two adiabatic states appear: the ground state with the double minimum and an excited state (Fig. 6.27).

Burdett's main point is coupling of the ground vibrational state with the two diabatic electronic states. It is during such vibrations that a dramatic change in electronic charge distribution is supposed to take place (strong *vibronic coupling*). The position of the vibrational level on the energy scale is said to be critical for superconductivity. If the position is substantially lower than the energy of the top of the barrier (b) one has to do with either of the two states localized in a given well. This corresponds to no communication between the wells and we have to do with either of the two different charge distributions (a "mixed valence" compound). If, on the other hand, the vibrational level corresponds to energy (c), much above the barrier top energy, one receives an averaged charge distribution, which does not change much during vibrations. According to Burdett the superconductivity appears when the vibrational level has the energy that is close to that of the top of the barrier (d) – this causes a strong coupling of the two diabatic electronic states through the vibrational state.

---

[88] J.K. Burdett, *Inorg. Chem.*, 32(1993)3915.

[89] Most probably the key phenomena take place in the copper–oxygen layers and may be described as a reversible reaction, $Cu^{3+}+O^{2-} \rightleftharpoons Cu^{2+\cdot}+O^{-\cdot}$ or $Cu^{2+\cdot}+O^{2-} \rightleftharpoons Cu^{+}+O^{-\cdot}$ (a dot means a unpaired electron). The presence of the unpaired electrons implies some ferro- and/or antiferromagnetic properties of these materials, which indeed have been discovered in the HTS phenomenon.

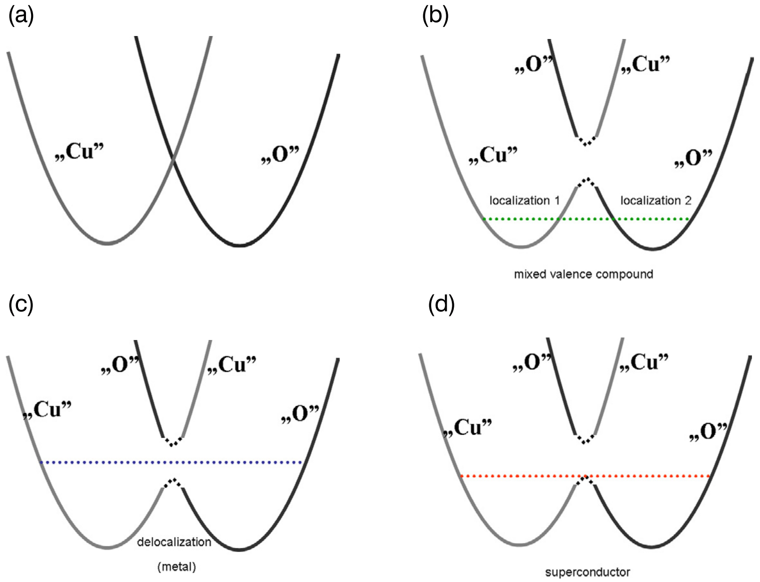


Fig. 6.27. Burdett's concept of superconductivity (schematic overview). (a) Two electronic diabatic energy curves (of comparable energies corresponding to their equilibrium positions) cross, resulting in two adiabatic energy curves (the ground and excited state, (b)–(d)). The diabatic states differ widely by the electronic charge distribution: one of them, denoted by the symbol "Cu," corresponds to the electron holes on the copper atoms, while the second one, denoted by "O," has such holes on the oxygen atoms. According to Burdett superconductivity has to do with the position of the lowest vibrational level of the ground electronic state. (b) The level is too low in energy, the vibrations are localized (either in the left- or in the right-hand side potential energy well; this is equivalent to a quasidegeneracy of the sum and difference of the delocalized vibrational states). The tunneling is marginal because of the exponential decay of the localized vibrational wave functions in the separating barrier. One has to do with an insulator in either of two coexisting states differing by the electronic charge distribution ("oxidation states"), which is known as a mixed valence compound. (c) The level is too high in energy, the vibrations are fully delocalized and proceed in the global potential energy well. The well details do not count much; one has to do with a state similar to averaging of the two states (a "metallic state"), i.e., both Cu and O have some averaged oxidation states when vibration occurs. (d) A "magic" position of the vibrational state – right at the height of the barrier. One may see this as two localized vibrational states that can tunnel easily through the barrier. The vibrations change the oxidation states of Cu and O, i.e., cause the electron transfer.

6.15.3 Photostability of proteins and DNA

How does it happen that life flourishes under the protection of the Sun, whereas it is well known our star emits some deadly radiation like charged particles and UV photons? We have two main protecting targets: Earth's magnetic field and Earth's atmosphere. Despite the atmospheric protection a substantial part of the UV radiation reaches the surface of our planet. Substances usually are not transparent for the UV, while absorption of a UV photon is often harmful for chemical bonds, hence they dissociation and/or create other bonds. This is desirable for producing vitamin D_3 in our body, but in many cases it could end up badly. For example, some substances, like DNA or proteins, have to be protected, for their destruction would destroy the basis of life. Therefore, how do these substances function so efficiently in the vibrant life processes? What is the additional target that protects them so well?

It turns out that this wonder target is the ubiquitous hydrogen bond, an important factor determining the three-dimensional shape of both DNA and proteins (see p. V2-424). The hydrogen bond $X–H\cdots Y$ has some special features also concerning its UV properties, and this holds independently of which electronegative atoms XY are involved. The explanation of the UV protection mechanism of the hydrogen bond given below comes from Sobolewski and Domcke.[90]

Fig. 6.28 shows three electronic energy hypersurfaces for the hydrogen bond $X–H\cdots Y$, visualized as sections along the proton transfer coordinate (PT) that describes the position of the proton in the hydrogen bridge. The ground state S_0 represents a diabatic state corresponding to the resonance structure $X–H\cdots Y$, in which two electrons are at X and two at Y. The energy of this singlet state has a minimum for PT close to X. An absorption of the UV photon causes a transition of the system to the lowest-energy singlet excited state (^1LE) with its electronic structure denoted as $(X–H\cdots Y)^*$. Most importantly, its energy curve intersects another singlet excited state, which corresponds to the electron transfer from X to Y (^1CT). The resulting conical intersection of the states ^1LE and ^1CT is shown as a double cone. The photon energy is large enough so the system reaches the conical intersection point and ends up on the CT energy curve, which means a single electron transfer symbolized by $(X–H)^+\cdots Y^-$. Since the minimum of the CT curve is shifted far away to the right, after going out from the conical intersection one has to do with a vibrational excited state denoted as an ionic "hot" structure: $[(X–H)^+\cdots Y^-]_{hot}$. The system slides down the CT curve, changing high potential energy to the vibrational energy and kinetic energy of the surrounding water molecules. This sliding down means at the same time moving a proton to the right; the system remains all the time in the ionic state. This however means now the radical-ionic structure of the type $X^\bullet\cdots(H–Y)^\bullet$. When sliding down the system meets the second conical intersection, which makes it possible to continue the motion on the ground-state curve. At this value of the PT one has the "hot" structure $[X^-\cdots(H–Y)^+]_{hot}$. Now the sliding down

---

[90] A. Sobolewski, W. Domcke, *Chem. Phys. Chem.*, 7(2006)561.

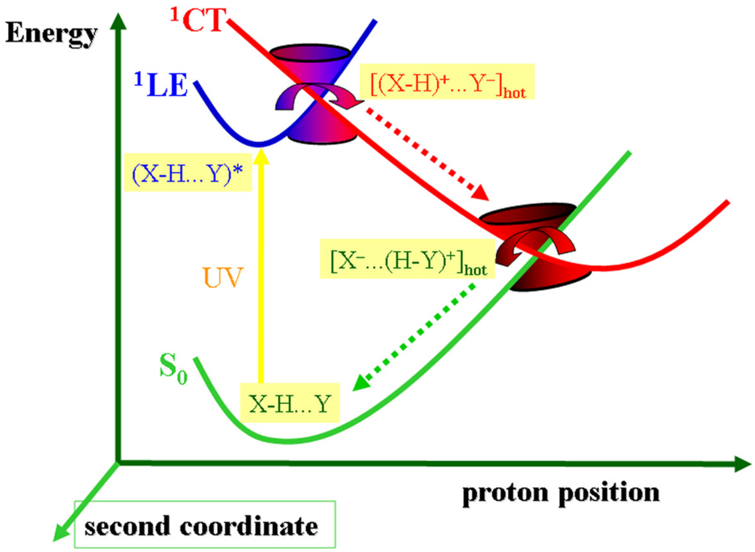


Fig. 6.28. The hydrogen bond is UV-stable because of the mobility of the bridging proton. The figure shows the electronic energy as a function of the proton position in the hydrogen bond $X\text{–}H \cdots Y$ (another coordinate that measures deviation of the proton from the XY axis is also marked). The electronic ground-state S_0 energy curve corresponds to the "four-electron" wave function corresponding to the bond pattern $X\text{–}H \cdots Y$ (one of what is known as resonance structures). A UV transition to the lowest excited singlet state $(X\text{–}H \cdots Y)^*$ (^1LE) is shown by a vertical arrow. The electronic energy curve for this state intersects (the conical intersection is shown as a double cone) a singlet diabatic state (^1CT) that corresponds to a transfer of an electron from X to Y. The excitation energy is sufficiently large to allow the system to attain the curve corresponding to the structure $(X\text{–}H)^+ \cdots Y^-$. After passing the conical intersection one has to do with a vibrationally excited state, which is symbolized by $[(X\text{–}H)^+ \cdots Y^-]_{hot}$. The proton continues its motion towards Y and the structure begins to be of the radical-ionic type: $X^\bullet \cdots (H-Y)^\bullet$. The system meets the second conical intersection, which allows it to attain the ground state. This time the proton moves towards X, while its electronic energy changes to the vibrational energy of the molecule and the surrounding water. Thus, the UV photon does not harm chemical bonds; its energy goes instead to heating the surrounding water. Courtesy of Professor A. Sobolewski.

means going left (the proton comes back) and transferring the vibrational energy to the water.

> Therefore, the net result is the following: the absorption of the UV photon, after some bouncing of the proton in the hydrogen bridge, results in heating the surrounding water, while the hydrogen bond stays in its ground state.

6.15.4 Muon-catalyzed nuclear fusion

Some molecules look really peculiar; they may contain a muon instead of an electron. The muon is an unstable particle with the charge of an electron and mass equal to 207 electronic masses.[91] For such a mass, assuming that nuclei are infinitely heavier than muons looks like a very bad approximation. Therefore, the calculations need to be nonadiabatic. The first computations for muonic molecules were performed by Kołos, Roothaan, and Sack[92] in 1960. The idea behind the project was muon-catalyzed fusion of deuterium (d) and tritium (t); the abbreviations pertain here to the nuclei only. This fascinating problem was proposed by Andrei Sakharov. Its essence is as follows.

Andrei Dimitriy Sakharov, (1921–1989) Russian physicist, father of the Soviet hydrogen bomb. During the final celebration of the H bomb project Sakharov expressed his hope that the bombs would never be used. A Soviet general answered coldly that it was not the scientists' business to decide such things. This was a turning point for Sakharov and he began his fight against the totalitarian system.
The idea of muon-induced fusion was conceived by Sakharov in 1945, in his first scientific paper, under the supervision of Tamm.

In 1957 David Jackson realized that muons may serve as catalysts.

If the electron in the molecule dt$e$ is replaced by a muon, immediately the dimension of the molecule decreases by a factor of about 200. Truly a giant nonadiabatic effect! How is this possible?

Well, the radius of the first Bohr orbit in the hydrogen atom (see p. 232) is equal to $a_0 = \frac{\hbar^2}{\mu e^2}$. After introducing atomic units, this formula becomes $a_0 = \frac{1}{\mu}$, and when we take into account that the reduced mass $\mu \approx m$ (where m stands for the electron mass) we get $a_0 \approx 1$. This approximation works for the electron, because in reality $\mu = 0.9995\ m$. If, in the hydrogen atom, instead an electron we have a muon, then μ would be equal to about 250 m. This, however, means that such a "muon Bohr radius" would be about 250 times smaller. Nuclear forces begin

---

[91] The muon was discovered in 1937 by C.D. Anderson and S.H. Neddermeyer. Its lifetime is about $2.2 \cdot 10^{-6}$ s. The muons belong to the lepton family (with the electron and τ particle, the latter with mass equal to about 3640 electronic masses). Nature created, for some unknown reasons, "more massive electrons." When the nuclear physicist Isidor Rabi was told about the incredible mass of the τ particle, he dramatically shouted: "*Who ordered that!?*"

[92] W. Kołos, C.C.J. Roothaan, R.A. Sack, *Rev. Mod. Phys.*, 32(1960)205.

to operate at such a small internuclear separation (strong interactions, Fig. 6.29a), and they are able to overcome the Coulombic barrier and stick the nuclei together by *nuclear fusion*. The muon, however, is released, and may serve as a catalyst in the next nuclear reaction.

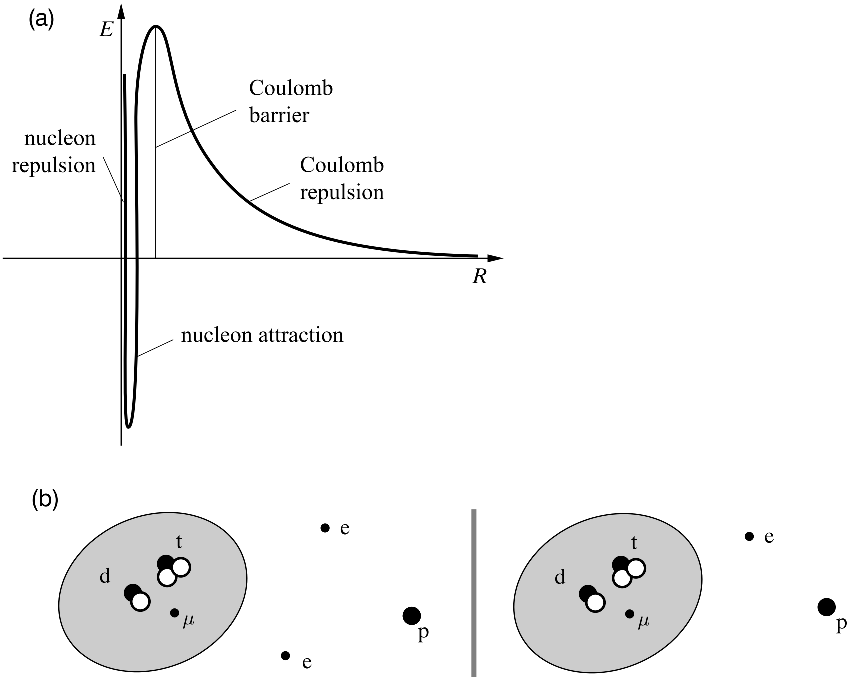


Fig. 6.29. On a relation between the electronic and nuclear structure. (a) The interaction energy potential of d and t as a function of the interparticle distance (R), taking the nuclear forces into account (an outline). At large R, of the order of nanometers, we have Coulombic repulsion; at distances of the order of femtometers the internuclear attractive forces (called the strong interaction) are switched on and overcome the Coulombic repulsion. At a distance of a fraction of a femtometer again we have a repulsion. (b) "Russian dolls" (outline): the analogs of H_2 and H_2^+.

Deuteron and tritium bound together represent a helium nucleus. One muon may participate in about 200–300 such muon-catalyzed fusion processes.[93] Everybody knows how much effort and money has been spent for decades (for the moment in vain) to ignite the nuclear synthesis $d + t \rightarrow He$. Muon-catalyzed fusion might be an alternative solution. If the muon project were successful, humanity would have access to a practically unlimited source of energy. Unfortu-

---

[93] The commercial viability of this process will not be an option unless we can demonstrate 900 fusion events for each muon. Fusion of about 10 grams of deuterium and 15 g of tritium would then be sufficient to supply the average person with electricity for life.

nately, theoretical investigations suggest that the experimental yield already achieved is about the maximum theoretical value.[94]

6.15.5 "Russian dolls" – or a molecule within molecule

Scrinzi and Szalewicz[95] carried out nonadiabatic calculations (p. 315) for a system of six particles: proton (p), deuterium (d), tritium (t), muon (μ), and two electrons (e) interacting by Coulombic forces (i.e., no nuclear forces assumed). It is not so easy to predict the structure of the system. It turned out that the resulting structure is a kind of "Russian doll"[96] (Fig. 6.29b): the muon has acted according to its mass (see above) and created tdμ with a dimension of about 0.02 Å. This system may be viewed as a partly split nucleus of charge $+1$ or, alternatively, as a mini model of the hydrogen molecular ion (scaled at 1:200). The "nucleus" serves as a partner to the proton and both create a system similar to the hydrogen molecule, in which the two electrons play their usual binding role, and the internuclear distance is about 0.7 Å. It turns out that the nonzero dimension of the "nucleus" makes a difference, and the energies computed with and without an approximation of the point-like nucleus differ. The difference is tiny (about 0.20 meV), but it is there.

It is quite remarkable that such small effects are responsible for the fate of the total system. The authors report that the relaxation of the "nucleus" dtμ (from the excited state to the ground state[97]) causes the ionization of the system: one of the electrons flies off. Such an effect, however, may excite those who study this phenomenon. How is it possible? The "nucleus" is terribly small when seen by an electron orbiting far away. How could the electron detect that the nucleus has changed its state and that it has no future in the molecule? Here, however, our intuition fails. For the electron, the most frequently visited regions of the molecule are nuclei. We will see this in Chapter 8, p. 510, but even the $1s$ state of the hydrogen atom (p. 232, the maximum of the orbital is at the nucleus) suggests the same. Therefore, no wonder the electron *could* recognize that something has abruptly changed on one of the nuclei and (being already excited) it felt receiving much more freedom, so much that it could leave the molecule.

We may pose the interesting question whether the "Russian doll" represents the global minimum of the particles system. We may imagine that the proton changes its position with the

---

[94] This has been the subject of a joint Polish-American project. More about this can be found in K. Szalewicz, S. Alexander, P. Froelich, S. Haywood, B. Jeziorski, W. Kołos, H.J. Monkhorst, A. Scrinzi, C. Stodden, A. Velenik, X. Zhao, in "*Muon Catalyzed Fusion*," eds. S.E. Jones, J. Rafelski, H.J. Monkhorst, AIP Conference Proceedings, 181(1989)254.

[95] A. Scrinzi, K. Szalewicz, *Phys. Rev. A*, 39(1989)4983.

[96] (((woman @ woman)@ woman)@).

[97] A. Scrinzi, K. Szalewicz, *Phys. Rev. A*, 39(1989)2855. The dtμ ion is created in the rovibrational state $J = 1$, $v = 1$, and then the system spontaneously goes to the lower-energy 01 or 00 states. The energy excess causes one electron to leave the system (ionization). This is an analog of the Auger effect in spectroscopy.

deuterium or tritium, i.e., new isomers (isotopomers[98]) appear. The authors did not study this question,[99] but investigated a substitution of the proton by deuterium and tritium (and obtained similar results).

Scrinzi and Szalewicz also performed some calculations for an analog of H_2^+: proton, deuterium, tritium, muon, and electron. Here the “Russian doll” looks wonderful (Fig. 6.29c); it is a four-level object:

- the molecular ion (the analog of H_2^+) is composed of *three* objects: the proton, the “split nucleus” of charge $+1$, and the electron;
- the “split nucleus” is also composed of *three* objects: d,t,μ (a *mini model of* H_2^+);
- the tritium is composed of *three* nucleons: the proton and the two neutrons;
- each of the nucleons is composed of *three* quarks (called the valence quarks).

Summary

- In the adiabatic and the Born–Oppenheimer approximations, the total wave function is taken as a *product* $\Psi = \psi_k(\mathbf{r}; R) f_k(\mathbf{R})$ of the function $f_k(\mathbf{R})$, which describes the motion of the nuclei (vibrations and rotations) and the function $\psi_k(\mathbf{r}; R)$ that pertains to the motion of electrons (and depends parametrically on the configuration of the nuclei; here we give the formulae for a diatomic molecule). This approximation relies on the fact that the nuclei are thousands of times heavier than the electrons.
- The function $\psi_k(\mathbf{r}; R)$ represents an eigenfunction of the electronic Hamiltonian $\hat{H}_0(R)$, i.e., the Hamiltonian $\hat{H}$, in which the kinetic energy operator for the nuclei is assumed to be zero (the *clamped nuclei Hamiltonian*).
- The eigenvalue of the clamped nuclei Hamiltonian depends on positions of the nuclei and in the Born–Oppenheimer approximation is mass-independent (the same for different isotopomers). This energy as a function of the configuration of the nuclei represents the potential energy for the motion of the nuclei (PES).
- The function $f_k(\mathbf{R})$ is a *product* of a spherical harmonic[100] Y_J^M that describes the *rotations* of the molecule (J and M stand for the corresponding quantum numbers) and a function that describes the *vibrations* of the nuclei.

---

[98] The situation is quite typical, although we very rarely think this way. Some people say that they observe *two different systems*, whereas others say that they see *two states of the same system*. This begins with the hydrogen atom – it looks different in its $1s$ and $3p_z$ states. We can easily distinguish two different conformations of cyclohexane, two isomers of butane, and some chemists would say these are different substances. Going much further, N_2 and CO represent two different molecules, or is one of them nothing but an excited state of the other? However strange it may sound for a chemist, N_2 represents an excited state of CO, because we may imagine a nuclear reaction of the displacement of a proton from one nitrogen to the other (and the energy curve per nucleon as a function of the atomic mass is convex). Such a point of view is better for viewing each object as a “new animal”: it enables us to see and use some relations among these animals.

[99] They focused their attention on tdμ.

[100] That is, of the eigenfunction for the rigid rotator.

- The diagram of the energy levels shown in Fig. 6.8c represents the basis of molecular spectroscopy. The diagram may be summarized in the following way:
  - The energy levels form some series separated by energy gaps, with no discrete levels. Each series corresponds to a single electronic state k, and the individual levels pertain to various vibrational and rotational states of the molecule in electronic state k.
  - Within the series for a given electronic state, there are groups of energy levels, each group characterized by a distinct vibrational quantum number ($v = 0, 1, 2, ...$), and within the group the states of higher and higher energies correspond to the increasing rotational quantum number J.
  - The energy levels fulfill some general relations:
    - increasing k corresponds to an electronic excitation of the molecule (UV-VIS);
    - increasing v pertains to a vibrational excitation of the molecule, and requires the energy to be *smaller by one or two orders of magnitude* than an electronic excitation (IR);
    - increasing J is associated with energy *smaller by one or two orders of magnitude* than a vibrational excitation (microwaves).
- Above the dissociation limit one has to do with a continuum of states of the dissociation products with kinetic energy. In such a continuum one may have also the resonance states, which may have wave functions that resemble those of stationary states, but differ from them by having finite lifetime.
- The electronic wave functions $\psi_k(\mathbf{r}; R)$ correspond to the energy eigenstates $E_k^0(R)$, which are functions of R. The energy curves[101] $E_k^0(R)$ for different electronic states k may cross each other, unless the molecule is diatomic and the two electronic states have the same symmetry.[102] In the latter case we have what is known as an avoided crossing (see Figs. 6.20 and 6.21).
- The adiabatic states represent the eigenfunctions of $\hat{H}_0(R)$. If electrons have enough time to follow the motion of the nuclei, we may apply the adiabatic function (which may change its chemical character when varying R). The diabatic states are not the eigenfunctions of $\hat{H}_0(R)$ and preserve their chemical character when changing R. If electrons are too slow to follow the nuclei, changing of R may result in keeping the same chemical character of the solution (diabatic state). In the adiabatic and in the diabatic approach the motion of the nuclei is described using a single PES.
- The nonadiabatic approach requires the use of several or many PESs when describing motion of the nuclei. The total wave function is a linear combination of the rovibronic functions with different and R-dependent amplitudes.
- For polyatomic molecules the energy hypersurfaces $E_k^0(\mathbf{R})$ can cross. The most important is *conical intersection* (Fig. 6.24) of the two diabatic hypersurfaces (I and II), i.e., those that (each individually) preserve a given pattern of chemical bonds. This intersection results in two adiabatic hypersurfaces ("lower PES" and "upper PES"). Each of the adiabatic hypersurfaces consists of two parts: one belonging to I and the second to II. Using a suitable coordinate system in the configurational space, we obtain *the adiabatic hypersurface splitting* (the difference of E_- and E_+) *when changing two coordinates* (ξ_1 *and* ξ_2) *only (the branching plane)*. The splitting begins by a linear dependence on ξ_1 and ξ_2, which gives a sort of cone (hence the name "conical intersection"). The other coordinates (the seam space) *alone* are unable to cause the splitting, although they may influence the opening angle of the cone and the energy.
- Conical intersection plays a prominent role in the photochemical reactions, because the excited molecule slides down the upper adiabatic hypersurface to the funnel (just the conical intersection

---

[101] As functions of R.

[102] That is, they transform according to the same irreducible representation.

point) and then, with a yield close to 100%, lands on the lower adiabatic hypersurface (assuming there is a mechanism for dissipation of the excess energy).

- The vibronic effects are at the basis of many important phenomena.

Main concepts, new terms

adiabatic approach (p. 315)
adiabatic approximation (p. 318)
adiabatic curve (p. 358, p. 349)
asymmetric top (p. 349)
avoided crossing (p. 363)
Berry phase (p. 371)
BFCS (p. 344)
Born–Oppenheimer approximation (p. 323)
branching plane (p. 369)
branching space (p. 369)
clamped nuclei Hamiltonian (p. 314)
conical intersection (p. 369)
continuum states (p. 353)
Coriolis energy (p. 347)
diabatic approach (p. 358)
diagonal correction for the motion of nuclei (p. 318)
dipole-bound electron (p. 379)
Eckart conditions (p. 347)
electronic energy (p. 316)
electronic Hamiltonian (p. 314)
electronic-vibrational-rotational spectroscopy (p. 332)
Franck–Condon rule (p. 374)
funnel effect (p. 374)
harpooning effect (p. 365)
HTS (p. 380)
infrared spectrum (p. 332)
microwave spectrum (p. 332)
molecular structure (p. 326)
moment of inertia (p. 349)
muon-catalyzed fusion (p. 386)
nonadiabatic approach (p. 315)
nonbound metastable states (p. 351)
nonbound states (p. 351)
noncrossing rule (p. 363)
nonradiative transitions (p. 374)
nuclear fusion (p. 386)
PES (p. 326)
photochemical reaction (p. 374)
photostability (p. 383)
potential energy curve (p. 320)
potential energy (hyper)surface (p. 320)
quasiharmonic approximation (p. 342)
RMCS (p. 344)
rotational energy (p. 332)
rotational structure (p. 332)
roto-translational coupling (p. 345)
seam space (p. 369)
SFCS (p. 311)
spherical top (p. 349)
split nucleus effect (p. 386)
superconductors (p. 381)
symmetric top (p. 349)
UV-VIS spectrum (p. 332)
vibrational energy (p. 336)
vibrational function (p. 332)
vibrational structure (p. 332)
vibronic coupling (p. 377)
vibro-translational coupling (p. 345)
vibro-rotational coupling (p. 345)
wave function “measurement” (p. 355)

From the research front

For the hydrogen molecule one may currently get a very high accuracy in predicting rovibrational levels. For example, exact analytic formulae have been derived[103] allowing one to compute the Born–Oppenheimer potential with an uncertainty smaller than 10^{-9} cm^{-1} and add the correction for the nonzero size of each nucleus (the latter correction shifts the rovibrational energy levels always by less than 10^{-4} cm^{-1}). The approach presented on p. 325 is able to produce the adiabatic diagonal correction and the nonadiabatic corrections for all rovibrational states of the ground electronic state with an accuracy better than 10^{-4} cm^{-1}. One is able to test the accuracy of not only the theory given in the present chapter, but also of its most sophisticated extensions, including quantum electrodynamics (QED). One may say that virtually for the first time QED can be confronted with the most accurate experiments beyond the traditional territory of the free electron and simple atoms (hydrogen, helium, lithium), i.e., with systems with more than one nucleus. For the hydrogen molecule one starts with an accurate solution to the Schrödinger equation[104] and then, circumventing the Dirac equation, one includes all the relativistic Breit–Pauli terms (the terms of the order of $(\frac{1}{c})^2$ and the terms of the order of $\frac{1}{c}$ vanish, where $c = \frac{1}{137.0359991}$ a.u.) and later the complete QED corrections of the order of $(\frac{1}{c})^3$ and the leading terms of $(\frac{1}{c})^4$. Just to show the accuracy achieved: for the $J = 0 \to 1$ rotational excitation the theory gives[105] 118.486812(9) cm^{-1}, while the most accurate experiment to date[106] gives 118.48684(10) cm^{-1}. Some theories trying to explain the presence of black matter need the nuclear forces operating at larger distances than they are traditionally believed to do. If these theories were true, there would be no such agreement between the theory and experiment; we would see a larger difference.

Ad futurum

The computational effort needed to calculate the PES for an M-atomic molecule is proportional to 10^{3M-6}. This strong dependence suggests that, for the next 20 years, it would be rather unrealistic to expect high-quality PES computations for $M > 10$. However, experimental chemistry offers high-precision results for molecules with hundreds of atoms. It seems inevitable that it will be possible to freeze the coordinates of many atoms. There are good reasons for such an approach, because indeed most atoms play the role of spectators in chemical processes. It may be that limiting ourselves to, say, 10 atoms will make the computation of rovibrational spectra feasible.

Additional literature

J. Hinze, A. Alijah, L. Wolniewicz, "Understanding the Adiabatic Approximation; the Accurate Data of H_2 Transferred to H_3^+," *Pol. J. Chem.*, 72(1998)1293.

---

[103] K. Pachucki, *Phys. Rev. A*, 82(2010)032509.

[104] The center of mass rests at the origin. The solution of the Schrödinger equation is numerically achieved, i.e., the nonadiabatic treatment is applied with very high and controlled accuracy.

[105] J. Komasa, K. Piszczatowski, G. Łach, M. Przybytek, B. Jeziorski, K. Pachucki, *J. Chem. Theor. Comput.*, 7(2011)3105.

[106] D.E. Jennings, S.L. Bragg, J.W. Brault, *Astrophys. J.*, 282(1984)L85. The uncertainty in parentheses is given in the units of the last digit reported.

This paper reports the derivation of the equation of motion for a *polyatomic* molecule. As the origin of the BFCS, unlike this chapter, the center of mass was chosen.[107]

W. Kołos, "Adiabatic Approximation and its Accuracy," *Advan. Quantum Chem.*, 5(1970)99.

Kołos was the No. 1 expert in the domain.

F. Bernardi, M. Olivucci, M.A. Robb, "Potential Energy Surface Crossings in Organic Photochemistry," *Chem. Soc. Rev*. p. 321–328, (1996).

A review article by the top experts in conical intersection problems.

W. Domcke, D.R. Yarkony, H. Köppel (eds.), "Conical Intersections: Electronic Structure, Dynamics and Spectroscopy," *Advanced Series in Physical Chemistry,* vol. 15, World Scientific Publishing, Singapore (2004).

A series of competent articles on the conical intersection problem.

Questions

1. The nonadiabatic theory for a diatomic ($\mathbf{r}$ denotes the electronic coordinates, $\mathbf{R}$ stands for the vector connecting nucleus b with nucleus a, $R \equiv \|\mathbf{R}\|$, N means the number of electrons, m is the electron mass, V represents the Coulombic interaction of all particles, and μ is the reduced mass of the two nuclei of masses M_a and M_b):
   a. the total wave function can be represented as $\Psi(\mathbf{r},\mathbf{R}) = \sum_k \psi_k(\mathbf{r}; R) f_k(\mathbf{R})$, where the functions ψ_k form a complete set in the Hilbert space for electrons (at a given R) and f_k are the coefficients depending on $\mathbf{R}$.
   b. in the expression $\Psi(\mathbf{r},\mathbf{R}) = \sum_k \psi_k(\mathbf{r}; R) f_k(\mathbf{R})$ the functions $f_k(\mathbf{R})$ describe rotations and vibrations of the molecule.
   c. as functions $\psi_k(\mathbf{r}; R)$ one may assume the eigenfunctions of the electronic Hamiltonian.
   d. may provide only some approximation of the solution Schrödinger equation.
2. Adiabatic approximation (for notation, see question 1):
   a. is also known as the Born–Oppenheimer approximation.
   b. the electronic Hamiltonian can be obtained from the total Hamiltonian by neglecting the kinetic energy operator for the nuclei.
   c. in the adiabatic approximation the total wave function represents a product $\psi_k(\mathbf{r};\mathbf{R}) f_k(\mathbf{R})$, where $\psi_k(\mathbf{r};\mathbf{R})$ stands for the eigenfunction of the clamped nuclei Hamiltonian for the configuration of the nuclei given by $\mathbf{R}$, while $f_k(\mathbf{R})$ denotes the wave function for the motion of the nuclei.
   d. E_k^0 as a function of $\mathbf{R}$ represents the eigenvalue of the clamped nuclei Hamiltonian that corresponds to the wave function $\psi_k(\mathbf{r};\mathbf{R})$.

---

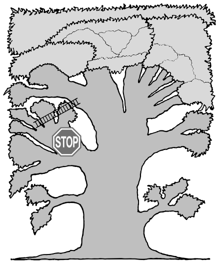
[107] We have chosen the center of the *ab* bond.

3. Consider a diatomic in the adiabatic approximation, where the origin of the coordinate system is in the geometric center of the molecule (at $\mathbf{R}/2$). The nuclei vibrate in the potential:
   a. $E_k^0(R) + J(J+1)\frac{\hbar^2}{2\mu R^2}$.
   b. $\left\langle \psi_k | \hat{H} \psi_k \right\rangle + (2J+1)\frac{\hbar^2}{2\mu R^2}$.
   c. $E_k^0(R) + H'_{kk} + J(J+1)\frac{\hbar^2}{2\mu R^2}$.
   d. $\left\langle \psi_k | \hat{H} \psi_k \right\rangle + J(J+1)\frac{\hbar^2}{2\mu R^2}$.
4. The potential energy curves for the motion of nuclei for electronic states computed by the Born–Oppenheimer approximation:
   a. may not intersect.
   b. have to intersect at an internuclear distance.
   c. cannot intersect if the corresponding eigenfunctions belong to the same irreducible representation of the symmetry group of the Hamiltonian.
   d. may intersect if the corresponding wave functions are of different symmetry.
5. The potential energy for the motion of the nuclei in the Born–Oppenheimer approximation:
   a. contains the eigenvalue of the clamped nuclei Hamiltonian.
   b. as a function of the configuration of the nuclei may exhibit many minima.
   c. contains the electronic energy.
   d. does not change after rotational excitations.
6. Due to the rotational excitation $J \to (J+1)$ of a diatomic of bond length R:
   a. one has to add to the potential energy a term proportional to $(2J+1)R^2$.
   b. the potential energy for vibrations changes.
   c. the molecule may dissociate due to the centrifugal force.
   d. the momentum of the molecule increases.
7. The adiabatic approximation:
   a. takes into account the finite mass of the nuclei.
   b. means the total wave function being a product of the electronic wave function and a wave function that describes the motion of the nuclei.
   c. as a consequence leads to the concept of a spatial shape of a molecule.
   d. is better satisfied by a molecule with muons instead of electrons.
8. Consider the basics of spectroscopy within the Born–Oppenheimer approximation:
   a. the electronic structure changes after absorbing microwaves.
   b. to excite vibrational levels (preserving the electronic state) one needs IR radiation.
   c. a red sweater looks red because of a dye that absorbs red light.
   d. microwaves can excite rotations of polar molecules.
9. At the conical intersection, the following directions in the space of the nuclear configurations cause splitting of E_+ and E_-:
   a. $\nabla(\bar{E}_1 - \bar{E}_2)$ and $\nabla(V_{12})$.
   b. $\nabla(\frac{\bar{E}_1+\bar{E}_2}{2})$ and $\nabla(V_{12})$.
   c. $\nabla(\frac{\bar{E}_1-\bar{E}_2}{2})$ and $\nabla(V_{12})$.
   d. any direction in the branching space.
10. At the conical intersection, the opening angle of the cone:
   a. equals zero.
   b. in general differs along the directions of $\nabla(\bar{E}_1 - \bar{E}_2)$ and $\nabla(V_{12})$.

c. depends on the point of the seam space.
d. is the same in the Born–Oppenheimer approximation for different isotopomers.

Answers

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



CHAPTER 7

Motion of Nuclei

> *If you are out to describe the truth, leave elegance to the tailor.*
> ***Albert Einstein***

Where are we?

We are on the most important side branch of the TREE.

An example

Which cyclohexane (C_6H_{12}) conformation (Fig. 7.1) is more stable: the "boat" or the "chair"? How do particular conformations look in detail (symmetry, interatomic distances, bond angles) when the electronic energy as a function of the positions of the nuclei attains a minimum value? How will the boat and chair conformations change if one of the hydrogen atoms is replaced by a phenyl substituent ($–C_6H_5$)? What will be the most stable conformation of the trimer $C_6H_{11}–(CH_2)_3–C_6H_{10}–(CH_2)_3–C_6H_{11}$?

What is it all about?

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00015-X

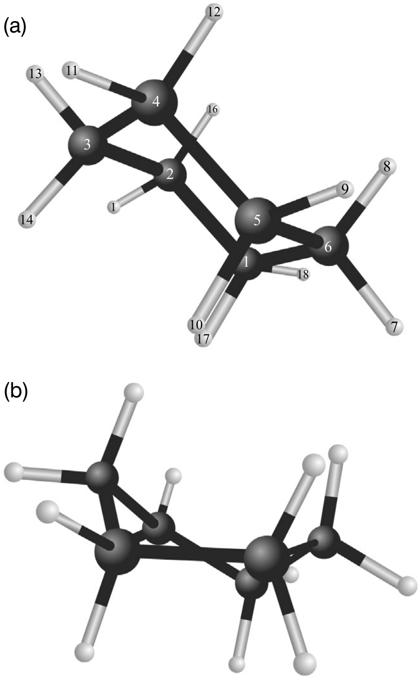


Fig. 7.1. The (a) chair and (b) boat conformations of cyclohexane. These geometries (obtained from arbitrary starting conformations) are optimized in the force field, which we will define in the present chapter. The force field indicates, in accordance with experimental results, that the chair conformation is the more stable by about 5.9 kcal/mol. Thus we obtain all the details of the atomic positions (bond lengths, bond angles, etc.). Note that the chair conformation obtained exhibits D_{3d} symmetry, while the boat conformation corresponds to D_2 (the boat has somewhat warped planks, because of repulsion of the two upper hydrogen atoms).

As shown in Chapter 6, the solution of the Schrödinger equation in the Born–Oppenheimer approximation can be divided into two tasks: the problem of electronic motion in the field of the clamped nuclei (this will be the subject of the next chapters) and the problem of *nuclear motion in the potential energy determined by the electronic energy*.

> The ground-state electronic energy $E_k^0(\mathbf{R})$ of Eq. (6.8) (where $k = 0$ means the ground state) will be (following tradition) denoted in short as $V(\mathbf{R})$, where $\mathbf{R}$ represents the vector of the nuclear positions.

The function $V(\mathbf{R})$ has quite a complex structure and exhibits many basins of stable conformations (as well as many maxima and saddle points). *The problem of the shape of* $V(\mathbf{R})$*, as well as of the nuclear motion on the* $V(\mathbf{R})$ *hypersurface, will be the subject of the present chapter.*

It will be seen that the electronic energy can be computed within sufficient accuracy as a function of $\mathbf{R}$ only for very simple systems (such as an atom plus a diatomic molecule system), for which quite a lot of detailed information can be obtained. In practice, for large molecules, we are limited to only some approximations to $V(\mathbf{R})$ called *force fields*. After accepting such an approximation we encounter the problem of optimization of the positions of the nuclei, i.e., of obtaining the most stable molecular conformation (or configuration[1]). Geometry of such a conformation is usually identified with a minimum on the electronic energy hypersurface, playing the role of the potential energy for the nuclei. Finding a "nearest" stable conformation from a starting geometry of the nuclear framework is the subject of *local molecular mechanics*. In practice we have the problem that the number of such minima is huge for medium and large molecules. The real challenge in such a case is finding the *most* stable structure, usually corresponding to the global minimum (*global molecular mechanics*) of $V(\mathbf{R})$.

Molecular mechanics does not deal with nuclear motion as a function of time or the kinetic energy of the system (related to its temperature). This is the subject of molecular dynamics, which means solving the Newton equation of motion for all the nuclei of the system interacting through potential energy $V(\mathbf{R})$. Various approaches to this question (of general importance) will be presented at the end of the chapter.

---

[1] Two conformations correspond to the same pattern of chemical bonds, and differ by rotations of the fragments about some of these bonds. On the contrary going from one configuration to another means changing the chemical bonds.

Why is this important?

The spatial structure of molecules in atomic resolution represents the most important information, which decides about chemical and physical properties of substances. Such key information is offered by a few experimental methods only: X-ray diffraction analysis, neutron diffraction analysis (both methods require crystals of good quality), and nuclear magnetic resonance. *Theory not only offers a much less expensive alternative way (by minimizing* $E_0^0(\mathbf{R}) \equiv V(\mathbf{R})$*), but in addition often reveals many additional structural details that explain the experimental results.*

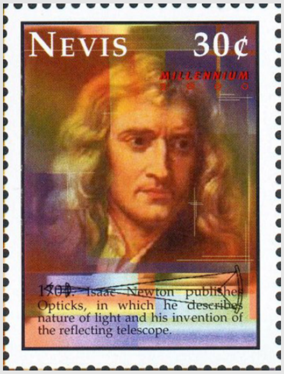
What is needed?

- Laplacian in spherical coordinates (Appendix H, p. 683, recommended),
- angular momentum operator and spherical harmonics (Chapter 4, recommended),
- harmonic oscillator (p. 217, necessary),
- Ritz method (Appendix M, p. 705, necessary),
- matrix diagonalization (Appendix L, p. 703, necessary),
- Newton equation of motion (necessary),
- Chapter 8 (an exception: the Car–Parrinello method needs some results which will be given in Chapter 8, marginally important),
- entropy, free energy, sum of states (necessary).

Classical works

There is no more classical work on dynamics than the monumental "*Philosophiae Naturalis Principia Mathematica*," Cambridge University Press, 1687 of Isaac Newton. ★ The idea of the force field was first presented by Mordechai Bixon and Shneior Lifson in *Tetrahedron* 23(1967)769 in a paper entitled "*Potential Functions and Conformations in Cycloalkanes*." ★ The paper by Berni Julian Alder and Thomas Everett Wainwright "*Phase Transition for a Hard Sphere System*" in *Journal of Chemical Physics*, 27(1957)1208 is considered the beginning of molecular dynamics. ★ The work by Aneesur Rahman "*Correlations in the Motion of Atoms in Liquid Argon*," published in *Physical Review A*, 136(1964)405 for the first time used a realistic interatomic potential (for 864 atoms). ★ The molecular dynamics of a small protein was first described in the paper by Andy McCammon, Bruce Gelin, and Martin Karplus under the title "*Dynamics of folded proteins*," *Nature*, 267(1977)585. ★ The simulated annealing method is believed to have been used first by Scott Kirkpatrick, Charles D. Gelatt, and Mario P. Vecchi in a work "*Optimization by Simulated Annealing*," *Science*, 220(1983)671. ★ The Metropolis criterion for the choice of the current configuration in the Monte Carlo method was given by Nicolas Constantine Metropolis, Arianna W. Rosenbluth, Marshal N. Rosenbluth, Augusta H. Teller, and Edward Teller in the paper "*Equations of State Calculations by Fast Computing Machines*" in *Journal of Chemical Physics*, 21(1953)1087. ★ The Monte Carlo method was used first by Enrico Fermi, John R. Pasta, and Stanisław Marcin Ulam during their stay in Los Alamos (E. Fermi, J.R. Pasta, S.M. Ulam, "*Studies of Non-Linear Problems*," vol. 1, *Los Alamos Reports*, LA-1940). Ulam is also the discoverer of cellular automata.

Isaac Newton (1643–1727), English physicist, astronomer, and mathematician, professor at Cambridge University, from 1672 member of the Royal Society of London, from 1699 Director of the Royal Mint – said to be merciless to the forgers. In 1705 Newton became a Lord. In the *opus magnum* mentioned above he developed the notions of space, time, mass, and force, gave three principles of dynamics and the law of gravity, and showed that the latter pertains to problems that differ enormously in their scale (e.g., the famous apple and the planets). Newton is also a founder of differential and integral calculus (independently from G.W. Leibnitz). In addition Newton made some fundamental discoveries in optics; among other things he was the first to think that light is composed of particles.



The first portrait of Newton (at the age of 46) by Godfrey Kneller is shown. The readers are encouraged to read the excellent book "Isaac Newton" by G.E. Christianson, Oxford University Press, New York, 1996.

7.1 Rovibrational spectra – an example of accurate calculations: atom–diatomic molecule

One of the consequences of adiabatic approximation is the potential energy hypersurface $V(\mathbf{R})$ for the motion of nuclei. To obtain the wave function for the motion of nuclei (and then to construct the total product-like wave function for the motion of electrons and nuclei) we have to solve the Schrödinger equation with $V(\mathbf{R})$ as the potential energy. This is what this hypersurface is for. We will find rovibrational energy levels and the corresponding wave functions, which will serve us to obtain rovibrational spectra (frequencies and intensities) to compare with experimental results.

7.1.1 Coordinate system and Hamiltonian

Let us consider a diatomic molecule AB plus a weakly interacting atom C (e.g., HH$\cdots$Ar or CO$\cdots$He), the total system in its electronic ground state. Let us center the origin of the body-fixed coordinate system[2] (with the axes oriented as in the space-fixed coordinate system,

---

[2] Any coordinate system is equally good from the point of view of mathematics, but its particular choice may make the solution easy or difficult. In the case of a weak C$\cdots$AB interaction (our case) the proposed choice of the origin is one of the natural ones.

see Appendix J, p. 691) in the center of mass of AB. The problem involves therefore $3 \times 3 - 3 = 6$ dimensions.

However strange it may sound, six is too much for contemporary (otherwise impressive) computer techniques. Let us subtract one dimension by assuming that no vibrations of AB occur (rigid rotator). The five-dimensional problem becomes manageable. The assumption about the stiffness of AB now also pays off because we exclude right away two possible chemical reactions C + AB→ CA + B and C + AB → CB + A, and admit therefore only some limited set of nuclear configurations – only those that correspond to a weakly bound complex C + AB. This approximation is expected to work better when the AB molecule is stiffer, i.e., has a larger force constant (and therefore vibration frequency).[3]

Carl Gustav Jacob Jacobi (1804–1851), German mathematical genius, son of a banker, graduated from school at the age of 12, then associated with the universities of Berlin and Königsberg. Jacobi made important contributions to number theory, elliptic functions, partial differential equations, and analytical mechanics.

We will introduce the *Jacobi coordinates* (Fig. 7.2): three components of vector $\mathbf{R}$ pointing to C from the origin of the coordinate system (the length R and angles Θ and Φ, both angles denoted by $\hat{\mathbf{R}}$) and the angles θ, ϕ showing the orientation $\hat{\mathbf{r}}$ of vector $\mathbf{r} = \overrightarrow{AB}$, altogether five coordinates – as there should be.

Now let us write down the Hamiltonian for the motion of the nuclei in the Jacobi coordinate system (with the stiff AB molecule with the AB equilibrium distance equal to r_{eq}).[4] We have

$$\hat{H} = -\frac{\hbar^2}{2\mu R^2}\frac{d}{dR}R^2\frac{d}{dR} + \frac{\hat{l}^2}{2\mu R^2} + \frac{\hat{j}^2}{2\mu_{AB} r_{eq}^2} + V,$$

where $\hat{l}^2$ denotes the operator of the square of the angular momentum of the atom C, $\hat{j}^2$ stands for the square of the angular momentum of the molecule AB, where

$$\hat{l}^2 = -\hbar^2\left[\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}\sin\Theta\frac{\partial}{\partial\Theta} + \frac{1}{\sin^2\Theta}\frac{\partial^2}{\partial\Phi^2}\right],$$

$$\hat{j}^2 = -\hbar^2\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right],$$

---

[3] A certain measure of this might be the ratio of the dissociation energy of AB to the dissociation energy of C$\cdots$AB. The higher the ratio, the better our model will be.

[4] As proposed in S. Bratož, M.L. Martin, *J. Chem. Phys.*, 42(1965)1051.

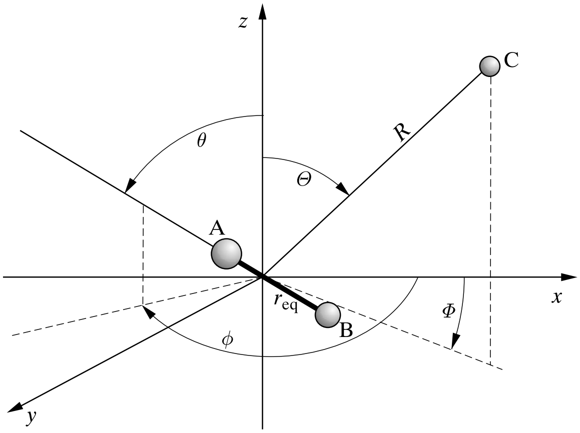


Fig. 7.2. The Jacobi coordinates for the C$\cdots$AB system. The origin is in the center of mass of AB (the distance AB is constant and equal to r_{eq}). The positions of atoms A and B are fixed by giving the angles θ,ϕ. The position of atom C is determined by three coordinates: R, Θ, and Φ. Altogether we have five coordinates: $R, \Theta, \Phi, \theta, \phi$ or (R, $\hat{\mathbf{R}}$ and $\hat{\mathbf{r}}$).

μ is the reduced mass of C and the mass of (A+B), μ_{AB} denotes the reduced mass of A and B, and V stands for the potential energy of the nuclear motion.

The expression for $\hat{H}$ is quite understandable. First of all, we have in $\hat{H}$ five coordinates, as there should be: R, two angular coordinates hidden in the symbol $\hat{\mathbf{R}}$, and two angular coordinates symbolized by $\hat{\mathbf{r}}$ – the four angular coordinates enter the operators of the squares of the two angular momenta. The first three terms in $\hat{H}$ describe the kinetic energy, V is the potential energy (the electronic ground-state energy depends on the nuclear coordinates). The kinetic energy operator describes the radial motion of C with respect to the origin (first term), the rotation of C about the origin (second term), and the rotation of AB about the origin (third term).

7.1.2 Anisotropy of the potential V

How to figure out the shape of V? Let us first make a section of V. If we freeze the motion of AB,[5] atom C would have (concerning the interaction energy) a sort of an energetic well around AB wrapping the AB molecule, caused by the C$\cdots$AB van der Waals interaction. The bottom of the well would be quite distant from the molecule (van der Waals equilibrium distance), while the shape determined by the bottom points would resemble the shape of AB, i.e., would

---

[5] That is, fixed the angles θ and ϕ.

be a little bit elongated. The depth of the well would vary depending on orientation with respect to the origin.

If V were isotropic, i.e., if atom C would have C$\cdots$AB interaction energy independent[6] of $\hat{\mathbf{r}}$, then of course we might say that there is no coupling between the rotation of C and the rotation of AB. We would then have a conservation law separately for the first and the second angular momentum and the corresponding commutation rules (cf. Chapter 2 and Appendix F), i.e.,

$$\left[\hat{H}, \hat{l}^2\right] = \left[\hat{H}, \hat{j}^2\right] = 0,$$

$$\left[\hat{H}, \hat{l}_z\right] = \left[\hat{H}, \hat{j}_z\right] = 0.$$

Therefore, the wave function of the total system would be the eigenfunction of $\hat{l}^2$ and $\hat{l}_z$ as well as of $\hat{j}^2$ and $\hat{j}_z$. The corresponding quantum numbers $l = 0, 1, 2, \ldots$ and $j = 0, 1, 2, \ldots$, which determine the squares of the angular momenta for $\hat{l}^2$ and $\hat{j}^2$, as well as the corresponding quantum numbers $m_l = -l, -l+1, \ldots, l$ and $m_j = -j, -j+1, \ldots, j$, which determine the projections of the corresponding angular momenta on the z axis, would be legal[7] quantum numbers (full analogy with the rigid rotator, Chapter 4). The rovibrational levels could be labeled using pairs of quantum numbers: (l, j). In the absence of an external field (no privileged orientation in space) any such level would be $(2l+1)(2j+1)$-tuply degenerate, since this is the number of different projections of both angular momenta on the z axis.

7.1.3 Adding the angular momenta in quantum physics

However, V *is not isotropic* (although the anisotropy is small). What then? Then, of all angular momenta, only the *total* angular momentum $\mathbf{J} = \mathbf{l} + \mathbf{j}$ is conserved (the conservation law results from the very foundations of physics, cf. Chapter 2).[8] Therefore, the vectors $\mathbf{l}$ and $\mathbf{j}$, when added to $\mathbf{J}$, would make all allowed angles: from minimum angle (the quantum number $J = l + j$), through larger angles[9] and the corresponding quantum numbers $J = l + j - 1, l + j - 2$, etc., up to the maximum angle, corresponding to $J = |l - j|$). Therefore,

---

[6] That is, the bottom of the well would be a sphere centered in the center of mass of AB and the well depth would be independent of the orientation.

[7] We use to say "*good*."

[8] Of course, the momentum has also been conserved in the isotropic case, but in this case the energy was identical independently of the quantum number J (resulting from different angles between $\mathbf{l}$ and $\mathbf{j}$).

[9] The projections of the angular momenta are quantized.

the number of all possible values of J (each corresponding to a different energy) is equal to the number of projections of the *shorter*[10] of the vectors $\mathbf{l}$ and $\mathbf{j}$ on the longer one, i.e.,

$$J = (l + j), (l + j - 1), \ldots, |l - j|. \tag{7.1}$$

For a given J there are $2J + 1$ projections of $\mathbf{J}$ on the z axis (because $|M_J| \leq J$); without any external field all these projections correspond to identical energy.

Please check that the number of all possible eigenstates is equal to $(2l + 1)(2j + 1)$, i.e., exactly what we had in the isotropic case. For example, for $l = 1$ and $j = 1$ the degeneracy in the isotropic case is equal to $(2l + 1)(2j + 1) = 9$, while for anisotropic V we would deal with five states for $J = 2$ (all of the same energy), three states corresponding to $J = 1$ (the same energy, but different from $J = 2$), and a single state with $J = 0$ (still another value of energy), altogether nine states. This means that switching anisotropy on partially removed the degeneracy of the isotropic level (l, j) and gave the levels characterized by quantum number J.

7.1.4 Application of the Ritz method

We will use the Ritz variational method (see Chapter 5, p. 271) to solve the Schrödinger equation. What should we propose as the expansion functions? It is usually recommended that we proceed systematically and choose first a complete set of functions depending on R, then a complete set depending on $\hat{\mathbf{R}}$, and finally a complete set that depends on the $\hat{\mathbf{r}}$ variables. Next, one may create the complete set depending on all five variables (these functions will be used in the Ritz variational procedure) by taking all possible products of the three functions depending on R, $\hat{\mathbf{R}}$, and $\hat{\mathbf{r}}$. There is no problem with the complete sets that have to depend on $\hat{\mathbf{R}}$ and $\hat{\mathbf{r}}$, as these may serve the spherical harmonics (the wave functions for the rigid rotator) $\{Y_l^m(\Theta, \Phi)\}$ and $\{Y_{l'}^{m'}(\theta, \phi)\}$, while for the variable R we may propose the set of harmonic oscillator wave functions $\{\chi_v(R)\}$.[11] Therefore, we may use the following as the variational function[12]:

$$\Psi(R, \Theta, \Phi, \theta, \phi) = \sum c_{vlml'm'} \chi_v(R) Y_l^m(\Theta, \Phi) Y_{l'}^{m'}(\theta, \phi),$$

---

[10] In the case of two vectors of the same length, the role of the shorter vector may be taken by either of them.

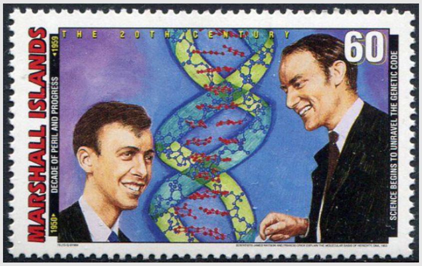
[11] See Chapter 4. Of course, our system does not represent any harmonic oscillator, but what counts is that the harmonic oscillator wave functions form a complete set (as the eigenfunctions of a Hermitian operator).

[12] The products $Y_l^m(\Theta, \Phi) Y_{l'}^{m'}(\theta, \phi)$ may be used to produce linear combinations that are automatically the eigenfunctions of $\hat{J}^2$ and $\hat{J}_z$, and have the proper parity (see Chapter 2). This may be achieved by using the Clebsch–Gordan coefficients (D.M. Brink, G.R. Satchler, *Angular Momentum*, Clarendon, Oxford, 1975). The good news is that this way we can obtain a smaller matrix for diagonalization in the Ritz procedure, the bad news is that the matrix elements will contain more terms to be computed. The method described above will give the same result as using the Clebsch–Gordan coefficients, because the eigenfunctions of the Hamiltonian obtained within the Ritz method will automatically be the eigenfunctions of $\hat{J}^2$ and $\hat{J}_z$ and have the proper parity.

where c are the variational coefficients and the summation goes over v, l, m, l', m' indices. The summation limits have to be finite in practical applications; therefore the summations go to some maximum values of v, l, and l' (m and m' vary from $-l$ to l and from $-l'$ to $+l'$). We hope (as always in quantum chemistry) that numerical results of a demanded accuracy will not depend on these limits. Then, as usual, the Hamiltonian matrix is computed and diagonalized (see p. 703), and the eigenvalues E_J as well as the eigenfunctions ψ_{J,M_J} of the ground and excited states are found. Each of the eigenfunctions will correspond to some J, M_J and to a certain parity. The problem is solved.

7.2 Force fields (FFs)

The middle of the 20th century marked the end of a long period of determining the building blocks of chemistry: chemical elements, chemical bonds, and bond angles. The lists of these are not definitely closed, but future changes will be rather cosmetic than fundamental. This made it possible to go one step further and begin to rationalize the structure of molecular systems as well as to foresee the structural features of the compounds to be synthesized. The crucial concept is based on the Born–Oppenheimer approximation and on the theory of chemical bonds and resulted in the spatial structure of molecules. The great power of such an approach was first proved by the construction of the DNA double-helix model by Watson and Crick. The first DNA model was build from iron spheres, wires, and tubes.



James Dewey Watson (left), born in 1928, American biologist, professor at Harvard University. Francis Harry Compton Crick (right) (1916–2004), British physicist, professor at Salk Institute in San Diego. Both scholars won the 1962 Nobel Prize for "*their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material.*" At the end of the historic paper J.D. Watson, F.H.C. Crick, *Nature*, 737(1953) (of about 800 words) the famous enigmatic but crucial sentence appears: "*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.*" The story behind the discovery is described in a colorful and nonconventional way by Watson in his book "*Double helix: A Personal Account of the Discovery of the Structure of DNA.*"

This approach created problems: one of the founders of force fields, Michael Levitt, recalls[13] that a model of a tRNA fragment constructed by him with 2000 atoms weighed more than 50 kg. And here shows the power of mathematics.

The experience accumulated paid off by proposing some approximate mathematical expressions for electronic energy, which is, as we know from Chapter 6, the potential energy of the motion of the nuclei. This is what we are going to discuss.

Suppose we have a molecule (a set of molecules can be treated in a similar way). We will introduce the *force field*, which will be a *scalar* field – a function $V(\mathbf{R})$ of the nuclear coordinates $\mathbf{R}$. The function $V(\mathbf{R})$ represents a generalization (from one dimension to $(3N-6)$ dimensions) of the function $E_0^0(R)$ of Eq. (6.8) on p. 316. The force acting on atom j occupying position x_j, y_j, z_j is computed as the proper components of the vector $\mathbf{F}_j = -\nabla_j V$, where

$$\nabla_j = \mathbf{i} \cdot \frac{\partial}{\partial x_j} + \mathbf{j} \cdot \frac{\partial}{\partial y_j} + \mathbf{k} \cdot \frac{\partial}{\partial z_j} \tag{7.2}$$

with $\mathbf{i}, \mathbf{j}, \mathbf{k}$ denoting the unit vectors along x, y, z, respectively.

FORCE FIELD

A force field represents a mathematical expression $V(\mathbf{R})$ for the electronic energy as a function of the nuclear configuration $\mathbf{R}$. Its gradient gives the forces acting on the atoms.

Of course, if we had to write down this scalar field in an accurate way, we would have to solve (with an accuracy of about 1 kcal/mol) the electronic Schrödinger equation (6.8) for every configuration $\mathbf{R}$ of the nuclei and take the eigenvalue (i.e., an analog of $E_0^0(R)$) as $V(\mathbf{R})$. This would take so much time, even for small systems composed of a few atoms, that we would abandon this method with a great feeling of relief. Even if such a calculation required huge computation time, it would give results which would have been quite simple in their overall features (assuming that the molecule has a pattern of chemical bonds). It just would turn out that V could be approximated by the following model function.

---

[13] M. Levitt, *Nature Struct. Biol.*, 8(2001)392.

- **Chemical bonds**. Suppose that a chemical bond between atoms X and Y had *a certain characteristic reference length* r_0 that would depend on the chemical character of the atoms X and Y. Then for this particular length $V(\mathbf{R})$ would be close to minimum. If the bond length were changed (shortened or elongated) to a certain value r, then the energy would increase, first according to the harmonic law (with force constant k_{XY}) and then some deviations from the harmonic approximation would appear.[14] A harmonic term of the kind $\frac{1}{2}k_{XY}(r-r_0)^2$ incorporated additively into V *replaces the true anharmonic dependence by a harmonic approximation* (assumption of small amplitudes) as if the two atoms had been bound by a harmonic spring (in the formula the atomic indices at symbols of distances have been omitted). The most effective feature is that the same formula $\frac{1}{2}k_{XY}(r-r_0)^2$ is assumed for all chemical bonds X–Y, *independently of some particular chemical neighborhood of a given X–Y bond*. For example, one assumes that any nondistorted single CC bond[15] has a characteristic reference length $r_0 = 1.523$ Å and a characteristic force constant $k_{XY} = 317\frac{\text{kcal}}{\text{mol}\cdot\text{Å}^2}$; similarly, some distinct parameters pertain to the C=C bond: $r_0 = 1.337$ Å, $k_{XY} = 690\frac{\text{kcal}}{\text{mol}\cdot\text{Å}^2}$, etc.[16]
- **Bond angles**. While preserving the distances r in the A–B and B–C bonds we may change the bond angle α =A–B–C, in this way changing the A$\cdots$C distance. A corresponding change of V has to be associated with such a change. The energy has to increase when the angle α deviates from a characteristic reference value α_0. The harmonic part of such a change may be modeled by $\frac{1}{2}k_{XYZ}(\alpha-\alpha_0)^2$ (the indices for angles are omitted), which is equivalent to setting a corresponding harmonic spring for the bond angle and requires small amplitudes $|\alpha-\alpha_0|$. The k_{XYZ} are assumed "universal," i.e., they depend on the chemical character of the X, Y, Z atoms, and do not depend on other details, such as the neighborhood of these atoms. For example, for the angle C–C–C $\alpha_0 = 109.47°$ and $k_{XYZ} = 0.0099\frac{\text{kcal}}{\text{mol}\cdot\text{degree}^2}$ (for all C–C–C fragments of the molecule), which means that to change the C$\cdots$C distance by varying angle is about an order of magnitude easier than to change a CC bond length.

---

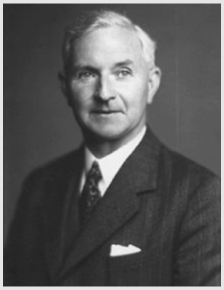
[14] These deviations from harmonicity (i.e., from the proportionality of force and displacement) are related to the smaller and smaller force needed to elongate the bond by a unit length and the larger and larger force needed to shorten the bond.

[15] That is, when all other terms in the force field equal zero.

[16] A CC bond involved in conjugated single and double bonds (e.g., in benzene) also has its own parameters.

- **Van der Waals interaction**. Two atoms X and Y, which do not form a chemical bond X–Y and do not participate in any sequence of bonds X–A–Y, still interact. There is nothing in the formulae introduced above that would prevent X and Y from *collapsing* without any change of V. However, when two such atoms approach at a distance smaller than the sum of their radii (the van der Waals radii, see p. V2-411), then V has to increase very greatly.[17] On the other hand, at large interatomic distances the two atoms have to attract each other by the dispersion interaction vanishing as r^{-6} (cf. Chapter V2-5, p. V2-353). Hence, there is an equilibrium distance r_e, at which the interaction energy attains a minimum equal to $-\varepsilon$. These features of the interaction are captured by the widely used Lennard-Jones potential $V_{LJ}(X,Y) = \varepsilon\left[\left(\frac{r_e}{r}\right)^{12} - 2\left(\frac{r_e}{r}\right)^{6}\right]$, where we skip for brevity the indices X, Y on the right-hand side. The Lennard-Jones potential given above is called LJ 12-6 (reflecting the powers involved). Sometimes other powers are used, leading to other "LJ *m*-*n*" potentials.[18] Due to their simplicity, LJ potentials are widely used (Fig. 7.3).

John E. Lennard-Jones (1894–1954) was professor of theoretical chemistry of the University of Cambridge, UK. The reader may find a historic picture of the theoretical chemistry team in *Intern. J. Quantum Chemistry*, S23(1989) p. XXXII.



- **Electrostatic interaction**. All the terms we have introduced to V so far do not take into account the fact that atoms carry net charges q_X and q_Y that have to interact electrostatically by Coulombic forces. To take this effect into account the electrostatic energy terms $q_X q_Y / r$ are added to V, where we assume the net charges q_X and q_Y are *fixed* (i.e., independent of the molecular conformation).[19]

---

[19] In some force fields the electrostatic forces depend on the dielectric constant of the neighborhood (e.g., solvent) despite the fact that this quantity has a macroscopic character and does not pertain to the nearest neighborhood of the interacting atoms. If all the calculations had been carried out taking the molecular structure of the solvent into account as well as the polarization effects, no dielectric constant would have been needed. If this is not possible, then the dielectric constant effectively takes into account the polarization of the medium (including reorientation of the solvent molecules). The next problem is *how* to introduce the dependence of the electrostatic interaction of two atomic charges on the dielectric constant. In some of the force fields we introduce a brute force kind of damping, namely, the dielectric constant is introduced into the denominator of the Coulombic interaction as equal to the interatomic distance. This is equivalent to saying that the electrostatic interaction is practically damped down for large distances.
In second generation force fields we explicitly take into account the induction interaction, e.g., the dependence of the atomic electric charges on molecular conformations. Such force fields, when explicitly taking the solvent molecules, should not introduce the dielectric constant.

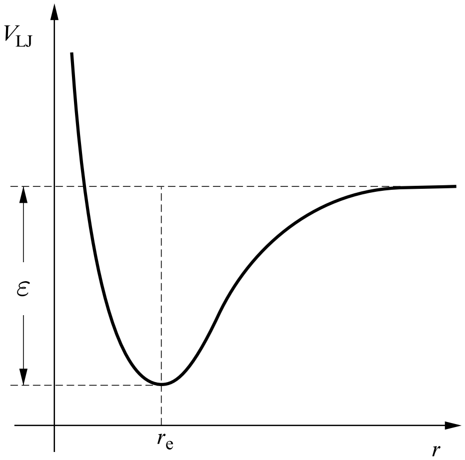


Fig. 7.3. The Lennard-Jones (LJ 12-6) potential. The parameter ε represents the depth of the potential well, while the parameter r_e denotes the minimum position. This r_e, corresponding to the *nonbonding* interaction of atoms X and Y, has no direct relation to the r_0 value pertaining to the *chemical* bond X–Y. The first is larger, by about an angstrom or so, than the second.

- **Torsional interactions**. In addition to all the terms described above we often introduce to the force field a torsional term $A_{X-Y-Z-W}\,(1-\cos n\omega)$ for each torsional angle ω showing how V changes when a rotation ω about the chemical bond Y–Z, in the sequence X–Y–Z–W of chemical bonds, takes place (n is the multiplicity of the energy barriers per single turn[20]). Some rotational barriers already result from the van der Waals interaction of the X and W atoms, but in practice the barrier heights have to be corrected by the torsional potentials to reproduce experimental values.
- **Mixed terms**. Besides the terms described above one often introduces some *coupling (mixed) terms*, e.g., bond–bond angle, etc. The reasoning behind this is simple. The bond angle force constant for X–Y–Z has to depend on the bond lengths X–Y and Y–Z, etc.

Summing up a simple force field might be expressed as shown in Fig. 7.4, where for the sake of simplicity the indices X, Y (at r, r_0), X, Y, Z (at α, α_0), and X, Y, Z, W (at ω) have been omitted. We obtain

$$V = \sum_{X-Y} \frac{1}{2} k_{XY}\,(r-r_0)^2 + \sum_{X-Y-Z} \frac{1}{2} k_{XYZ}\,(\alpha-\alpha_0)^2 + \sum_{X\ldots Y} V_{LJ}\,(X,Y) +$$
$$\sum_{X,Y} \frac{q_X q_Y}{r} + \sum_{tors} A_{X-Y-Z-W}\,(1-\cos n\omega) + \text{ coupling terms (if any).}$$

---

[20] For example, $n = 3$ for ethane.

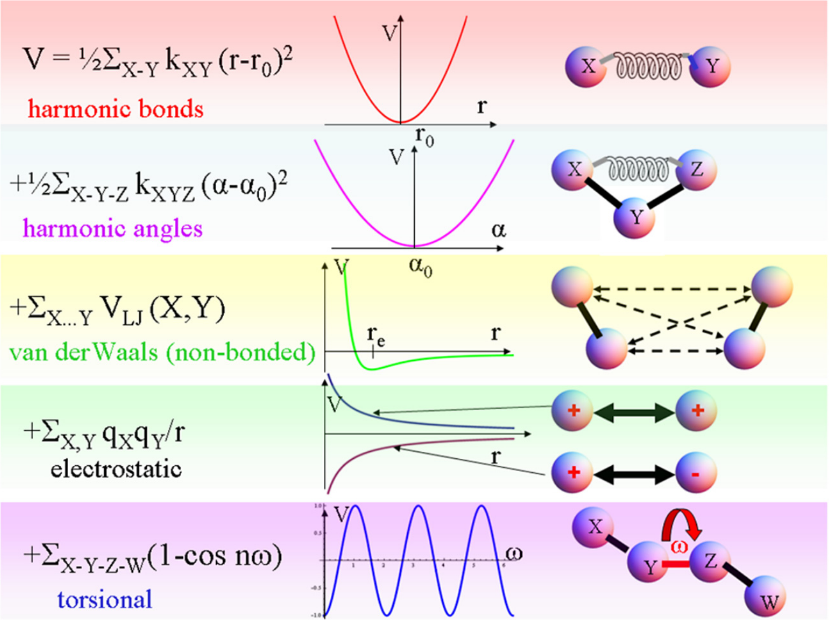


Fig. 7.4. The simplest force field in a mnemonic presentation.

Hence, some simple formulae help us to figure out how the electronic energy $E_0^0(\mathbf{R}) = V(\mathbf{R})$ looks as a function of the configuration $\mathbf{R}$ of the nuclei. Our motivation is as follows:

- Economy of computation: *ab initio* calculations of the electronic energy for larger molecules would have been many orders of magnitude more expensive.
- In addition, a force field gives $V(\mathbf{R})$ in the form of a simple formula for *any* positions $\mathbf{R}$ of the nuclei, while the calculation of the electronic energy would give us $V(\mathbf{R})$ numerically, i.e., for some *selected* nuclear configurations.

7.3 Local molecular mechanics

7.3.1 Bonds that cannot break

It is worth noting that the force fields correspond to a fixed (and unchangeable during computation) system of chemical bonds. The chemical bonds are treated as springs, most often satisfying Hooke's[21] law (harmonic), and are therefore unbreakable.[22] Similarly, the bond angles are

---

[21] Robert Hooke, British physicist and biologist (1635–1703).

[22] There are a few such force fields in the literature. They give similar results, as far as their main features are considered. The force field concept was able to clarify many observed phenomena and even fine effects. It may also fail as with anything confronted with the real world.

forced to satisfy Hooke's law. Such a force field is known as *flexible molecular mechanics*. To decrease the number of variables, we sometimes use *rigid molecular mechanics*[23], in which the bond lengths and the bond angles are fixed at values close to experimental ones, but the torsional angles are free to change. The argument behind such a choice is that a quantity of energy that is able to make only tiny changes in the bond lengths and larger but still small changes in the bond angles can make large changes in the torsional angles, i.e., the torsional variables determine the overall changes of the molecular geometry. Of course, the second argument is that a smaller number of variables means lower computational costs.

Molecular mechanics represents a method of finding a stable configuration of the nuclei by using a minimization of $V(\mathbf{R})$ with respect to the nuclear coordinates (for a molecule or a system of molecules). The essence of molecular mechanics is that we roll the potential energy hypersurface slowly downhill from a starting point chosen (corresponding to a certain starting geometry of the molecule) to the "nearest" energy minimum corresponding to the final geometry of the molecule.

The "rolling down" is carried out by a minimization procedure that traces point-by-point the trajectory in the configurational space, e.g., in the direction of the negative gradient vector calculated at any consecutive point. The minimization procedure represents a mechanism showing how to obtain the next geometry from the previous one. The procedure ends when the geometry ceases to change (e.g., the gradient vector has zero length[24]). The geometry attained is called the equilibrium or stable geometry. The rolling described above is more like a crawling down with large friction, since in molecular mechanics the kinetic energy is always zero and the *system is unable to go uphill* of V.

A lot of commercial software offers force field packages. Unfortunately, the results depend to quite a significant degree on the force field chosen. Even using the same starting geometry we may obtain final (equilibrium) results that differ very much from one another. Usually the equilibrium geometries obtained in one force field do not differ much from those from another one, but the corresponding energies may be very different. Therefore, the most stable geometry (corresponding to the lowest energy) obtained in a force field may turn out to be less stable in another one, thus leading to different predictions of the molecular structure.

---

[23] The rigid molecular mechanics was a very useful tool for Paul John Flory (1910–1985), American chemist, professor at the universities of Cornell and Stanford. Using such mechanics, Flory developed a theory of polymers that explained their physical properties. In 1974 he obtained the Nobel Prize "*for his fundamental achievements, both theoretical and experimental, in the physical chemistry of macromolecules*."

[24] The gradient is also equal zero at energy maxima and energy saddle points. To be sure that a minimum really has been finally attained we have to calculate (at the particular point suspected to be a minimum) a Hessian, i.e., the matrix of the second derivatives of V, and then diagonalize it (cf. p. 703) and check whether the eigenvalues obtained are all positive.

A big problem in molecular mechanics is that the final geometry is very close to the starting one. We start from a boat (chair) conformation of cyclohexane and obtain a boat (chair) equilibrium geometry. The very essence of molecular mechanics, however, is that when starting from some, e.g., distorted boat (chair) conformation, we obtain the perfect, beautiful equilibrium boat (chair) conformation, which may be compared with experimental results. Molecular mechanics is extremely useful in conformational studies of systems with a small number of stable conformations, either because the molecule is small, it is rigid, or its overall geometry is fixed. In such cases all, or all "reasonable,[25]" conformations can be investigated and those of lowest energy can be compared with experimental results.

7.3.2 Bonds that can break

Harmonic bonds cannot be broken and therefore molecular mechanics with harmonic approximation is unable to describe chemical reactions. When instead of harmonic oscillators we use the Morse model (p. 224), then the bonds can be broken.

And yet we most often use the harmonic oscillator approximation. Why? There are a few reasons:

- the Morse model requires many computations of the exponential function, which is expensive[26] when compared to the harmonic potential;
- the Morse potential requires three parameters, while the harmonic model needs only two parameters;
- in most applications the bonds do not break and it would be very inconvenient to obtain breaking due to, for instance, a particular starting point;
- a description of chemical reactions requires not only the possibility of breaking bonds, but also a realistic, i.e., quantum chemical, computation of the charge distributions involved (cf. p. 433); the Morse potential would be too simplistic for such purposes.

7.4 Global molecular mechanics

7.4.1 Multiple minima catastrophe

If the number of local minima is very large (and this may happen even for medium-sized molecules) or even "astronomic," then exploring the whole conformational space (all pos-

---

[25] A very dangerous word!

[26] Each time requires a Taylor expansion calculation.

sible geometries) by finding all possible minima using a minimization procedure becomes impossible. Hence, we may postulate another procedure which may be called *global molecular mechanics* and could find the global minimum (the most stable conformation) starting from any point in the configurational space.

If the number of local minima is small, there is in principle no problem with using theory. Usually it turns out that the quantum mechanical calculations are feasible, often even at the *ab initio* level. A closer insight leads, however, to the conclusion that only some extremely accurate and expensive calculations would give the correct energy sequence of the conformers, and only for quite small molecules up to a dozen atoms. This means that for larger molecules we are forced to use molecular mechanics. For molecules with a few atoms we might investigate the whole conformational space by sampling it by a stochastic or systematic procedure, but this approach soon becomes prohibitive for larger molecules.

For such larger molecules we encounter difficulties which may only be appreciated by individuals who have made such computations themselves. We may say, in short, that virtually nothing helps us with the huge number of conformations to investigate. According to Schepens[27] the number of conformations found is proportional to the time spent conducting the search. It is worth noting that this means catastrophe, because for a 20-amino acid oligopeptide the number of conformations is of the order[28] of 10^{20}, and for a hundred amino acids it is of the order of 10^{100}. Also methods based on molecular dynamics (cf. p. 427) do not solve the problem, since they could cover only a tiny fraction of the total conformational space.

7.4.2 Is it the global minimum which counts?

The goal of conformational analysis is to find those conformations of the molecule which are observed under experimental conditions. At temperatures close to 300 K the lowest-energy conformations prevail in the sample, i.e., first of all those corresponding the global minimum of the potential energy V.

---

[27] Wijnand Schepens, PhD thesis, University of Ghent, 2000.

[28] The difficulty of finding a particular conformation among 10^{20} conformations is a real horror. Maybe the example below will show the severity of the problem that has been encountered. A single grain of sand has a diameter of about 1 mm. Let us try to align 10^{20} of such sand grains side by side. What will the length of such a chain of grains be? Let us compute: 10^{20} mm $= 10^{17}$ m $= 10^{14}$ km. One light year is $300\,000$ km/s $\times$ 3600 s $\times 24 \times 365 \simeq 10^{13}$ km. Hence, the length is about 10 light years, i.e., longer than the round trip from our Sun to the nearest star, Alpha Centauri. This is what the thing looks like.

We may ask whether indeed the global minimum of the potential energy decides the observed experimental geometry. Let us neglect the influence of the solvent (neighborhood). A better criterion would be the global minimum of the *free energy*, $E - TS$, where the entropic factor would also enter.[29] A wide potential well means a higher density of vibrational states, a narrow well means a lower density of states (cf. Eq. (4.26), p. 226; a narrow well corresponds to a large α). If the global minimum corresponds to a wide well, the system in such a well is additionally stabilized by entropy.

For large molecules, there is a possibility that, due to the synthesis conditions, the molecule is trapped in a local minimum (*kinetic minimum*), different from the global minimum of the free energy (*thermodynamic minimum*) (Fig. 7.5).

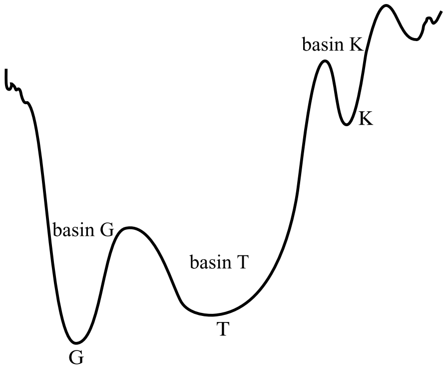


Fig. 7.5. The basins of the *thermodynamic minimum* (T), of the *kinetic minimum* (K), and of the *global minimum* (G). The deepest basin (G) should not correspond to the thermodynamically most stable conformation (T). Additionally, the system may be caught in a kinetic minimum (K), from which it may be difficult to tunnel to the thermodynamic minimum basin. Explosives and fullerenes may serve as examples of K.

For the same reason that the diamonds (kinetic minimum) in your safe do not change spontaneously into graphite (thermodynamic minimum), a molecule imprisoned in the kinetic minimum may rest there for a very long time (when compared with the time of experiment). We will cite two examples of kinetically trapped molecules. Explosive materials represent some high-energy substances, which when stimulated in some way leave their high-energy minimum. In a small fraction of a second they transform to (mainly gaseous) products.

---

[29] According the famous formula of Ludwig Boltzmann, entropy $S = k_B \ln \Omega(E)$, where Ω is the number of the states available for the system at energy E. The more states, the larger the entropy.

Christian Anfinsen (1916–1995) obtained the Nobel Prize in 1972 "*for his work on ribonuclease, especially concerning the connection between the amino acid sequence and the biologically active conformation*." He made an important contribution showing that after denaturation (a large change of conformation) some proteins fold back spontaneously to their native conformation. ©The Nobel Foundation.

Despite these complications, we generally assume in conformational analysis that the global minimum and other low-energy conformations play the most important role. In living matter, taking a definite (native) conformation is sometimes crucial. It has been shown[30] that the native conformation of natural enzymes has much lower energy than those of other conformations (energy gap). Artificial enzymes with stochastic amino acid sequences do not usually have this property, resulting in no well-defined conformation.

According to the present author global molecular mechanics is one of the most important challenges not only in chemistry, but in natural sciences in general. This is because optimization is fundamental in virtually all domains and, if it is treated mathematically, usually transforms into a problem of finding a global minimum.[31]

7.5 Small amplitude harmonic motion – normal modes

The hypersurface $V(\mathbf{R})$ has, in general (especially for large molecules), an extremely complex shape with many minima, each corresponding to a stable conformation. Let us choose *one* of those minima and ask *what kind of motion the molecule undergoes when only small displacements from the equilibrium geometry are allowed*. In addition we assume that the potential energy for this motion is a *harmonic* approximation of $V(\mathbf{R})$ in the neighborhood of the minimum.[32] Then we obtain the *normal vibrations* or *normal modes*.

---

[30] E.I. Shakanovich, A.M. Gutin, *Proc. Natl. Acad. Sci. USA*, 90(1993)7195; A. Šali, E.I. Shakanovich, M. Karplus, *Nature*, 369(1994)248.

[31] The reader may find my adventure with this problem in L. Piela in *Handbook of Global Optimization*, vol. 2, (2002)461, P.M. Pardalos, H.E. Romeijn (eds.), *Kluwer Academic Publishers.*

[32] We may note *en passant* that a similar philosophy prevailed in science until quite recent times: take only the linear approximation and forget about nonlinearities. It turned out, however, that the nonlinear phenomena (cf. Chapter V2-7) are really fascinating.

NORMAL MODES

A normal mode represents a harmonic oscillation (of a certain frequency) of all the atoms of the molecule about their equilibrium positions with the same phase for all the atoms (i.e., all the atoms attain their equilibrium position at the same time).

7.5.1 Theory of normal modes

Suppose we have at our disposal an analytical expression for $V(\mathbf{R})$ (e.g., the force field), where $\mathbf{R}$ denotes the vector of the Cartesian coordinates of the N atoms of the system (it has $3N$ components). Let us assume (Fig. 7.6) that the function $V(\mathbf{R})$ has been minimized in the configurational space, starting from an initial position $\mathbf{R}_i$ and going downhill until a minimum position $\mathbf{R}_0$ has been reached, with $\mathbf{R}_0$ corresponding to one of many minima the V function may possess[33] (we will call the minimum the "closest" to the $\mathbf{R}_i$ point in the configurational space). All the points $\mathbf{R}_i$ of the configurational space that lead to $\mathbf{R}_0$ represent the *basin of the attractor*[34] $\mathbf{R}_0$.

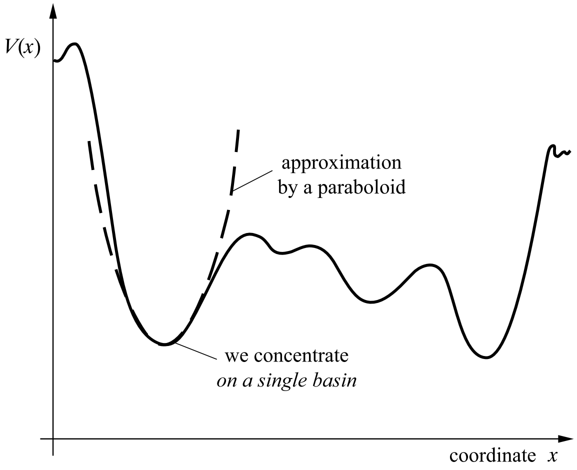


Fig. 7.6. A schematic (one-dimensional) view of the hypersurface $V(x)$ that illustrates the choice of a particular basin of V related to the normal modes to be computed. The basin chosen is then approximated by a paraboloid in $3N$ variables. This gives the $(3N-6)$ modes with nonzero frequencies and six "modes" with zero frequencies.

---

[33] These are improper minima, because a translation or rotation of the system does not change V.

[34] The total configurational space consists of a certain number of such basins.

From this time on, all other basins of the function $V(\mathbf{R})$ *have "disappeared from the theory" – only motion in the neighborhood of* $\mathbf{R}_0$ is to be considered.[35] If someone is aiming to apply harmonic approximation and to consider small displacements from $\mathbf{R}_0$ (as we do), then it is a good idea to write down the Taylor expansion of V about $\mathbf{R}_0$ (hereafter instead of the symbols $X_1, Y_1, Z_1, X_2, Y_2, Z_2, \ldots$ for the atomic Cartesian coordinates we will use a slightly more uniform notation: $\mathbf{R} = (X_1, X_2, X_3, X_4, X_5, X_6, ..X_{3N})^T$), i.e.,

$$V(\mathbf{R}_0+\mathbf{x}) = V(\mathbf{R}_0) + \sum_i \left(\frac{\partial V}{\partial x_i}\right)_0 x_i + \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_0 x_i x_j + \ldots, \tag{7.3}$$

where $\mathbf{x} = \mathbf{R} - \mathbf{R}_0$ is the vector with the *displacements* of the atomic positions from their equilibria ($x_i = X_i - X_{i,0}$ for $i = 1, \ldots, 3N$), while the derivatives are computed at $\mathbf{R} = \mathbf{R}_0$.

In $\mathbf{R}_0$ all the first derivatives vanish. According to the harmonic approximation, the higher-order terms denoted as "$+\ldots$" are neglected. In effect we have

$$V(\mathbf{R}_0+\mathbf{x}) \cong V(\mathbf{R}_0) + \frac{1}{2}\sum_{ij}\left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_0 x_i x_j. \tag{7.4}$$

In matrix notation we have $V(\mathbf{R}_0+\mathbf{x}) = V(\mathbf{R}_0) + \frac{1}{2}\mathbf{x}^T\mathbf{V}''\mathbf{x}$, where $\mathbf{V}''$ is a square matrix of the Cartesian *force constants*, $(\mathbf{V}'')_{ij} = \left(\frac{\partial^2 V}{\partial x_i \partial x_j}\right)_0$ computed at the minimum.

The Newton equations of motion for all the atoms of the system can be written in matrix form as ($\ddot{\mathbf{x}}$ means the second derivative with respect to time t)

$$\mathbf{M}\ddot{\mathbf{x}} = -\mathbf{V}''\mathbf{x}, \tag{7.5}$$

where $\mathbf{M}$ is the diagonal matrix of the atomic masses (the numbers on the diagonal are $M_1, M_1, M_1, M_2, M_2, M_2, \ldots$), because we calculate the force component along the k axis as $-\frac{\partial V}{\partial x_k} = -\frac{1}{2}\sum_j \left(\frac{\partial^2 V}{\partial x_k \partial x_j}\right)_0 x_j - \frac{1}{2}\sum_i \left(\frac{\partial^2 V}{\partial x_i \partial x_k}\right)_0 x_i = -\sum_j \left(\frac{\partial^2 V}{\partial x_k \partial x_j}\right)_0 x_j = -(\mathbf{V}''\mathbf{x})_k$.

We may use the relation $\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}} = \mathbf{M}$ in order to write Eq. (7.5) in a slightly different way:

$$\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\ddot{\mathbf{x}} = -\mathbf{M}^{\frac{1}{2}}\mathbf{M}^{-\frac{1}{2}}\mathbf{V}''\mathbf{M}^{-\frac{1}{2}}\mathbf{M}^{\frac{1}{2}}\mathbf{x}, \tag{7.6}$$

where $\mathbf{M}^{\frac{1}{2}}$ is a matrix similar to $\mathbf{M}$, but its elements are the square roots of the atom masses instead of the masses, while the matrix $\mathbf{M}^{-\frac{1}{2}}$ contains the inverse square roots of the masses. The last equation, after multiplying from the left by $\mathbf{M}^{-\frac{1}{2}}$, gives

---

[35] For another starting conformation $\mathbf{R}_i$ we might obtain another minimum of $V(\mathbf{R})$. This is why the choice of $\mathbf{R}_i$ has to have a definite relation to that which is observed experimentally.

$$\ddot{\mathbf{y}} = -\mathbf{A}\mathbf{y}, \tag{7.7}$$

where $\mathbf{y} = \mathbf{M}^{\frac{1}{2}}\mathbf{x}$ and $\mathbf{A} = \mathbf{M}^{-\frac{1}{2}}\mathbf{V}''\mathbf{M}^{-\frac{1}{2}}$.

Let us try to find the solution in the form[36]

$$\mathbf{y} = \mathbf{c}_1 \exp(+i\omega t) + \mathbf{c}_2 \exp(-i\omega t),$$

where the vectors $\mathbf{c}_i$ (of the dimension $3N$) of the complex coefficients are time-independent. The coefficients $\mathbf{c}_i$ depend on the initial conditions as well as on the matrix $\mathbf{A}$. If we decide to find a solution, in which at time $t = 0$ all the atoms are at equilibrium, i.e., $\mathbf{y}(t = 0) = \mathbf{0}$, then we obtain the relation $\mathbf{c}_1 = -\mathbf{c}_2$, leading to the formula

$$\mathbf{y} = \mathbf{L}\sin(\omega t), \tag{7.8}$$

where the vector[37] $\mathbf{L}$ and ω depend on the matrix $\mathbf{A}$. Vector $\mathbf{L}$ is determined only to the accuracy of a multiplication constant, because multiplication of $\mathbf{L}$ by any number does not interfere with satisfying (7.7).

When we insert the proposed solution (7.8) in (7.7), we immediately obtain that ω and $\mathbf{L}$ have to satisfy the following equation:

$$(\mathbf{A} - \omega^2\mathbf{1})\mathbf{L} = \mathbf{0}. \tag{7.9}$$

The values of ω^2 represent the eigenvalues,[38] while the $\mathbf{L}$ are the eigenvectors of the matrix $\mathbf{A}$. There are $3N$ eigenvalues, and each of them corresponds to its eigenvector $\mathbf{L}$. This means that we have $3N$ normal modes, each mode characterized by its angular frequency $\omega = 2\pi\nu$ (ν is the frequency) and its vibration amplitudes $\mathbf{L}$. Hence, it would be natural to assign a normal mode index $k = 1, \ldots, 3N$ for ω and $\mathbf{L}$. Therefore we have

$$\boxed{(\mathbf{A} - \omega_k^2\mathbf{1})\mathbf{L}_k = 0. \tag{7.10}}$$

The diagonalization of $\mathbf{A}$ (p. 703) is an efficient technique for solving the eigenvalue problem using commercial computer programs (diagonalization is equivalent to a rotation of the coordinate system, see Fig. 7.7).

---

[36] This form (with $\omega = a + ib$) allows for a constant solution ($a = b = 0$), an exponential growth or vanishing ($a = 0, b \neq 0$), oscillations ($a \neq 0, b = 0$), or oscillatory growing or oscillatory vanishing ($a \neq 0, b \neq 0$). For $\mathbf{R}_0$ denoting a minimum $\det \mathbf{A} > 0$ and this ensures a solution with $a \neq 0, b = 0$.

[37] Equal to $2i\mathbf{c}_1$, but since $\mathbf{c}_1$ is unknown, as for the time being is $\mathbf{L}$, we can say goodbye to $\mathbf{c}_1$ without feeling any discomfort whatsoever.

[38] $\mathbf{A}$ is a symmetric matrix, hence its eigenvalues ω^2 and therefore also $\omega = a + ib$ are real ($b = 0$). Whether ω are positive, negative, or zero depends on the hypersurface V at $\mathbf{R}_0$ (see Fig. 7.7).

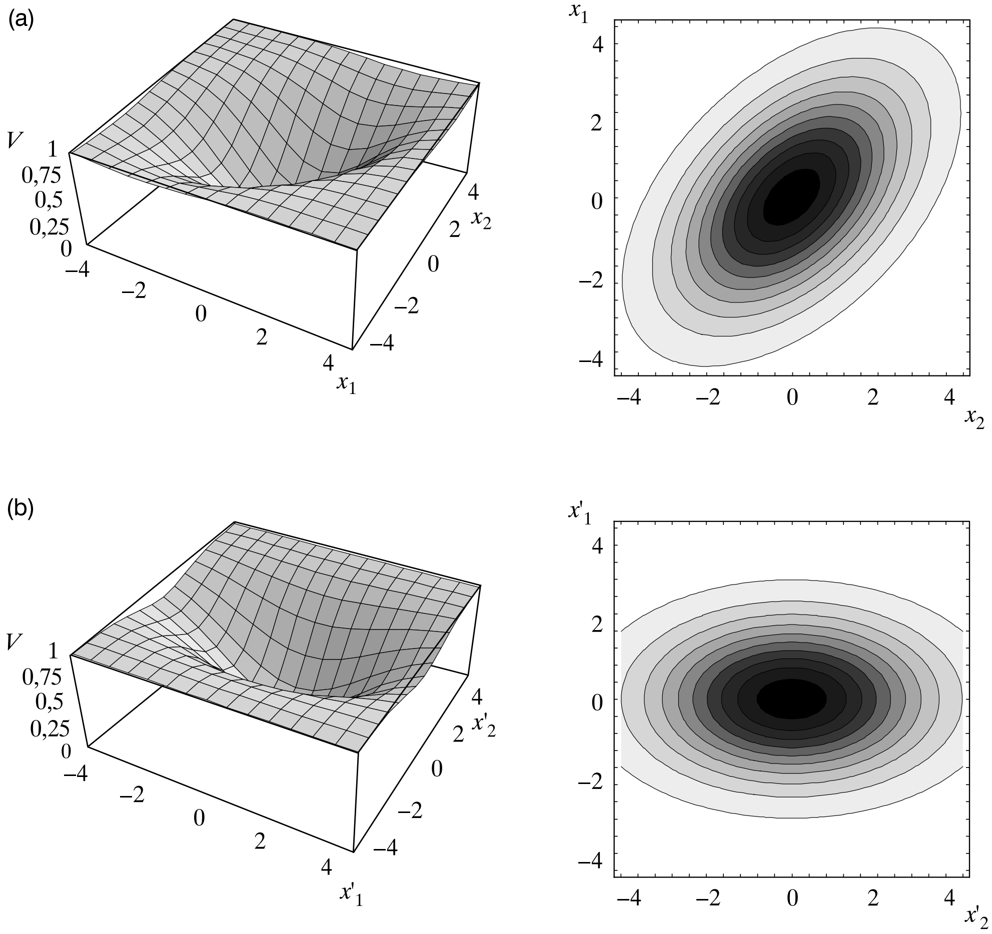


Fig. 7.7. A ball oscillating in a potential energy well (scheme). (a) and (b) The normal vibrations (normal modes) about a point $\mathbf{R}_0 = \mathbf{0}$ being *a minimum* of the potential energy function $V(\mathbf{R}_0 + \mathbf{x})$ of two variables $\mathbf{x} = (x_1, x_2)$. This function is first approximated by a quadratic function, i.e., a paraboloid $\tilde{V}(x_1, x_2)$. Computing the normal modes is equivalent to such a *rotation of the Cartesian coordinate system* (a), that the new axes (b) x_1' and x_2' become the principal axes of any section of $\tilde{V}$ by a plane $\tilde{V} = const$ (i.e., ellipses). Then we have $\tilde{V}(x_1, x_2) = V(\mathbf{R}_0 = \mathbf{0}) + \frac{1}{2}k_1\left(x_1'\right)^2 + \frac{1}{2}k_2\left(x_2'\right)^2$. The problem then becomes equivalent to the two-dimensional harmonic oscillator (cf. Chapter 4) and separates into two independent one-dimensional oscillators (normal modes): one of angular frequency $\omega_1 = 2\pi\nu_1 = \sqrt{\frac{k_1}{m}}$ and the other with angular frequency $\omega_2 = 2\pi\nu_2 = \sqrt{\frac{k_2}{m}}$, where m is the mass of the oscillating particle. (c) and (d) Here we show what would happen if $\mathbf{R}_0$ corresponded not to a minimum, but to (c) a maximum or (d) the saddle point. For a maximum (c) k_1 and k_2 in $\tilde{V}(x_1', x_2') = V(\mathbf{0}) + \frac{1}{2}k_1\left(x_1'\right)^2 + \frac{1}{2}k_2\left(x_2'\right)^2$ would be both *negative*, and therefore the corresponding normal "vibrations" would have both imaginary frequencies, while for the saddle point (d) only one of the frequencies would be imaginary.

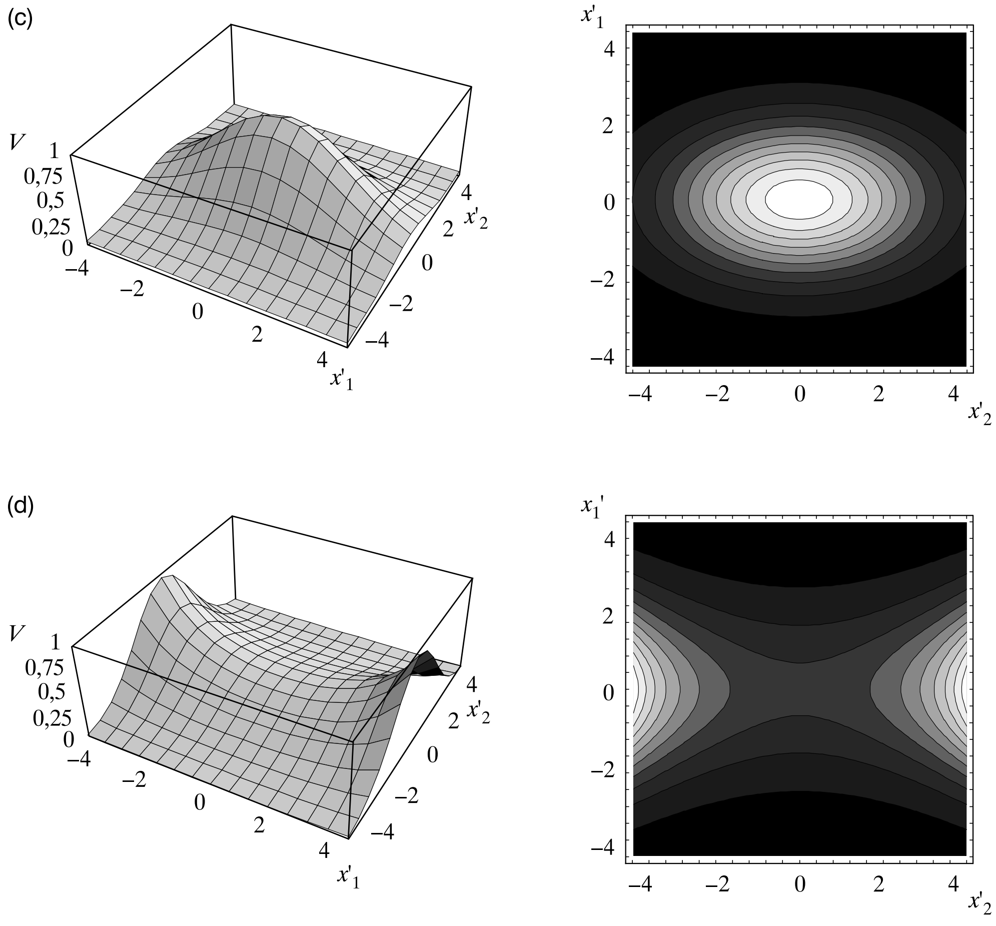


Fig. 7.7. (*continued*)

This is equivalent to replacing V by a $3N$-dimensional paraboloid with origin $\mathbf{R}_0$. The normal mode analysis means such a rotation of the coordinate system that it will make the new axes coincide with the principal axes of the paraboloid.

There will be six frequencies (five for a linear molecule) equal to zero. They are connected to the translation and rotation of the molecule in space: three translations along x, y, z and three rotations about x, y, z (two in the case of a linear molecule). Such free translations/rotations do not change the energy and may therefore be thought of as corresponding to zero force constants.

If we are interested in what the particular atoms are doing when a single mode k is active, then the displacements from the equilibrium position as a function of time are expressed as

$$\mathbf{x}_k = \mathbf{M}^{-\frac{1}{2}}\mathbf{y}_k = \mathbf{M}^{-\frac{1}{2}}\mathbf{L}_k Q_k \sin(\omega_k t), \tag{7.11}$$

where $Q_k \in (-\infty, \infty)$, because the displacements $\mathbf{x}_k$ are not bound by the length of vector $\mathbf{L}_k$, which has been arbitrarily set to 1. Therefore, if one shifts all atoms according to how they behave in the normal mode k, one should insert the variable Q_k, which tunes the displacement amplitude.

A given atom participates in all vibrational modes. Even if any vibrational mode makes all atoms move, some atoms move more than others. It may happen that a particular mode changes mainly the length of *one of the chemical bonds (stretching mode)*, another mode moves another bond, another changes a particular *bond angle (bending mode)*, etc.

> This means that some chemical bonds or some functional groups may have *characteristic* vibration frequencies, which is of great importance for the identification of these bonds or groups in spectral analysis.

Table 7.1. Characteristic frequencies (wave numbers, in cm^{-1}) typical for some chemical bonds (stretching vibrations) and bond angles (bending vibrations). This is of great importance for chemical analysis.

Bond	Vibration	Wave number
C–H	stretching	2850–3400
H–C–H	bending	1350–1460
C–C	stretching	700–1250
C=C	stretching	1600–1700
C≡C	stretching	2100–2250
C=O	stretching	1600–1750
N–H	stretching	3100–3500
O–H	stretching	3200–4000

In Table 7.1 typical ("*characteristic*") frequencies for some particular chemical bonds are reported. The characteristic frequencies are of great importance in chemical analysis, where the wave numbers $\bar{\nu}$ are defined by the relation

$$\omega = 2\pi\nu = 2\pi\bar{\nu}c, \tag{7.12}$$

with c being the velocity of light and ν the frequency. The wave number is the number of the wave lengths covering a distance of 1 cm.

Example 1 (Water molecule). A single water molecule has $3 \times 3 = 9$ normal modes. Six of them would have the angular frequencies ω equal zero (they correspond to three free translations and three free rotations of the molecule in space). Three normal modes remain; the vectors $\mathbf{x}$ of Eq. (7.11) for these modes can be described as follows (Fig. 7.8, the corresponding wave numbers have been given in parentheses[39]):

- one of the modes means a *symmetric* stretching of the two OH bonds ($\bar{\nu}_{sym} = 3894$ cm^{-1}),
- the second mode corresponds to a similar, but *antisymmetric* motion, i.e., when one of the bonds shortens and the other one extends and *vice versa*[40] ($\bar{\nu}_{asym} = 4029$ cm^{-1}),
- the third mode is called the bending mode and corresponds to an oscillation of the HOH angle about the equilibrium value ($\bar{\nu}_{bend} = 1677$ cm^{-1}).

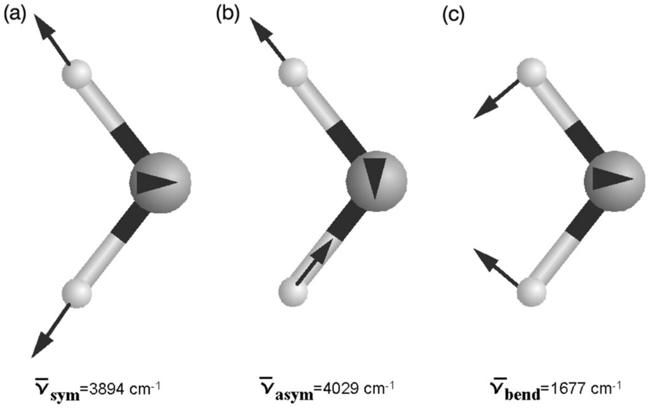


Fig. 7.8. The normal modes of the water molecule: (a) symmetric, (b) antisymmetric, and (c) bending. The arrows indicate the directions and proportions of the atomic displacements.

Example 2 (Water dimer). Now let us take *two interacting water molecules*. First, let us ask how many minima we can find on the electronic ground-state energy hypersurface. Detailed calculations have shown that there are two such minima (Fig. 7.9). The global minimum corresponds to the configuration characteristic for the *hydrogen bond* (cf. p. V2-415). One of the molecules is a donor, the other is an acceptor of a proton (Fig. 7.9a). A local minimum of smaller stability appears when one of the water molecules serves as a donor of two protons, while the other serves as an acceptor of them called the *bifurcated* hydrogen bond[41] (Fig. 7.9b).

---

[39] J. Kim, J.Y. Lee, S. Lee, B.J. Mhin, K.S. Kim, *J. Chem. Phys.*, 102(1995)310. This paper reports normal mode analysis for potential energy hypersurfaces computed by various methods of quantum chemistry. I have chosen the coupled cluster method (see Chapter V2-2) CCSD(T) as an illustration.

[40] The shortening has the same value as the extension. This is a result of the harmonic approximation, in which both shortening and lengthening require the same energy.

[41] See, e.g., a theoretical analysis given by R.Z. Khaliullin, A.T. Bell, M. Head-Gordon, *Chem. Eur. J.*, 15(2009)851 at the DFT level (see Chapter V2-3).

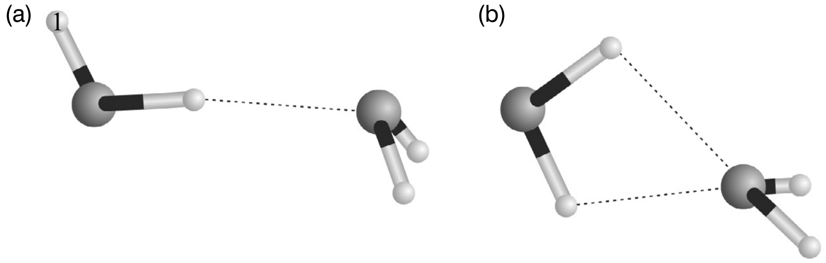


Fig. 7.9. The water dimer and the configurations of the nuclei that correspond to minima of the two basins of the potential energy V. (a) The global minimum corresponds to a single hydrogen bond O–H$\cdots$O. (b) The local minimum corresponds to the bifurcated hydrogen bond.

Now, *we decide* to focus on the global minimum potential well. We argue that for thermodynamic reasons, this particular well will be most often represented among water dimers. This potential energy well has to be approximated by a paraboloid. The number of degrees of freedom is equal to $6 \times 3 = 18$ and this is also the number of normal modes to be obtained. As in Example 2, six of them will have zero frequencies and the number of "true" vibrations is 12. This is the number of normal modes, each with its frequency ω_k and the vector $\mathbf{x}_k = \mathbf{M}^{-\frac{1}{2}}\mathbf{L}_k \sin(\omega_k t)$ that describes the atomic motion. The two water molecules, after forming the hydrogen bond, have not lost their individual features (in other words the OH vibration is *characteristic*).

In dimer vibrations we will find the vibration frequencies of individual molecules changed a little by the water–water interaction. These modes should appear in pairs, but the two frequencies should differ (the role of the two water molecules in the dimer is different). The proton acceptor has something attached to its heavy atom, the proton donor has something attached to the light hydrogen atom. Let us recall that in the harmonic oscillator, the reduced mass is relevant, which therefore is almost equal to the mass of the *light* proton. If something attaches to this atom, it means a considerable lowering of the frequency. This is why lower frequencies correspond to the proton donor. The computed frequencies[42] are the following:

- two stretching vibrations with frequencies 3924 cm^{-1} and 3904 cm^{-1} (both antisymmetric, the higher frequency corresponds to the proton acceptor, the lower to the proton donor);
- two stretching vibrations with frequencies 3796 cm^{-1} and 3704 cm^{-1} (both symmetric, again the higher frequency corresponds to the proton acceptor, the lower to the proton donor);
- two bending vibrations with frequencies 1624 cm^{-1} (donor bending) and 1642 cm^{-1} (acceptor bending).

---

[42] R.J. Reimers, R.O. Watts, *Chem. Phys.*, 85(1984) 83.

Thus, among 12 modes of the dimer we have discovered six modes which are related to the individual molecules: four O–H stretching and two H–O–H bending modes. Now, we have to identify the remaining six modes. These are the intermolecular vibrations (Fig. 7.9a):

- stretching of the hydrogen bond O–H$\cdots$O (the vibration of two water molecules treated as entities): 183 cm^{-1};
- bending of the hydrogen bond O–H$\cdots$O in the plane of the figure: 345 cm $^{-1}$;
- bending of the hydrogen bond O–H$\cdots$O in the plane perpendicular to the figure: 645 cm^{-1};
- rocking of the hydrogen atom H_1 perpendicular to the figure plane: 115 cm^{-1};
- rocking of the second water molecule (the right-hand side of the figure) in the figure plane: 131 cm^{-1};
- rocking of the second water molecule (the right-hand side of the figure) about its symmetry axis: 148 cm^{-1}.

As we can see, the intermolecular interactions have made the "intramolecular" vibration frequencies decrease,[43] while the "intermolecular" frequencies have very low frequencies. The last effect is, of course, nothing strange, because a change of intermolecular distances does require a small expenditure of energy (which means small force constants). Note that the simple Morse oscillator model considered in Chapter 4 (p. 228) gave the correct order of magnitude of the intermolecular frequency of two water molecules (235 cm^{-1} as compared to the above, much more accurate, result 183 cm^{-1}).

Example 3 (The formaldehyde molecule $H_2C{=}O$). Each molecule of this useful substance is planar and consists of only four atoms; therefore the number of the normal vibrational modes should be equal to $3 \times 4 - 6 = 6$. The quantum mechanical calculations based on the ground-state electronic energy as a function of the nuclear coordinates[44] gave indeed six normal modes. Fig. 7.10 shows the experimental IR absorption spectrum. One sees six light frequencies ("lines") for which a significant absorption occurs. Each of these lines is interpreted as an excitation of one of the six normal oscillators (modes),[45] each corresponding to transition from the ground vibrational state with $v = 0$ to the first excited state with $v = 1$ of the corresponding harmonic oscillator. We have already some experience and may predict that the highest frequencies will correspond to the C–H stretching vibrations in CH_2, because of the

---

[43] This is how the hydrogen bonds behave. This seemingly natural expectation after attaching an additional mass to a vibrating system is legitimate when assuming that the force constants have not increased. An interesting example of the opposite effect for a wide class of compounds has been reported by Pavel Hobza and Zdenek Havlas (P. Hobza, Z. Havlas, *Chem. Rev.*, 100(2000)4253).

[44] I have used the Gaussian package with the coupled cluster method (CCSD(T), a procedure known as a rather reliable one [see Chapter V2-2]; also a reasonable quality of the atomic basis set 6-311G(d,p) was used [see Chapter 8]).

[45] As one can see these spectral lines exhibit certain widths (related to various accompanying rotational excitations).

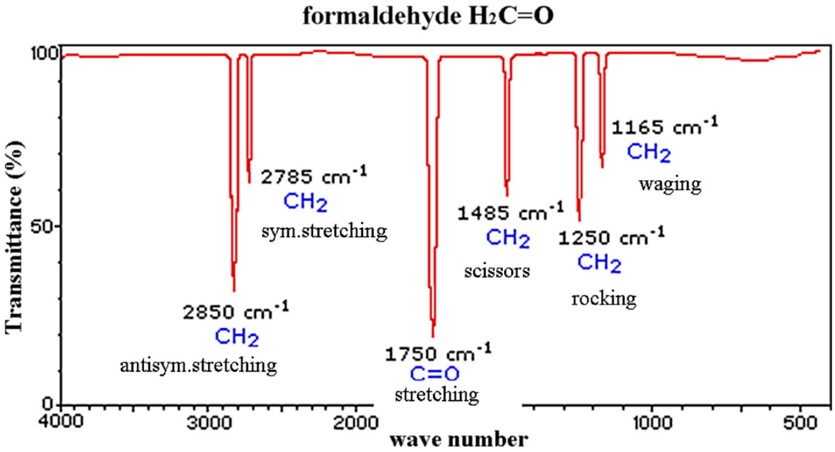


Fig. 7.10. The vibrational modes concept is related to the experimental IR spectrum. An example of the formaldehyde molecule (a planar $H_2C=O$ system). The number of vibrational modes $3N - 6 = 6$ is identical with the number of absorption lines (each line corresponding to the vibrational excitation $v = 0 \to 1$). If someone is interested in the intensity of these absorptions, one would need to compute some integrals involving the corresponding vibrational wave functions.

low mass of hydrogen atoms. One of the modes is symmetric (calculation: 2934 cm^{-1}, experiment: 2785 cm^{-1}), the other antisymmetric (calculation: 2999 cm^{-1}, experiment: 2850 cm^{-1}), similarly as for the water molecule. We expect a stretching vibration of the C=O bond to have a much lower frequency, because of much heavier atoms involved (calculated: 1769 cm^{-1}, experiment: 1750 cm^{-1}). Bending vibrations are of lower energy than the stretching ones (for atoms of comparable masses). Indeed, for the scissor-like mode of CH_2 the computation gives 1546 cm^{-1}, while the experimental result is 1485 cm^{-1}. Next, the rocking vibration of CH_2 predicted by theory (that moves the CH_2 group within the molecular plane) has a frequency of 1269 cm^{-1}, whereas measurement gives 1250 cm^{-1}. Finally, a wagging motion (that moves the entire CH_2 group out of the molecular plane) corresponds to the calculated 1183 cm^{-1} and the measured 1165 cm^{-1} frequencies.

7.5.2 Zero-vibration energy

The computed minimum of V (using any method, either quantum mechanical or force field) does not represent the energy of the system for exactly the same reason why the bottom of the parabola (the potential energy) does not represent the energy of the harmonic oscillator (cf. p. 217). The reason is the kinetic energy contribution.

If all the normal oscillators are in their ground states ($v_j = 0$, called the "zero-vibrations"), then the energy of the system is the energy of the bottom of the parabola V_{min} plus the zero-vibration energy (we assume no rotational contribution)

$$E = V_{\text{min}} + \frac{1}{2}\sum_j \left(h\nu_j\right). \tag{7.13}$$

In the above formula it has been assumed that the vibrations are harmonic. This assumption usually makes the frequencies higher by several percent (cf. p. 228).

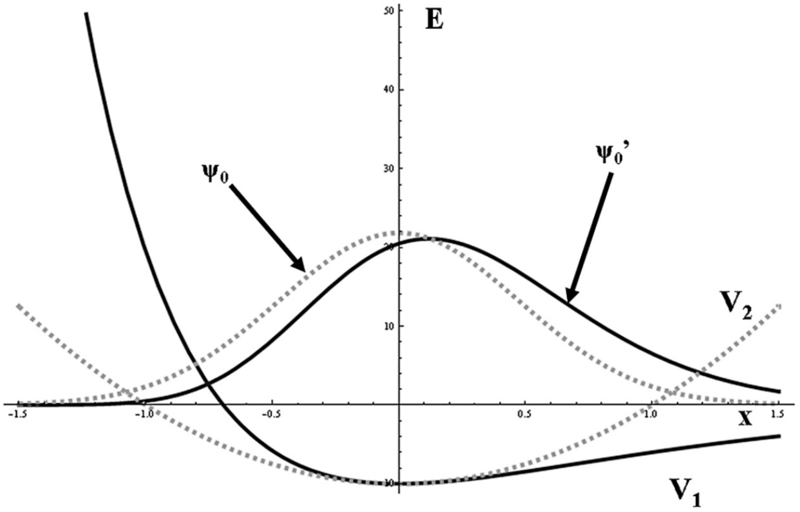


Fig. 7.11. The ground-state vibrational wave function ψ_0' of the anharmonic oscillator (of potential energy V_2) is *asymmetric and shifted* towards positive values of the displacement when compared to the wave function ψ_0 for the harmonic oscillator with the same force constant (the potential energy V_1).

Taking anharmonicity into account is a much more difficult task than normal mode analysis. Note (Fig. 7.11) that in such a case the position of the minimum of V does not correspond to the mean value of the interatomic distance due to the asymmetry of V.

7.6 Molecular dynamics

In all the methods described above there is no such a thing as temperature. It looks as if all experiments were conducted at 0 K. It is difficult to tolerate such a situation.

This is what molecular dynamics is for. The idea is very simple.

If we knew the potential energy V as a function of the position ($\mathbf{R}$) of all the atoms (we may think here about a force field as its approximation[46]), then all the forces the atoms undergo

---

[46] See p. 411.

could be easily computed. If $\mathbf{R} = (X_1, X_2, \ldots . X_{3N})^T$ denotes the coordinates of all the N atoms (X_1, X_2, X_3 are the x, y, z coordinates of atom 1, X_4, X_5, X_6 are the x, y, z coordinates of atom 2, etc.), then $-\frac{\partial V}{\partial X_1}$ is the x component of the force atom 1 undergoes, $-\frac{\partial V}{\partial X_2}$ is the y component of the same force, etc. When a force field is used, all this can be easily computed even analytically.[47] We had the identical situation in molecular mechanics, but there we were just interested in making these forces equal to zero (through obtaining the equilibrium geometry). In molecular dynamics we are interested in time t, the velocity of the atoms (in this way temperature will come into play), and the acceleration of the atoms.

> Our immediate goal is collecting the atomic positions as functions of time, i.e., of the system trajectory.

The Newton equation tells us that, knowing the force acting on a body (e.g., an atom), we may compute the acceleration the body undergoes. We have to know the mass, but there is no problem with that.[48] Hence the i-th component of the acceleration vector is equal to

$$a_i = -\frac{\partial V}{\partial X_i} \cdot \frac{1}{M_i} \tag{7.14}$$

for $i = 1, 2, \ldots, 3N$ ($M_i = M_1$ for $i = 1, 2, 3$, $M_i = M_2$ for $i = 4, 5, 6$, etc.).

Now, let us assume that at $t = 0$ all the atoms have the initial coordinates[49] $\mathbf{R}_0$ and the initial velocities $\mathbf{v}_0$. Now we assume that the forces calculated act *unchanged* during a short period Δt (often 1 femtosecond or 10^{-15} sec). We know what should happen to a body (atom) if under influence of a constant force during time Δt. Each atom undergoes a uniformly variable motion, the new position may be found in the vector

$$\mathbf{R} = \mathbf{R}_0 + \mathbf{v}_0 \Delta t + \mathbf{a}\frac{\Delta t^2}{2}, \tag{7.15}$$

---

[47] That is, an analytical formula can be derived.

[48] We assume that what moves is the nucleus (see the discussion on p. 325). In molecular dynamics we do not worry about the fact that the nucleus moves together with its electrons, because the electron mass is about 1840 times smaller than the proton mass. Thus, the mass is negligible in view of other approximations which have been made.

[49] Where could these coordinates be taken from? To tell the truth, almost from a "hat." "Almost" – because some essential things will be ensured. First, we may quite reasonably conceive the geometry of a molecule, because we know which atoms form the chemical bonds, their reasonable lengths, the reasonable values of the bond angles, etc. That is, however, not all we would need for larger molecules. What do we take as dihedral angles? This is a difficult case. Usually we take a conformation which we could call "reasonable." In a minute we will take a critical look at this problem. The next question is the velocities. Having nothing better at our disposal, we may use a random number generator, ensuring however that the velocities are picked out according to the Maxwell–Boltzmann distribution suitable for a given temperature T of the laboratory, e.g., 300 K. In addition, we will make sure that the system does not rotate or fly off somewhere.
In this way we have our starting position and velocity vectors $\mathbf{R}_0$ and $\mathbf{v}_0$.

and its new velocity in the vector

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{a}\Delta t, \tag{7.16}$$

where the acceleration **a** is a vector composed of the acceleration vectors of all the N atoms

$$\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_N)^T, \tag{7.17}$$

$$\begin{aligned} \mathbf{a}_1 &= (-\frac{\partial V}{\partial X_1}, -\frac{\partial V}{\partial X_2}, -\frac{\partial V}{\partial X_3}) \cdot \frac{1}{M_1}, \\ \mathbf{a}_2 &= (-\frac{\partial V}{\partial X_4}, -\frac{\partial V}{\partial X_5}, -\frac{\partial V}{\partial X_6}) \cdot \frac{1}{M_2}, \\ &\text{etc.} \end{aligned}$$

Everything on the right-hand side of (7.15) and (7.16) is known. Therefore the new positions and the new velocities are easy to calculate.[50] Now, we may use the new positions and velocities as starting ones and repeat the whole procedure over and over. This makes it possible to go along the time axis in a step-like way, in practice reaching even nanosecond times (10^{-9} sec), which means millions of such steps. The procedure described above simply represents the numerical integration of $3N$ differential equations. If $N = 2000$, then the task is impressive. It is so straightforward, because we are dealing with a numerical (not analytical) solution.[51]

7.6.1 What does molecular dynamics offer us?

The computer simulation makes the system evolve from the initial state to the final one. The position **R** in $3N$-dimensional space becomes a function of time and therefore $\mathbf{R}(t)$ represents the trajectory of the system in the configurational space. A similar statement pertains to $\mathbf{v}(t)$. Knowing the trajectory means that we know the smallest details of the motion of all the atoms. Within the approximations used, we can therefore answer any question about this motion. For example, we may ask about some mean values, like the mean value of the total energy, potential energy, or kinetic energy, or the distance between atom 4 and atom 258, etc. All these quantities may be computed at any step of the procedure, added up, and then divided by the number of steps, giving the mean values we require. In this way we may obtain the theoretical prediction of the mean value of the interatomic distance and then compare it to, say, the nuclear magnetic resonance result.

---

[50] In practice we use a more accurate computational scheme called the *leap frog algorithm*.

[51] By the way, if somebody gave us the force field for galaxies (this is simpler than for molecules), we could solve the problem as easily as in our case. This is what astronomers often do.

In this way we may search for some correlation of motion of some atoms or groups of atoms, i.e., the *space correlation* ("when this group of atoms is shifted to the left, then another group is most often shifted to the right") or the *time correlation* ("when *this* thing happens to the functional group G_1, then after a time τ *that* most often takes place with another functional group G_2") or *time autocorrelation* ("when *this* happens to a certain group of atoms, then after time τ *the same* most often happens to the same group of atoms"). For example, is the x coordinate of atom 1, i.e., X_1, correlated to the y coordinate of atom 41, i.e., X_{122}, or are these two quantities independent? The answer to this question is given by the correlation coefficient $c_{1,122}$ calculated for M simulation steps in the following way:

$$c_{1,122} = \frac{\frac{1}{M}\sum_{i=1}^{M}\left(X_{1,i} - \langle X_1\rangle\right)\left(X_{122,i} - \langle X_{122}\rangle\right)}{\sqrt{\left(\frac{1}{M}\sum_{i=1}^{M}\left(X_{1,i} - \langle X_1\rangle\right)^2\right)\left(\frac{1}{M}\sum_{i=1}^{M}\left(X_{122,i} - \langle X_{122}\rangle\right)^2\right)}},$$

where $\langle X_1\rangle$ and $\langle X_{122}\rangle$ denote the mean values of the coordinates indicated, and the summation goes over the simulation steps. It is seen that any deviation from independence means a nonzero value of $c_{1,122}$. What could be more correlated to the coordinate X_1 than the same X_1 (or $-X_1$)? Of course, absolutely nothing. In such a case (in the formula we replace $X_{122,i} \rightarrow X_{1,i}$ and $\langle X_{122}\rangle \rightarrow \langle X_1\rangle$), we obtain $c_{1,1} = 1$ or -1. Hence, c always belongs to $[-1, 1]$, where $c = 0$ means independence and $c \pm 1$ means maximum dependence.

Has molecular dynamics anything to do with reality?

If the described procedure were applied without any modification, then most probably we would have bad luck and our $\mathbf{R}_0$ would be located on a slope of the hypersurface V. Then the solution of the Newton equations would reflect what happens when a point (representing the system) is placed on a slope – it will slide downhill. The point would go faster and faster and soon the vector $\mathbf{v}$ would not correspond to the room temperature, but, say, to 500 K. Of course, despite such a high temperature the molecule would not disintegrate, because this is not a real molecule but one operating with a force field that usually corresponds to unbreakable chemical bonds.[52] Although the molecule will not fall apart, such a large T has nothing to do with the temperature of the laboratory. This suggests that after some number of steps we should check whether the atomic velocities still correspond to the proper temperature. If not, it is recommended to scale all the velocities by multiplying them by a factor in order to make them corresponding again to the desired temperature. For this reason, the only goal of the first part of a molecular dynamics simulation is known as "thermalization," in which the error connected to the nonzero Δt is averaged and the system is forced stepwise (by scaling) to behave as what is called the canonical ensemble. The canonical ensemble preserves the number of molecules, the

---

[52] This pertains to a single molecule bound by chemical bonds; a system of several molecules could fall apart.

volume, and the temperature (simulating contact with a thermostat at temperature T). In such a "thermalized" system total energy fluctuations are allowed.

The thermalization done, the next (main) stage of molecular dynamics, i.e., the harvesting of data (trajectory), begins.

7.6.2 What to worry about?

- During simulation, the system has to have enough time to wander through all parts of the phase space[53] that are accessible in the experimental conditions (with which the simulation is to be compared). We are never sure that it happens. We have to check whether the computed mean values depend upon the simulation time. If they do not, then probably everything is in order – we have a description of the equilibrium state.
- The results of the molecular dynamics (the mean values) should not depend on the starting point, because it has been chosen arbitrarily. This is usually satisfied for small molecules and their collections. For large and flexible molecules we usually start from the vector $\mathbf{R}_0$ found from X-ray-determined atomic positions. Why? Because *after* the molecular dynamics we will still stay close to this (all in all experimental) conformation. If the simulation started from another conformation, it would result in a conformation close to this new starting point. This is because even with the most powerful computers, simulation times are too short. In such a way we have a simulation of one conformation evolution rather than a description of the thermodynamic equilibrium.
- The simulation time in molecular dynamics is limited on one side by the power of computers and on the other side by the time step Δt, which is not infinitesimally small, and creates an error that cumulates during the simulation (as a result the total energy may vary too much and the system may be heading into nonphysical regions of the phase space).

7.6.3 Molecular dynamics of nonequilibrium processes

Thermalization is not always what we want. We may be interested in what happens when a DNA molecule being anchored to a solid surface by one of its functional end groups is distorted by pulling the other end of the molecule. Such molecular dynamics results may nowadays be compared to the corresponding experiment.

Another example. A projectile hits a wall ("armor"), Fig. 7.12. The projectile is being composed of Lennard-Jones atoms (with some ε_p and $r_{e,p}$, p. 409); we assume the same for the wall (for

---

[53] The Cartesian space of all atomic positions and momenta.

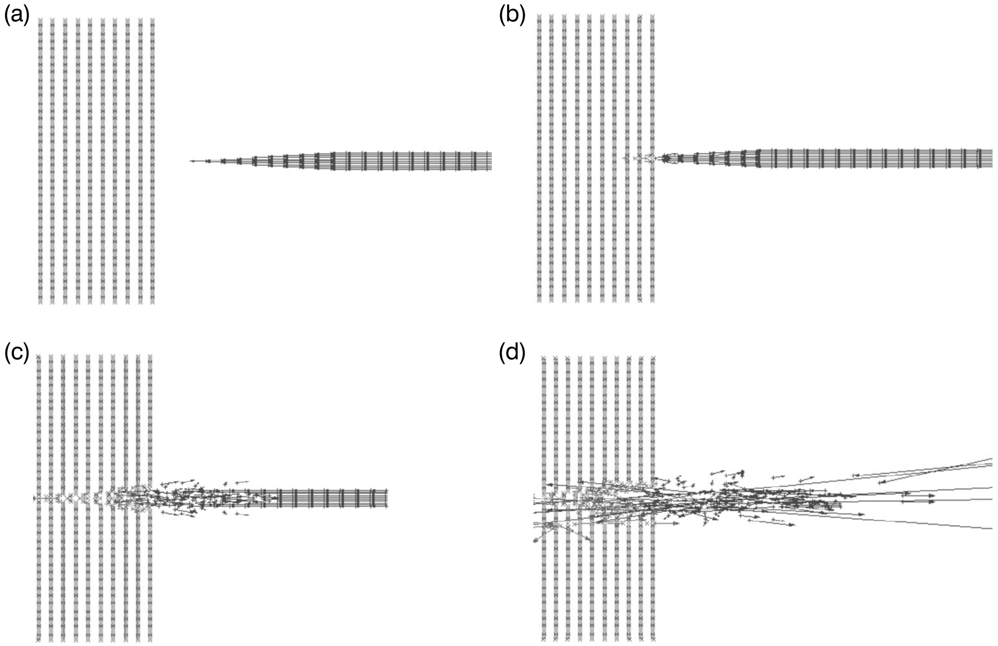


Fig. 7.12. A projectile hitting an armor – an example of simulating nonequilibrium processes in molecular dynamics. (a) On the right side is a projectile flying left, on the left side the armor, both objects built of particles ("atoms") that interact according to the Lennard-Jones interatomic potential. The Lennard-Jones parameters for the projectile atoms are $\varepsilon = 0.2$ a.u., $r_e = 2.3$ a.u., those of the armor are $\varepsilon = 0.1$ a.u., $r_e = 2.6$ a.u. (the projectile–armor interaction parameters correspond to their mean values). The projectile atoms are denoted by "+," the armor atoms by "×." At the starting point shown the armor atoms all have a velocity equal to zero, while all projectile atoms have only the x component of velocity (x is the left–right axis), which is equal to $v_x = -55\,000\frac{km}{h}$ (i.e., the missile is heading right into the armor). (b) An early stage of the hit. The top of the missile undergoes a destruction; the armor atoms in the epicenter accelerate (arrows) and change their position. It is remarkable that this motion is for the time being limited only to some small region and that one can see something which is similar to an intact missile ("a virtual missile"!), which is built, however, from the atoms of both colliding objects! (c) Half of the missile is already destroyed; one can see quite a lot of particles that bounce off the armor and go back very fast, and the missile was able to crush the armor width. In spite of this, the abovementioned "virtual missile" is still visible. (d) The armor is destroyed, a lot of debris is flying off; one does not see any virtual missile anymore. The results were obtained by Mr. Marcin Gronowski under the supervision of the author, during an undergraduate course at the Chemistry Department, University of Warsaw.

other values of the parameters, let us make the wall less resistant than the projectile: $\varepsilon_w < \varepsilon_p$ and $r_{e,w} > r_{e,p}$). Altogether we may have hundreds of thousands or even millions of atoms (i.e., millions of differential equations to solve). Now, we prepare the input $\mathbf{R}_0$ and $\mathbf{v}_0$ data. The wall atoms are assumed to have stochastic velocities drawn from the Maxwell–Boltzmann distribution for room temperature. The same for the projectile atoms, but additionally they have a constant velocity component along the direction pointing to the wall. At first, nothing particularly interesting happens – the projectile flies towards the wall with a constant velocity (while all the atoms of the system vibrate). Of course, the time the projectile hits the wall is the most interesting. Once the front part of the projectile touches the wall, the wall atoms burst into space in a kind of eruption, the projectile's tip looses some atoms as well, and the spot on the wall hit by the projectile vibrates and sends a shock wave and concentric waves traveling within the wall. A violent (and instructive) movie.

Among more civil applications, we may think of the interaction of a drill and a steel plate, to plan better drills and better steel plates, as well as about other microtools which have a bright future.

7.6.4 Quantum classical molecular dynamics

Typical molecular dynamics does not allow for breaking chemical bonds and the force fields which allow this give an inadequate, classical picture, so a quantum description is sometimes a must. The systems treated by molecular dynamics are usually quite large, which excludes a full quantum mechanical description.

For enzymes (natural catalysts) researchers proposed[54] joining the quantum and the classical description by making the precision of the description dependent on how far the region of focus is from the enzyme active center (where the reaction the enzyme facilitates takes place). They proposed dividing the system enzyme + solvent into three regions:

- region I represents the active center atoms,
- region II is composed of the other atoms of the enzyme molecule,
- region III is the solvent.

Region I is treated as a quantum mechanical object and described by the proper time-dependent Schrödinger equation, region II is treated classically by the force field description and the corresponding Newton equations of motion, and region III is simulated by a continuous medium (no atomic representation) with a certain dielectric permittivity.

The regions are coupled by their interactions: the quantum mechanical region I is subject to the external electric field produced by region II evolving according to its molecular dynamics

---

[54] P. Bała, B. Lesyng, J.A. McCammon, in "*Molecular Aspects of Biotechnology: Computational Methods and Theories*," Kluwer Academic Publishers, p. 299(1992). A similar philosophy stands behind the Morokuma ONIOM procedure: M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, K. Morokuma, *J. Phys. Chem.*, 100(1996)19357.

as well as that of region III, region II feels the charge distribution changes region I undergoes through electrostatic interaction.

7.7 Simulated annealing

The goal of molecular dynamics may differ from simply computing some mean values, e.g., we may try to use molecular dynamics to find regions of the configurational space for which the potential energy V is particularly low.[55] From a chemist's point of view, this means trying to find a particularly stable structure (conformation of a single molecule or an aggregate of molecules). To this end, molecular dynamics is sometimes coupled with an idea of Kirkpatrick et al.,[56] taken from an ancient method of producing metal alloys of exceptional quality (the famous steel of Damascus) and serving to find the minima of arbitrary functions. The idea behind simulated annealing is simple.

This goal is achieved by a series of heating and cooling procedures (called the annealing protocol). First, a starting configuration is chosen that, to the best of our knowledge, is of low energy and the molecular dynamics simulation is performed at a high temperature T_1. As a result, the system (represented by a point $\mathbf{R}$ in the configuration space) rushes through a large manifold of configurations $\mathbf{R}$, i.e., wanders over a large portion of the hypersurface $V(\mathbf{R})$. Then, a lower temperature T_2 is chosen and the motion slows down, the visited portion of the hypersurface shrinks and hopefully corresponds to some regions of low values of V – the system is confined in a large superbasin (the basin composed of individual minima basins). Now the temperature is raised to a certain value $T_3 < T_1$, thus allowing the system eventually to leave the superbasin and to choose another one, maybe of lower energy. While the system explores the superbasin, it is cooled again, this time to temperature $T_4 < T_2$, and so forth. Such a procedure does not give any guarantee of finding the global minimum of V, but there is a reasonable chance of getting a configuration with lower energy than the start. The method, being straightforward to implement, is very popular. Its successes are spectacular, although sometimes the results are disappointing. The highly prized swords made in ancient Damascus using annealing prove that the metal atoms settle down in quasioptimal positions, forming a solid state of low energy, very difficult to break or deform.

---

[55] Like in the global molecular mechanics.

[56] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, *Science*, 220(1983)671.

7.8 Langevin dynamics

Paul Langevin (1872–1946), French physicist, professor at the Collège de France. His main achievements are in the theory of magnetism and in relativity theory. His PhD student Louis de Broglie made a breakthrough in quantum theory.

In molecular dynamics we solve Newton equations of motion for all atoms of the system. Imagine we have a large molecule in an aqueous solution (biology offers us important examples). We have no chance to solve Newton equations because there are too many of them (a lot of water molecules). What do we do then? Let us recall that we are interested in the macromolecule; the water molecules are interesting only as a medium that changes the conformation of the macromolecule. The changes may occur for many reasons, but the simplest is the most probable – just the fact that the water molecules in their thermal motion hit the atoms of the macromolecule. If so, their role is reduced to a source of chaotic strikes. The main idea behind Langevin dynamics is to ensure that the atoms of the macromolecule indeed feel some random hits from the surrounding medium *without taking this medium into consideration explicitly. This is the main advantage of the method.*

A reasonable part of this problem may be incorporated into the Langevin equation of motion:

$$M_i \ddot{X}_i = -\frac{\partial V}{\partial X_i} + F_i\,(t) - \gamma_i M_i \dot{X}_i, \tag{7.18}$$

for $i = 1, 2, \ldots, 3N$, where besides the force $-\nabla V$ resulting from the potential energy V for the macromolecule alone, we have an additional stochastic force $\mathbf{F}\,(t)$, the magnitude and direction of which, keeping the force related to the temperature and ensuring its isotropic character, are drawn. The coefficient γ_i is a friction coefficient and the role of friction is proportional to atomic velocity.

The Langevin equations are solved in the same way as those of molecular dynamics, with the additional stochastic force drawn using a random number generator.

7.9 Monte Carlo dynamics

Las Vegas, Atlantic City, and Monte Carlo are notorious among upright citizens for day and night use of random number generators such as billiards, roulette, and card games.

Because of this, the idea and even the name of Monte Carlo have been accepted in mathematics, physics, chemistry, and biology. The key concept is that a random number, when drawn successively many times, may serve to create a sequence of system snapshots.

All this began with an idea of a mathematician from Lwów, Poland (now Lviv in the Ukraine), Stanisław Marcin Ulam.

Stanisław Ulam (1909–1984), first associated with the University of Lwów, then professor at Harvard University, the University of Wisconsin, the University of Colorado, and Los Alamos National Laboratory. In Los Alamos Ulam solved the most important bottleneck in the development of the hydrogen bomb by suggesting that pressure is the most important factor and that sufficient pressure could be achieved by using an *atomic* bomb as a detonator. Using this idea and an idea of Edward Teller about further amplification of the ignition effect by implosion of radiation, both scholars designed the hydrogen bomb. They both owned the US patent for H-bomb production.

According to the *Ulam Quarterly Journal* (http://www.ulam.usm.edu/editor.html), Ulam's contribution to science includes logic, set theory, measure theory, probability theory, computer science, topology, dynamic systems, number theory, algebra, algebraic and arithmetic geometry, mathematical biology, control theory, mathematical economy, and mathematical physics. He developed and coined the name of the Monte Carlo method, and also the cellular automata method (described at the end of this chapter). Stanisław Ulam wrote a very interesting autobiography, "*Adventures of a Mathematician*." During the A-bomb project Ulam lamented that he always worked with symbols, not numbers; he was driven so low as to use them now, and even worse: these are numbers with decimal points!

The picture shows one of the "magic places" of the world of science, the *Szkocka Café*, Akademicka street, Lwów, now a bank at Prospekt Szewczenki 27, where, before the Second World War, young Polish mathematicians, among them the mathematical genius Stefan Banach, made a breakthrough. It is since called the "Polish school of mathematics."

Perhaps an example will best explain the Monte Carlo method. I have chosen the methodology introduced to the protein folding problem by Andrzej Koliński and Jeffrey Skolnick.[57] In a version of this method we use a simplified model of the real protein molecule, a polymer composed of monomeric peptide units –NH–CO–CHR–, as a chain of point-like entities NH–CO–CH from which protrude points representing various side chains R. The polymer chain goes through the vertices of a crystallographic lattice (the side chain points can also occupy only the lattice vertices), which excludes a lot of unstable conformations and enables us to focus on those chemically/physically relevant. The lattice representation speeds up computation by several orders of magnitude.

The reasoning goes as follows. The nonzero temperature of the water the protein is immersed in makes the molecule acquire random conformations all the time. The authors assumed that a given starting conformation is modified by a series of random micromodifications. An allowed micromodification has to be chosen so as to obey three rules: these have to be:

- chemically/physically acceptable;
- always local, i.e., they have to pertain to a small fragment of the protein, because in the future we would like to simulate the kinetics of the protein chain (how a conformational change evolves);
- when repeated, they should be able to transform any conformation into any other conformation of the protein.

This way we are able to modify the molecular conformation, but we want the protein to move, i.e., to have the *dynamics* of the system, i.e., a sequence of molecular conformations, each one derived from the previous one in a physically acceptable way.

To this end we have to be able to write down the energy of any given conformation. This is achieved by assigning an energy award (i.e., energy lowering) if the configuration corresponds to intramolecular energy gain (e.g., trans conformation, the possibility of forming a hydrogen bond or a hydrophobic effect, see Chapter V2-5), and an energy penalty for intramolecular repulsion (e.g., cis conformation, or when two fragments of the molecule are to occupy the same space). It is, in general, better if the energy awards and penalties are accompanied by an ingredient that reflects the complex reality of the Protein Data Bank, the most extended data base. This data base may serve to add what is known as the *statistical interaction potential*. The potential is a correction computed from the frequency of finding two kinds of amino acids close in space (e.g., glutamic acid and lysine; there are 20 natural amino acids) in the Protein Data Bank. If the frequency is larger than just the random one, we deduce an attraction has to

---

[57] J. Skolnick, A. Koliński, *Science*, 250(1990)1121.

occur between them and assign an energy gain that takes into account the value of this higher frequency.[58]

Now we have to let the molecule move. We start from an arbitrarily chosen conformation and calculate its energy E_1. Then, a micromodification, or even a series of micromodifications (this way the calculations go faster), is drawn from the micromodifications list and applied to the molecule. Thus a new conformation is obtained with energy E_2. Now the most important step takes place. We decide to *accept* or to *reject* the new conformation according to the *Metropolis criterion*[59], which gives the probability of the acceptance as

$$P_{1\to 2} = \begin{cases} 1 & \text{if } E_2 \le E_1, \\ a = \exp\left(-\frac{(E_2 - E_1)}{k_B T}\right) & \text{if } E_2 > E_1. \end{cases}$$

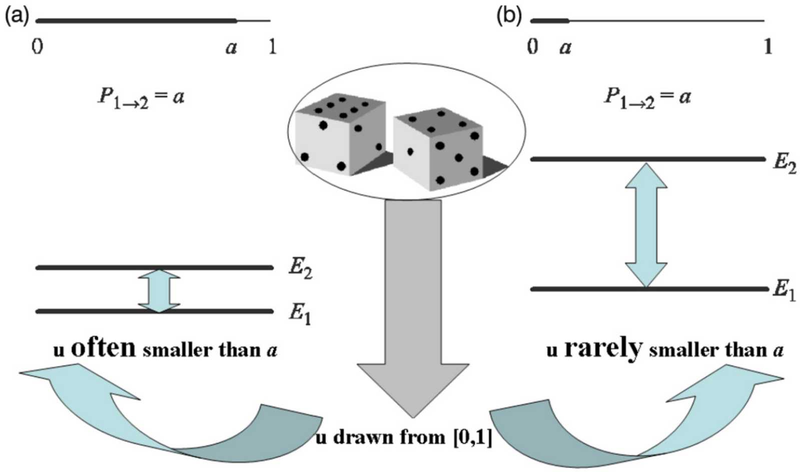


Fig. 7.13. Metropolis algorithm. (a) If E_2 is only a little higher than E_1, then the Metropolis criterion often leads to accepting the new conformation (of energy E_2). (b) However, if the energy difference is large, then the new conformation is accepted only rarely.

Well, we have a *probability* but what we need is a clear decision: *to be or not to be in state 2*. This is where the Monte Carlo spirit comes in (see Fig. 7.13). By using a random number generator we draw a random number[60] u from section [0, 1] and compare it with the number a. If $u \le a$, then we accept the new conformation; otherwise conformation 2 is rejected (and we

---

[58] Such an approach may be certainly considered as a bit too pragmatic. People believe that before coming to a true force field that will give us correct answers (in the distant future), we have a lot of time to explore what kind of corrections the current force fields need to get reasonable answers.

[59] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, *J. Chem. Phys.*, 21(1953)1087.

[60] The current situation is like this: there are deterministic computer programs, which are claimed to generate random numbers. According to John von Neumann, those who use such programs “live in the state of sin.”

forget about it). The whole procedure is repeated over and over again: drawing micromodifications → a new conformation → comparison with the old one by the Metropolis criterion → accepting (the new conformation becomes the current one) or rejecting (the old conformation remains the current one), → drawing micromodifications, etc.

The Metropolis criterion is one of those mathematical tricks a chemist has to know about. Note that the algorithm always accepts conformation 2 if $E_2 \leq E_1$ and therefore will have a tendency to lower the energy of the current conformation. On the other hand, when $E_2 > E_1$ the algorithm may decide to increase the energy by accepting the higher-energy conformation 2. If $\frac{(E_2 - E_1)}{k_B T} > 0$ is small, the algorithm accepts the new conformation very easily (Fig. 7.13a), at a given $E_2 - E_1$ the easier the higher the temperature. On the other hand, an attempt of a very high jump (Fig. 7.13b) in energy may be successful in practice only at very high temperatures. The algorithm prefers higher-energy conformations to the same extent as the Boltzmann distribution. Thus, grinding the mill of the algorithm on and on (sometimes it takes months on the fastest computers of the world) and making statistics of the number of accepted conformations as a function of energy, we arrive at the Boltzmann distribution as it should be in thermodynamic equilibrium.

Thus as the mill grinds we can make a movie. The movie would reveal how the protein molecule behaves at high temperature: the protein acquires practically any new conformation generated by the random micromodifications and it looks as if the molecule is participating in a kind of rodeo. However, we decide the temperature. Thus, let us decide to lower the temperature. Until a certain temperature we will not see any qualitative change in the rodeo, but at a sufficiently low temperature we can recognize that something has happened to the molecule. From time to time some local structures typical of the secondary structures of proteins (the α-helices and the zigzag-type β-strands, the latter like to bind together laterally by hydrogen bonds) emerge and vanish, emerge again, etc.

When the temperature decreases, at a certain critical value, T_{crit}, all of a sudden a stable structure emerges (an analog of the so-called native structure, i.e., the one ensuring the molecule can perform its function in nature).

The structure vibrates a little, especially at the ends of the protein, but further cooling does not introduce anything new. The native structure exhibits a unique secondary structure pattern along the polymeric chain (i.e., definite sections of the α and β structures) which packs together into a unique *tertiary structure*. In this way a highly probable scenario for the coil-globular *phase transition* was demonstrated for the first time by Koliński and Skolnick.

One of the most successful variations of the basic Monte Carlo algorithm described above is what is known as Monte Carlo with Replica Exchange (MCRE). The idea is to speed up the exploration of the conformational space by making parallel computations for the same system ("replica"), but each computation differs by the assumed and fixed temperature, from very low to very high. Then, during the simulation one stochastically exchanges the replicas in a way that is analogous to the Metropolis criterion (instead of E_1 and E_2 one considers $\frac{E_1}{k_B T_1}$ and $\frac{E_2}{k_B T_2}$). This simple idea makes it easier to overcome even large energy barriers.

Predicting the three-dimensional structure of globular proteins from the sequence of amino acids[61] and using a sophisticated algorithm[62] based on all relevant physical interactions and adjusted using the accumulated structural knowledge of proteins (nearly 100 000 structures in the Protein Data Bank, no bias towards the target) is feasible nowadays with a remarkable accuracy. An example is shown in Fig. 7.18 on p. 451.

Example 4 (Conformational autocatalysis as a model of prion disease propagation). In biology there is a concept of "contagious misfolded proteins" (prions), which are supposed to increase their quantity spontaneously at the expense of the native fold of the same protein. This phenomenon is suspected to be the cause of dangerous diseases such as Creutzfeldt–Jakob's in humans or the "mad cow" disease in cattle.

The elaborated force field and the above described MCRE technique proved to predict the three-dimensional structure of many globular proteins to a very good accuracy (e.g., of the order of 2 Å of root mean square [r.m.s.] deviation). Having in hand such a tool one may consider questions that are related to prion disease propagation.

Imagine a protein with an amino acid sequence designed in such a way that the protein folds to a certain conformation (A) being its lowest-energy conformation ("native"). The sequence has however also another state (conformation B), which widely differs structurally from A and is metastable. Thus, there is a frustration introduced on purpose in the molecule.[63] The protein will usually fold to A, but if some external factors are present it may prefer to fold to B. These external factors might be simply the intermolecular interaction of several such molecules. For example, if the lowest energy of the dimer corresponded to the BB conformation, a molecule in conformation A, when in contact with another molecule in conformation B, might change from AB to BB. This would mean elimination of the most stable conformation of a protein molecule just by contact with a contagious, metastable ("misfolded") form, which resembles spreading of a kind of "conformational disease."

---

[61] This problem is sometimes called the "second genetic code" in the literature. This name reflects the final task of obtaining information about protein function from the "first genetic code" (i.e., DNA), which encodes the sequence of amino acids.

[62] J. Skolnick, Y. Zhang, A.K. Arakaki, A. Koliński, M. Boniecki, A. Szilagyi, D. Kihara, *Proteins*, 53(2003), Suppl. 6469–479.

[63] E. Małolepsza, M. Boniecki, A. Koliński, L. Piela, *Proc. Natl. Acad. Sciences (USA)*, 102(2005)7835.

If such a scenario were possible, then is the theory strong enough to predict the necessary amino acid sequence that would behave like that? Well, let us consider an oligopeptide composed of 32 amino acids only, small enough for making many series of computations feasible. There is already some computational experience that serves as a guideline in planning the amino acid sequence.

First of all, to make the effect seen clearly we plan to have A and B structurally distinct as much as possible. Let one of them be of α-helical (A), the other (B) of β-sheet character, both quite compact ones ("globular"). To induce this compactness let us locate two glycines in the middle of the sequence, because we want the chain to bend easily there, changing its direction in space (a "hairpin-like" structure). Glycine is unique among the 20 amino acids given by Nature; it is known for its flexibility, since its side chain is a very small object – the hydrogen atom (no sterical hindrance). Two other glycines have been put at positions 8 and 25, exactly where we plan to have some increased flexibility necessary for the planned β-sheet (we plan four β-strands bound together by the hydrogen bonds). We hope these latter glycines will not break the two α-helices to be formed and packed in the hairpin-like structure. Moreover, we would like the native structure (A) to have an additional stability. To this end we will introduce the interactions within the four pairs of glutamic acid and lysine (they interact through the electrostatic attraction $\text{–NH}_3^+ \cdots {}^-\text{OOC–}$) in the positions that ensure they are close in space, but only in a pattern typical for an α-helix. However, to increase also a chance of forming β structures (to make the B structure stable), the sequence has been enriched by a pattern of hydrophobic valines and isoleucines (they are known for forming what is called the *valine-leucine zipper*) in the amino acid sequence, their positions ensuring a strong interaction within a β structure. Although the water structure (necessary for the hydrophobic effect) is not taken into account explicitly in the force field used, it is present there implicitly through the force field, e.g., the statistical amino acid interaction potential from the Protein Data Bank (which do take it into account).

To complete such a design, the amino acid sequence has been tuned, increasing the propensity towards α-helices or towards β-sheets. The calculations have been repeated independently for many sequences, each time checking the behavior of single molecules as well as dimers and trimers. Among the final 14 sequences only one, i.e.,

GVEIAVKGAEVAAKVGGVEIAVKAGEVAAKVG

(G = glycine, V = valine, E = glutamic acid, I = isoleucine, A = alanine, K = lysine), exhibited the desired conformational autocatalysis.

It turned out in the MCRE procedure that the oligopeptide molecule always attains the global minimum conformation in the form of a two-helix bundle (Fig. 7.14a,b) if the temperature is

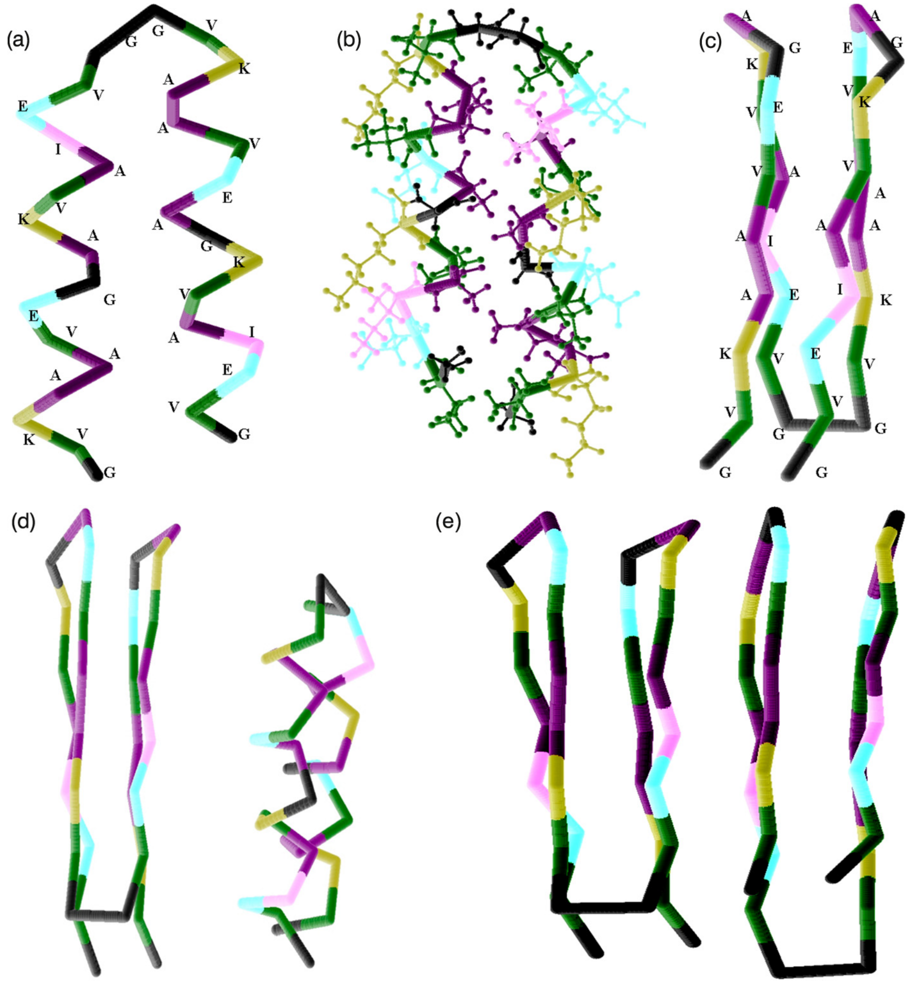


Fig. 7.14. Conformational autocatalysis. (a) A two α–helix bundle conformation of a protein with a particular amino acid sequence: GVEIAVKGAEVAAKVGGVEIAVKAGEVAAKVG (G = glycine, V = valine, E = glutamic acid, I = isoleucine, A = alanine, K = lysine). This three-dimensional structure is shown in a simplified way by representing a single amino acid as a point connected to its neighbors. (b) The all heavy atom representation of such structures can be recovered, for example for the two α-helix bundle. The structure corresponds to the global minimum of the protein molecule. Note the stabilizing interaction of the hydrophobic amino acids (AA, VA, VV) between the helices, as well as the electrostatic KE stabilization within them. (c) A β-barrel conformation (a metastable one) of the same protein. The stabilization of this structure also comes from similar interactions, this time acting between the β-strands. (d) The starting configuration for the MCRE procedure: the frozen β-barrel (left) interacts with the two-α-helix bundle (right), which is subject to MCRE dynamics. (e) The final result after a long MCRE run: the α-helix bundle has unfolded and then refolded to a β-barrel conformation, thus producing two interacting β-barrels (autocatalysis).

higher than a certain minimal temperature. Below this minimal temperature not only the two-helix bundle, but also a (very different from the latter one) four-member β-barrel (Fig. 7.14c) is stable. If two protein molecules interact, one of them frozen (for whatever reasons, chemical or physical) in its metastable β-barrel conformation, while the second molecule is free to move, the second molecule practically always folds to the β-barrel, interacting very strongly with the frozen β-barrel (Fig. 7.14d). This happens even when the second molecule starts from its native, i.e., α-helical form: the α-helical form unfolds and then folds to the β-barrel (Fig. 7.14e). It has been also demonstrated that a third protein molecule, when in presence of the two frozen β-barrels, folds to the β-barrel form, which fits very well to the two, forming a stack of three β-barrel molecules. Formation of such stacks is notorious for prion diseases.

This model may be seen as a prototype of prion disease propagation, in which a metastable "incorrect conformation" spontaneously spreads out in the system.

7.10 Car–Parrinello dynamics

Despite the fact that the present textbook assumes that the reader has completed a basic quantum chemistry course, we do not profit from this too extensively. Car–Parrinello dynamics is an exception. It positively belongs to the present chapter, while borrowing heavily from the results of Chapter 8. If the reader feels uncomfortable with this, the section may just be omitted.

We have already listed some problems associated with the otherwise nice and powerful molecular dynamics. We have also mentioned that the force field parameters (e.g., the net atomic charges) do not vary when the conformation changes or when two molecules approach, whereas everything has to change. Car and Parrinello[64] thought of a remedy in order to make the parameters change "in flight."

Let us assume the one-electron approximation.[65] Then the total electronic energy $E_0^0(\mathbf{R})$ is (in the Born–Oppenheimer approximation) not only a function of the positions of the nuclei, but also a functional of the spin orbitals $\{\psi_i\}$: $V = V(\mathbf{R}, \{\psi_i\}) \equiv E_0^0(\mathbf{R})$.

The function $V = V(\mathbf{R}, \{\psi_i\})$ will be minimized with respect to the positions $\mathbf{R}$ of the nuclei and the spin orbitals $\{\psi_i\}$ depending on the electronic coordinates.

---

[64] R. Car, M. Parrinello, *Phys. Rev. Letters*, 55(1985)2471.

[65] The approximation will be described in Chapter 8 and is based on assuming the wave function in the form of a single Slater determinant built of orthonormal spin orbitals. Car and Parrinello gave their procedure for the density functional theory (DFT). As will be seen in Chapter V2-3, there also a single Slater determinant plays an important role.

If we are going to change the spin orbitals, we have to take care of their orthonormality at all stages of the change.[66] For this reason Lagrange multipliers appear in the equations of motion (Appendix O). We obtain the following set of Newton equations for the motion of N nuclei:

$$M_I \ddot{X}_I = -\frac{\partial V}{\partial X_I} \quad \text{for} \ \ I = 1, ..., 3N,$$

and we obtain the following equation of motion for each spin orbital (each corresponding to the evolution of one electron probability density in time):

$$\mu \ddot{\psi}_i = -\hat{F}\psi_i + \sum_{j=1} \Lambda_{ij}\psi_j, \tag{7.19}$$

where $\mu > 0$ is a *fictitious parameter*[67] for the electron, $\hat{F}$ is a Fock operator (see Chapter 8, p. 478), and Λ_{ij} are the Lagrange multipliers to ensure the orthonormality of the spin orbitals ψ_j.

Both equations are quite natural. The first (Newton equation) says that a nucleus has to move in the direction of the force acting on it ($-\frac{\partial V}{\partial X_I}$) and the larger the force and the smaller the mass, the larger the acceleration achieved. Good! The left-hand side of the second equation and the first term on the right-hand side say the following: let the spin orbital ψ_i change in such a way that the orbital energy has a tendency to go down (in the sense of the mean value). How on earth does this follow from the equation? From a basic course in quantum chemistry (this will be repeated in Chapter 8) we know that the orbital energy may be computed as the mean value of the operator $\hat{F}$ with the spin orbital ψ_i, i.e., $\left\langle \psi_i | \hat{F}\psi_i \right\rangle$. To focus our attention, let us assume that $\delta\psi_i$ is localized in a small region of space (see Fig. 7.15).

From Fig. 7.15, it is seen that it would be desirable to have the acceleration $\ddot{\psi}_i$ with the same sign as $-\hat{F}\psi_i$. This is equivalent to increasing the changes that lower the corresponding orbital energy and to suppressing the changes that make it higher. The ψ_i spin orbitals obtained in the numerical integration have to be corrected to satisfy orthonormality conditions, as is ensured by the second term in (7.19).

The prize from the elegance of the Car–Parrinello method is the computation time, which allows one to treat systems currently up to a few hundreds of atoms (while molecular dynamics may even deal with a million of atoms). The integration interval has to be decreased by a factor of 10 (i.e., 0.1 fs instead of 1 fs), which allows us to reach simulation times of the order of 10–100 picoseconds instead of (in classical molecular dynamics) nanoseconds.

---

[66] Because the formulae they satisfy are valid under this condition.

[67] We may call it "mass." In practical applications μ is large, usually taken as a few hundreds of the electron mass, because ... this ensures the agreement of theory and experiment.

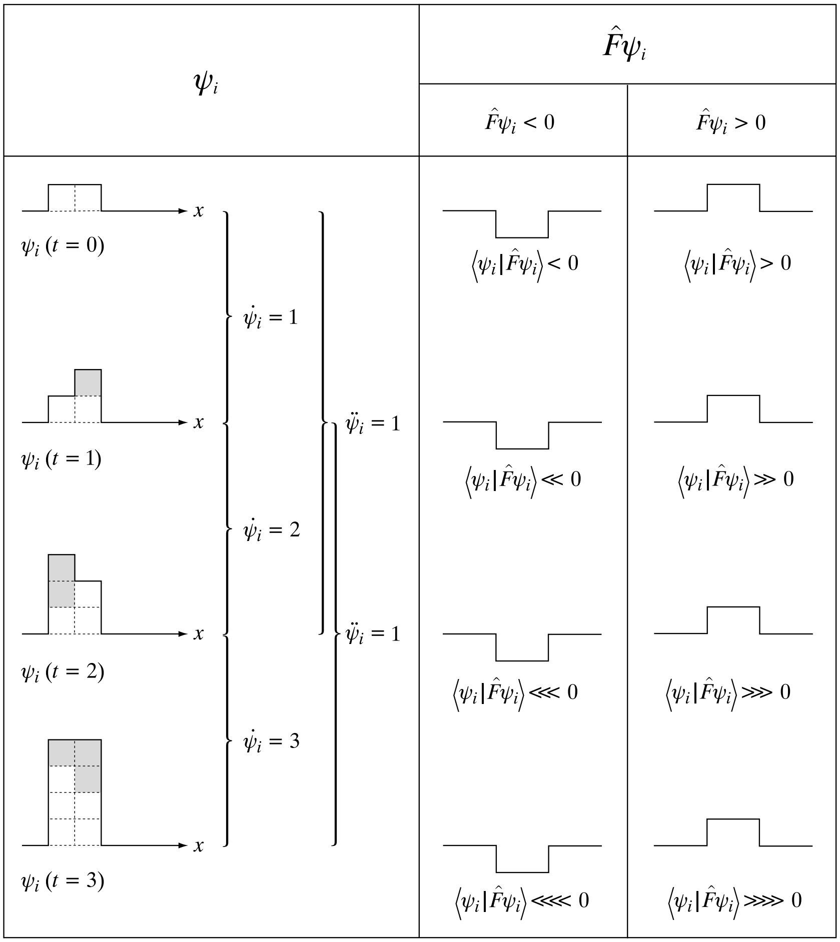


Fig. 7.15. A scheme showing why the acceleration $\ddot{\psi}_i$ of the spin orbital ψ_i has to be of the same sign as that of $-\hat{F}\psi_i$. Time (arbitrary units) goes from up ($t = 0$) downwards ($t = 3$) where the time step is $\Delta t = 1$. On the left-hand side the changes (localized in one-dimensional space, x is the electron coordinate) of ψ_i are shown in a schematic way (in single small square units). It is seen that the velocity of the change is not constant and the corresponding acceleration is equal to 1. Now let us imagine for simplicity that function $\hat{F}\psi_i$ has its nonzero values precisely where $\psi_i \neq 0$ and let us consider two cases (last two columns): (a) $\hat{F}\psi_i < 0$ and (b) $\hat{F}\psi_i > 0$. In such a situation we may easily foresee the sign of the mean value of the energy $\left\langle\psi_i|\hat{F}\psi_i\right\rangle$ of an electron occupying spin orbital ψ_i. In situation (a) the conclusion for changes of ψ_i is: *keep that way* or, in other words, even *increase* the acceleration $\ddot{\psi}_i$, making it proportional to $-\hat{F}\psi_i$. In (b) the corresponding conclusion is: *suppress* these changes or, in other words, *decrease the acceleration*, e.g., making it negative as $-\hat{F}\psi_i$. Thus, in *both* cases we have $\mu\ddot{\psi}_i = -\hat{F}\psi_i$, which agrees with Eq. (7.19). In both cases there is a trend to lower orbital energy $\varepsilon_i = \left\langle\psi_i|\hat{F}\psi_i\right\rangle$.

7.11 Cellular automata

Another powerful tool is the cellular automata method, invented by John (in his Hungarian days Janos) von Neumann and Stanisław Marcin Ulam (under the name of "cellular spaces"). Cellular automata are mathematical models in which space and time both have a granular structure (in molecular dynamics only time has such a structure). A cellular automaton consists of a periodic lattice of cells (nodes in space). In order to describe the system locally, we assume that every cell has its "state" representing a vector of N components. Each component is a Boolean variable, i.e., a variable having a logical value (e.g., "0" for "*false*" and "1" for "*true*").

A cellular automaton evolves using some *propagation and collision (or actualization) rules* that are always of a *local character*. The local character means that (at a certain time step t and a certain cell) the variables change their values depending only on what happened at *the cell* and at its *neighbors* at time step $t - 1$. The propagation rules dictate what would happen next (for each cell) with variables on the cell and on the nearest neighbor cells for each cell independently. But this may provoke a collision of the rules, because a Boolean variable on a cell may be forced to change by the propagation rules related to two or more cells. We need a unique decision and this comes from the collision, or actualization, rules.

Example 5 (Gas lattice model). One of the simplest examples pertains to a lattice model of a gas. Let the lattice be regular two-dimensional (Fig. 7.16).

Propagation rules

There are a certain number of point-like particles of equal mass which may occupy the nodes (cells) only and have only unit velocities pointing either in North–South or East–West directions, thus reaching the next row or column after a unit of time. We assign each cell a state which is a four-dimensional vector of Boolean variables. The first component tells us whether there is a particle moving North on the node (Boolean variables take 0 or 1), the second moving East, the third South, and the fourth West. There should be no more than one particle going in one direction at any node; therefore a cell may correspond to 0, 1, 2, 3, or 4 particles. Any particle is shifted by one unit in the direction of the velocity vector.

Collision rules

If two particles are going to occupy the same state component at the same cell, the two particles are annihilated and a new pair of particles is created with drawn positions and velocities. Any two particles which meet at a node with opposite velocities acquire the velocities that are opposite to each other and perpendicular to the old ones (the "lateral outgoing," see Fig. 7.16).

This primitive model has nevertheless an interesting property. It turns out that such a system attains an equilibrium state. No wonder that this approach with more complex lattices and rules became popular. Using the cellular automata we may study an extremely wide range of

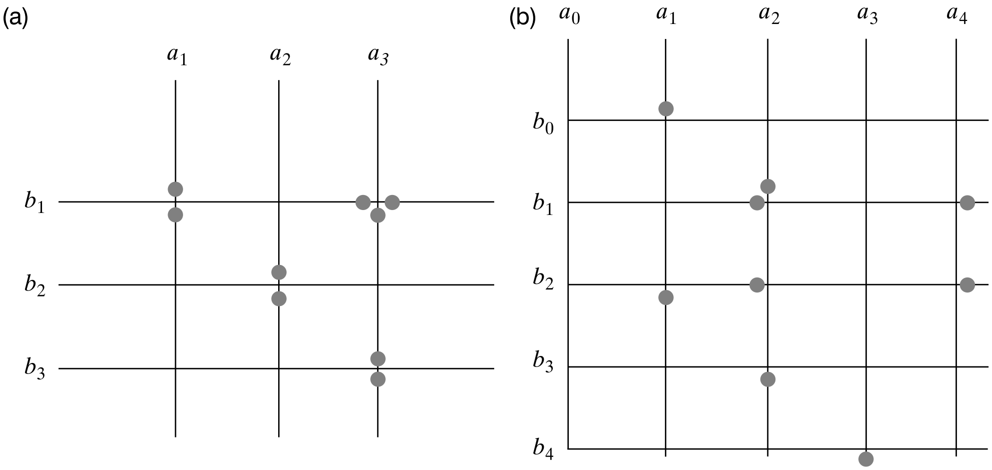


Fig. 7.16. Operation of a cellular automaton – a model of gas. The particles occupy the lattice nodes (cells). Their displacement from the node symbolizes which direction they are heading on with the velocity equal to 1 length unit per time step. On the left scheme (a) the initial situation is shown. On the right scheme the result of the one-step propagation and one-step collision is shown. Collision only took place in one case (at a_3b_2) and the collision rule has been applied (of the lateral outgoing). The game would become more dramatic if the number of particles were larger and if walls as well as appropriate propagation rules (with walls) were introduced.

phenomena, such as turbulent flow of air along a wing surface, electrochemical reactions, etc. It is a simple and powerful tool of general importance.

For physically relevant states, the propagation and collision rules for the behavior of such a set of cells, as time goes on, may mirror what would happen with a physical system. This is why cellular automata are appealing. Another advantage is that due to the locality mentioned above, the relevant computer programs may be effectively parallelized, which usually significantly speeds up computations. The most interesting cellular automata are those for which the rules are of a nonlinear character (cf. Chapter V2-7). There are generally four qualitatively different cellular automata evolutions: static, repetitive, chaotic, and intelligent. Well, let us taste the latter one.

Example 6 (Langton's ant). A two-dimensional gas model given above may be useful, but is essentially quite boring when compared to a two-dimensional cellular automaton known as Langton's ant[68]: an infinite regular lattice of unit squares (cells) having two colors possible, say, black and white. We start from all cells white. At each step a single cell is a special one: this is where the ant resides. The propagation rules are as follows. The ant moves by a unit

---

[68] Ch.G. Langton, *Physica D: Nonlinear Phenomena*, 22(1986)120.

length in four possible directions (N, E, S, W). At each time step, if the ant resides in a white (black) cell, it turns left (right) by 90° and enters the neighboring cell, while changing the color of the old cell.

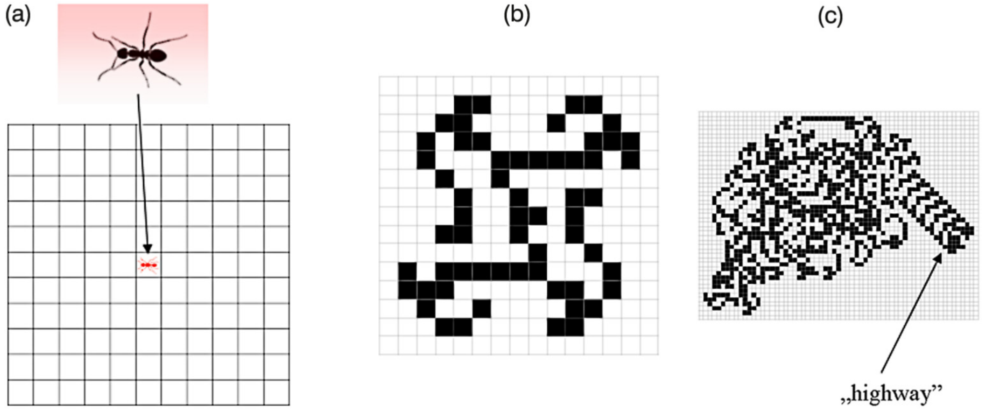


Fig. 7.17. An example of cellular automaton: the Langton ant. (a) The starting point. (b) After 386 steps. (c) After 10 647 steps – we see already the highway under construction! E.W. Weisstein, “Langton’s Ant.” From MathWorld – A Wolfram Web Resource.

This extremely simple rule is able to lead from a monochromatic plane (Fig. 7.17a) to an extraordinary pattern. After several dozens of steps the pattern looks rather a bit chaotic (Fig. 7.17b) and it is still looking like that until some 10 000 steps. It seems that nothing special can be expected. Then quite abruptly the ant begins to build and replicate a very peculiar pattern with an artifact form called “highway” (right-hand side of Fig. 7.17c), its length increasing every 104 steps!

This example shows that complexity may at least sometimes emerge from very simple rules, which is encouraging for natural sciences trying to understand the enormous complexity we observe in the Universe and in ourselves.

Summary

- *Detailed* information about a molecule (in our case: a three-atom complex $C \cdots AB$) may be obtained making use of the potential energy hypersurface for the nuclear motion computed as the ground-state electronic energy (however, even in this simplified case, the AB distance has been frozen). After constructing the basis functions appropriate for the five variables and applying the Ritz method of solving the Schrödinger equation for the motion of nuclei we obtain the rovibrational levels and corresponding wave functions for the system. This allows us to compute the IR and microwave spectra.
- However, in chemical practice we are forced to construct an approximation to the potential energy hypersurface for the motion of the nuclei by designing what is known as *force field*, or a simple

expression for the electronic energy as a function of the position of the nuclei. Most often in proposed force fields we assume harmonicity of the chemical bonds and bond angles ("springs"). The hypersurface obtained often has a complex shape with many local minima.

- *Molecular mechanics* (should have the adjective "local") represents
  - choice of the starting configuration of the nuclei (a point in the configuration space), and
  - sliding downhill from the point (configuration) to the "nearest" local minimum, which corresponds to a stable conformation with respect to small displacements in the configurational space.
- *Global molecular mechanics* means
  - choice of the starting configuration of the nuclei,
  - finding the global (the lowest-energy) minimum, i.e., the most stable configuration of the nuclei.

While the local molecular mechanics represents a standard procedure, the global one is still *in statu nascendi*.

- Any of the potential energy minima can be approximated by a paraboloid. Then, for N nuclei, we obtain $(3N-6)$ *normal modes* (i.e., harmonic and having the same phase) of the molecular vibrations. This represents important information about the molecule, because it is sufficient to calculate the IR and Raman spectra frequencies (cf. p. 605). Each of the normal modes makes all the atoms move, but some atoms may move more than others. It often happens that a certain mode is dominated by the vibration of a particular bond or functional group and therefore the corresponding frequency is *characteristic* for this bond or functional group, which may be very useful in chemical analysis.
- Molecular mechanics does not involve atomic kinetic energy, *molecular dynamics* does. Molecular dynamics represents a method of solving the Newton equations of motion[69] for all the atoms of the system. The forces acting on each atom at a given configuration of the nuclei are computed (from the potential energy V assumed to be known[70]) as $\mathbf{F}_j = -\nabla_j V$ for atoms $j = 1, 2, \ldots, N$. The forces known, we calculate the acceleration vector, and from that the velocities and the new positions of the atoms after an assumed time increment. The system starts to evolve as time goes on. Important ingredients of the molecular dynamics procedure are:
  - choice of starting conformation,
  - choice of starting velocities,
  - thermalization at a given temperature (with velocity adjustments to fulfill the appropriate Maxwell–Boltzmann distribution),
  - harvesting the system trajectory,
  - conclusions derived from the trajectory.
- In molecular dynamics (also in the other techniques listed below) there is the possibility of applying a sequence (protocol) of cooling and heating intervals in order to achieve a low-energy configuration of the nuclei (*simulated annealing*). The method is very useful and straightforward to apply.
- Besides molecular dynamics, there are other useful techniques describing the motion of the system:

---

[69] We say sometimes: integration.

[70] Usually it is a force field.

- Langevin dynamics, which allows taking into account the surrounding solvent ("at virtually no cost"),
- Monte Carlo dynamics – a powerful technique basing on drawing and then accepting/rejecting random configurations by using the *Metropolis criterion*; the criterion says that if the energy of the new configuration is lower, the configuration is accepted, and if it is higher, it is accepted with a certain probability;
- Car–Parrinello dynamics allows for the electron structure to be changed "in flight" when the nuclei move;
- cellular automata – a technique of general importance, which divides the total system in cells. Each cell is characterized by its state being a vector with its components being Boolean variables. There are propagation rules that change the state, as time goes on, and collision rules, which solve conflicts of the propagation rules. Both types of rules have a local character. Cellular automata evolution may have some features in common with thermodynamic equilibria.

Main concepts, new terms

angular momenta addition (p. 404)
annealing protocol (p. 434)
autocorrelation (p. 430)
Boolean variables (p. 446)
Car–Parrinello algorithm (p. 443)
cellular automaton (p. 446)
characteristic frequency (p. 416)
entropy (p. 413)
force field (p. 406)
free energy (p. 415)
global minimum (p. 413)
global optimization (p. 413)
Jacobi coordinate system (p. 402)
kinetic minimum (p. 413)
Langevin dynamics (p. 435)
Lennard-Jones potential (p. 409)
Metropolis algorithm (p. 438)
molecular dynamics (p. 427)
molecular mechanics (p. 411)
Monte Carlo dynamics (p. 435)
normal modes (p. 416)
rovibrational spectrum (p. 401)
simulated annealing (p. 434)
spatial correlation (p. 430)
thermalization (p. 430)
thermodynamic minimum (p. 413)
time correlation (p. 430)
torsional potential (p. 409)

From the research front

The number of atoms taken into account in molecular dynamics may nowadays reach a million. The real problem is not the size of the system, but rather its complexity and the wealth of possible structures, with too large a number to be investigated. Some problems may be simplified by considering a quantum mechanical part in the details and a classical part described by Newton equations. Another important problem is to predict the three-dimensional structure of proteins, starting from the available amino acid sequence. Every two years from 1994 a Critical Assessment of techniques for protein Structure Prediction (CASP) has been organized in California. CASP is a kind of scientific competition, in which theoretical laboratories (knowing only the amino acid sequence) make blind predictions about three-dimensional protein structures about to be determined in experimental laboratories. Most of the theoretical methods are based on the similarity of the sequence to a sequence from the Protein Data Bank of three-dimensional structures; only some of the methods are related to chemical physics.[71] Fig. 7.18 shows two examples of the accuracy achieved for quite large molecules (about 3000 heavy atoms).

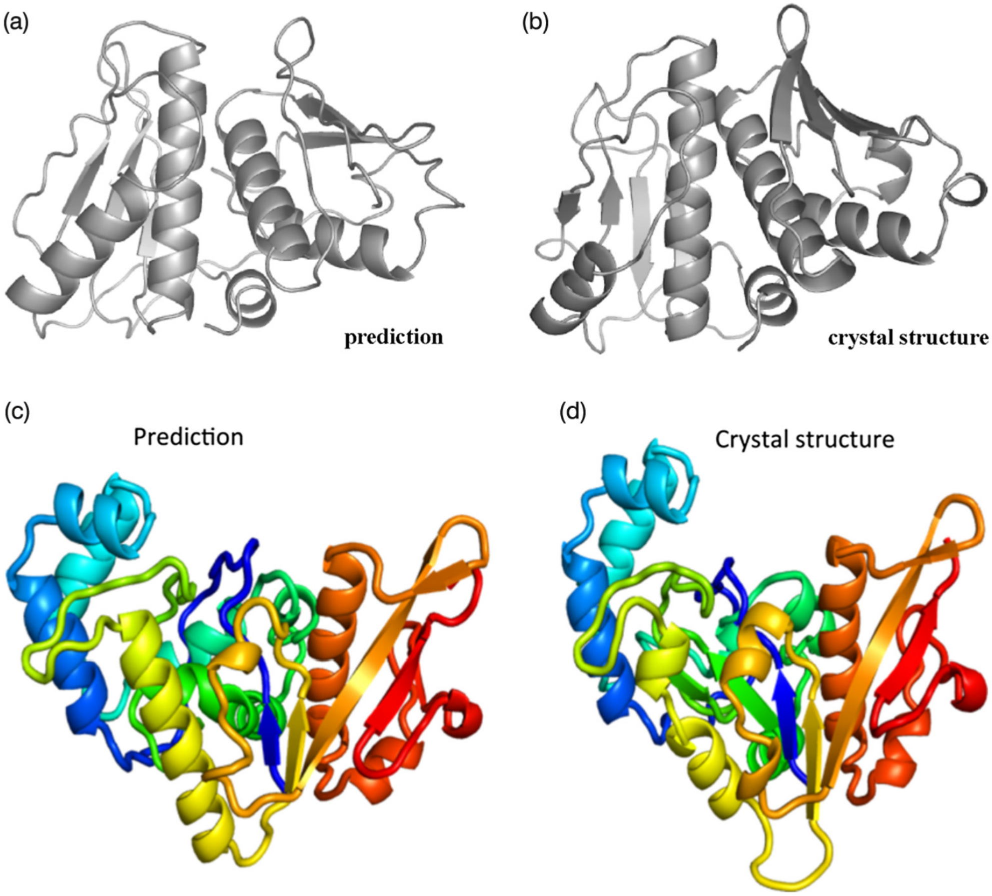


Fig. 7.18. A picture is worth a thousand words. Reproduced by courtesy of Professors Andrzej Koliński and Krzysztof Bujnicki (a,b) and Professor David Baker (c,d).

The first example (Fig. 7.18a,b) corresponds to the putative nitroreductase (206 amino acids), one of the 1877 proteins of the bacterium *Thermotoga maritima*, which lives in geothermal marine sediments. The

---

[71] More details, e.g., in http://predictioncenter.llnl.gov/casp6/Casp6.html.

energy expression which was used in theoretical calculations resulting in Fig. 7.18a takes into account the physical interactions (such as hydrogen bonds, hydrophobic interactions, etc.) as well as an empirical potential deduced from representative proteins' experimental structures deposited in the Brookhaven Protein Data Bank (no bias towards the target protein). The X-ray-determined atomic positions for the corresponding crystal lead to Fig. 7.18b.

The second protein (T0806 in the 2014 CASP11 competition) represents a chain of 223 amino acids, and its amino acid sequence had only a 5% similarity to any known protein structure. The predicted structure[72] is given in Fig. 7.18c. In this case the prediction methodology had both automatic and human-assisted protocols, some of them based on the particularly successful "coevolution-derived contacts," i.e., the amino acid–amino acid contacts through space that are preferred in protein families.

Both theory and experiment give the target molecules as containing quite a few α-helices and some β-pleated sheets (wide arrows). These (typical for proteins) secondary structure elements interact in a complex way and form the unique (native) tertiary structure, which is able to perform its biological function.

Ad futurum

The maximum size of the systems investigated by the molecular mechanics and molecular dynamics methods will increase systematically to several million atoms in the near future. A critical problem will be the choice of the system to be studied as well as the question to be asked. Very probably nonequilibrium systems will become more and more important, e.g., concerning impact physics, properties of materials subject to various stresses, explosions, self-organization, and first of all chemical reactions. At the same time the importance of molecular dynamics simulations of microtools of dimensions of tens of thousands angstrom will very probably increase.

Treating nanostructures at a mesoscale will become more and more popular. The CASP successes described above show that even complex problems, with complexity beyond our imagination, may become solvable. The protein three-dimensional structure is necessary for producing new generation drugs and with this multibillion-dollar market any method of prediction is good (if more or less reliable). Note that it seems that to solve the problem we do not need electrons, quantum mechanics, even nuclei. What turned out to be decisive was treating each amino acid as *a single bead* (coarse-grained models), create some protocols to profit fully from the known protein structures from the Protein Data Bank and an extensive use of the Monte Carlo method. This is a hint about what to do with other problems of enormous structural complexity, but with a limited number of kinds of structural elements. Fig. 7.19 shows a general philosophy for how to attack problems which exhibit a hierarchy of energy basins.

There is a chance we will be efficient (in producing things). Is efficiency at the heart of science? Maybe understanding is closer to what the human endeavor we call science is? Well, but what does understanding mean, when one has a dozen or a hundred of reasons to explain why a phenomenon occurs? This is the kind of questions we will have to do with, when using our faster and faster computers.

---

[72] D. Baker et al., *Proteins*, 84(2016)67.

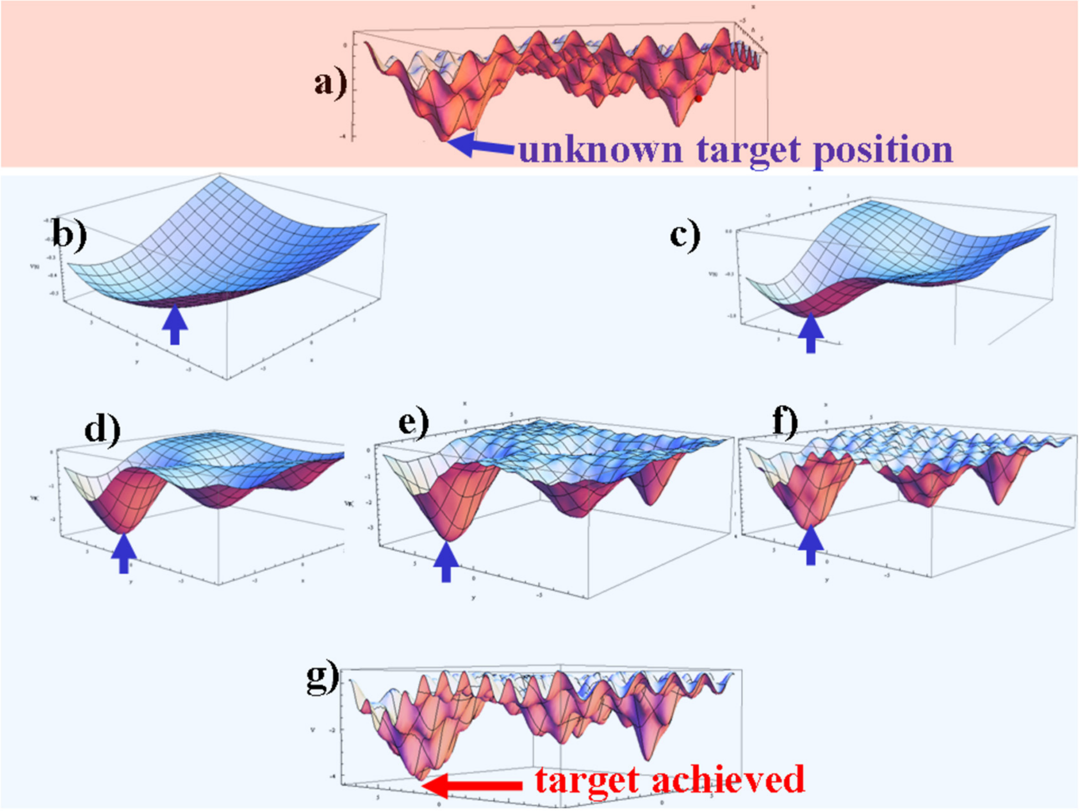


Fig. 7.19. A general philosophy of combatting the multiple minima problem in a sequence of approximations. (a) The conformational energy as a function of configuration of the nuclei (PES for the ground electronic state). Even if we knew this function, it would be hopeless to find the corresponding lowest-energy conformation (the global minimum), because of an astronomic number of misleading and irrelevant minima. (b) A very crude coarse-grained PES allows one to find a low energy minimum (hopefully related to the position of the global one). (c)–(f) Starting from the position found, one systematically improves the PES (including describing more and more accurately some essential physical effects, profiting from analogies found in data banks, etc.), finding better and better approximations to the global minimum. Each time the PES becomes more complicated, but we are already closer to the target. (g) The target is hopefully achieved. Even a suboptimal solution is of great value.

Additional literature

A.R. Leach, "Molecular Modelling. Principles and Applications," Longman, 1996.

A "Bible" of theoretical simulations.

M.P. Allen, D.J. Tildesley, "Computer Simulations of Liquids," Oxford Science Publications, Clarendon Press, Oxford, 1987.

A book with a more theoretical flavor.

Questions

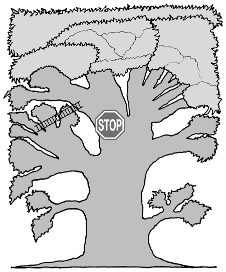
1. The energy $V(\mathbf{R})$ as a potential energy for the motion of the nuclei in the adiabatic approximation:
   a. allows one to identify the most stable configuration of the nuclei as such a point of the configurational space for which $\nabla V(\mathbf{R}) = \mathbf{0}$.
   b. differs for two isotopomers.
   c. when in the Born–Oppenheimer approximation gives the same $V(\mathbf{R})$ for two isotopomers.
   d. the gradient $\nabla V(\mathbf{R})$ calculated for a stable configuration $\mathbf{R}$ of the nuclei is equal to $\mathbf{0}$.
2. The symbols in the Hamiltonian $\hat{H} = -\frac{\hbar^2}{2\mu R^2}\frac{d}{dR}R^2\frac{d}{dR} + \frac{\hat{l}^2}{2\mu R^2} + \frac{\hat{j}^2}{2\mu_{AB} r_{eq}^2} + V$ for a triatomic system C$\cdots$AB mean:
   a. V stands for the ground-state electronic energy as a function of the configuration of the nuclei.
   b. $R =$ the C–A distance.
   c. $r_{eq} =$ the A–B distance.
   d. $\hat{j}^2$ means the operator of the square of the angular momentum of C with respect to the center of mass of AB.
3. A force field:
   a. represents an approximate ground-state electronic energy as a function of the configuration of the nuclei.
   b. represents an approximation for the total energy of the molecular system.
   c. allows to compute the approximate forces acting on each nucleus at a given nuclear configuration.
   d. means an electric field created by the molecule.
4. The normal mode frequencies for a molecule:
   a. pertain to a particular minimum of the potential energy (they will be different for another minimum).
   b. are calculated assuming that the vibrational amplitudes are small.
   c. take into account a small anharmonicity.
   d. are different for different isotopomers.
5. The equation for the normal modes $(\mathbf{A}-\omega_k^2\mathbf{1})\mathbf{L}_k=\mathbf{0}$ ($\mathbf{M}$ is a diagonal matrix of atom masses):
   a. if $\mathbf{A}$ is computed for a point for which the gradient of the potential energy is zero, it may happen that ω_k is imaginary.
   b. if $\mathbf{A} = \mathbf{M}^{-\frac{1}{2}}\mathbf{V}''\mathbf{M}^{-\frac{1}{2}}$, where $\mathbf{V}''$ stands for the matrix of second derivatives computed for a minimum of the potential energy, $\omega_k \geq 0$ for all k.
   c. the vectors $\mathbf{L}_k$ represent the columns of $\mathbf{A}$.
   d. $Tr(\mathbf{V}'') < 0$, computed at a point where $\nabla V = \mathbf{0}$, means that the point is a saddle point.
6. One of the following sets of the wave numbers (cm^{-1}) in an IR spectrum may correspond to vibrations of the following bonds: C–H, C–C, C=C:
   a. 2900, 1650, 800.
   b. 800, 2900, 1650.

   c. 1650, 800, 2900.
   d. 2900, 800, 1650.

7. In the simulated annealing method in molecular mechanics:
   a. one adapts the velocity of particles to the Maxwell distribution for a given temperature (which varies during the simulation).
   b. the goal is to find the most stable structure.
   c. one carries out computation with slowly and monotonically decreasing temperature.
   d. one alternatively increases and decreases the temperature.
8. In the Metropolis algorithm used in the Monte Carlo method (for temperature T) a new configuration of the nuclei is accepted:
   a. only if its energy is lower than the energy of the current configuration.
   b. always.
   c. always, if its energy is lower than the energy of the current one.
   d. sometimes, even if its energy is higher than the energy of the current configuration.
9. In the Langevin dynamics the atoms of the solvent molecules:
   a. represent a source of stochastic forces acting on the molecule under consideration.
   b. are responsible for a friction felt by moving atoms of the molecule under consideration.
   c. are treated differently than the atoms of the molecule under consideration.
   d. are treated in the same way as the atoms of the molecule under consideration.
10. In the Car–Parrinello dynamics:
    a. the nuclei move according to the Newton dynamics, while for the electrons one solves the Schrödinger equation for each configuration of the nuclei.
    b. the electronic charge distribution adjusts to the current positions of the nuclei, while the position of the nuclei changes due to the electronic charge distribution.
    c. each electron is considered as having the mass of 1 a.u.
    d. when the nuclei move, the electronic charge distribution changes in such a way as to decrease the electronic energy at any position of the nuclei.

Answers

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



CHAPTER 8

Orbital Model of Electronic Motion in Atoms and Molecules

> *Everything should be made as simple as possible, but not simpler.*
> ***Albert Einstein***

Where are we?

We are in the upper part of the main trunk of the TREE.

An example

What is the electronic structure of atoms? How do atoms interact in a molecule? Two *neutral* moieties (say, hydrogen atoms) attract each other with a large force *of a similar order of magnitude* as the Coulombic forces between two ions. This is quite surprising. What pulls these neutral objects to one another? These questions are at the foundations of chemistry.

What is it all about?

Ideas of Quantum Chemistry
https://doi.org/10.1016/B978-0-44-464246-2.00016-1

RESULTS OF THE HARTREE–FOCK METHOD

- Localization in practice – computational technique
- The chemical bonds of σ, π, δ symmetry
- Electron pair dimensions and the foundations of chemistry
- Hybridization or mixing one-center AOs

A minimal model of a molecule (▲) **p. 571**

- Valence Shell Electron Pair Repulsion (VSEPR)

The isolobal analogy (△) **p. 578**

The Born–Oppenheimer (or adiabatic) approximation is the central point of this book (note its position in the TREE). Thanks to the approximation, we can consider *separately* two *coupled* problems concerning molecules:

- the motion of the electrons at fixed positions of the nuclei (to obtain the electronic energy),
- the motion of nuclei in the potential representing the electronic energy of the molecule (see Chapter 7).

From now on we will concentrate on the *motion of the electrons at fixed positions of the nuclei* (the Born–Oppenheimer approximation, p. 320).

To solve the corresponding Eq. (6.18), we have at our disposal the variational and the perturbation methods. The latter one should have a reasonable starting point (i.e., an unperturbed system). This is not the case in the problem we want to consider at the moment. Thus, only the variational method remains. If so, a class of the trial functions should be proposed. In this chapter the trial wave function will have a very specific form, bearing significant importance for the theory. We mean here the so-called Slater determinant, which is composed of molecular orbitals (MOs). At a certain level of approximation, each MO is a "home" for two electrons. We will now learn how to get the optimum MOs (Hartree–Fock method). In spite of some quite complex formulae, which will appear below, the main idea behind them is extremely simple. It can be expressed in the following way.

Let us consider road traffic; the cars (electrons) move at fixed positions of buildings (nuclei). The safe motion of the cars looks very complex indeed (as it does for the electrons). Is there any way at all to design such a complex many-body and safe motion? Well, one may radically simplify the problem and use the so-called mean-field approximation (paying necessarily the price of lower quality). In the mean-field approximation method we focus on the motion of *one car only*, considering its motion in such a way that the car avoids *those streets that are known as most jammed*. In this chapter we will treat the electrons in a similar manner (leaving the difficulties of considering the correlation of the motions of the electrons to Chapter V2-2). Now, the electrons will not feel the true electric field of the other electrons (as they should in a precise approach), but rather their static *mean* electric field, i.e., averaged over their motions. The idea looks hopeless, but it turns out to be a very effective and fruitful approach.

Translating this into quantum mechanical language, the underlying assumptions of the mean-field method for N identical particles (here: electrons) are the following:

- there is a certain "effective" *one-particle* operator $\hat{F}(i)$ of an identical mathematical form for all particles $i = 1, 2, ...N$, which has the eigenfunctions ϕ_k, i.e., $\hat{F}\phi_k = \varepsilon_k\phi_k$, such that
- $\left\langle \Psi|\hat{H}\Psi\right\rangle \approx \left\langle \tilde{\Psi}|\hat{H}^{ef}\tilde{\Psi}\right\rangle$, where $\tilde{\Psi}$ (normalized) is a wave function that approximates the exact wave function Ψ for the total system, $\hat{H}$ is the electronic Hamiltonian (in the clamped nuclei approximation, Chapter 6), and $\hat{H}^{ef} = \sum_{i=1}^{N} \hat{F}(i)$. In such a case the eigenvalue equation $\hat{H}^{ef}\Pi_{i=1}^{N}\phi_i(i) = E_0\Pi_{i=1}^{N}\phi_i(i)$ holds, and the approximate total energy is equal to $E_0 = \sum_{i=1}^{N}\varepsilon_k$, *as if the particles were independent*.

Any mean-field method needs to solve two problems:

- How should $\tilde{\Psi}$ be constructed using N eigenfunctions ϕ_k?
- What is the form of the one-particle effective operator $\hat{F}$?

These questions will be answered in the present chapter.

Such effectively independent, yet interacting particles, are called quasiparticles or – as we sometimes use to say – bare particles dressed up by the interaction with others.

It is worth remembering that the mean-field method bears several different names in chemistry:

- one-determinant approximation,
- one-electron approximation,
- one-particle approximation,
- molecular orbital method,
- independent-particle approximation,
- mean-field approximation,
- Hartree–Fock method,
- self-consistent field method (as regards practical solutions).

It will be shown how the mean-field method implies that milestone of chemistry: the periodic table of chemical elements.

Next, we will endeavor to understand *why* two atoms create a chemical bond, and also what affects the ionization energy and the electron affinity of a molecule.

Then, still within the MO scheme, we will show how we can reach a localized description of a molecule, with chemical bonds between *some* atoms, with the inner electronic shells, and the lone electronic pairs. The last terms are elements of a rich and very useful language commonly used by chemists.

Why this is important?

Contemporary quantum chemistry uses better methods than the mean field, described in this chapter. We will get know them in Chapters V2-2 and V2-3. Yet all these methods *start from the mean-field approximation* and in most cases they only perform cosmetic changes in energy and in electron distribution.

For example, the methods described here yield about 99% of the total energy of a system.[1] There is one more reason for why this chapter is important. Methods beyond the one-electron approximation are – computationally – very time consuming (hence they may be applied only to small systems), while the molecular orbital approach is the "daily bread" of quantum chemistry. It is a sort of standard method, and the standards have to be learned.

What is needed?

- Postulates of quantum chemistry (Chapter 1, necessary),
- operator algebra, Hermitian operators (Appendix B, p. 595, necessary),
- complete set of functions (Chapter 1, necessary),
- Hilbert space (Appendix B, p. 595, recommended),
- determinants (Appendix A, p. 589, absolutely necessary),
- Slater–Condon rules (Appendix N, p. 707, only the results are needed),
- Lagrange multipliers (Appendix O, p. 719, necessary),
- Mulliken population analysis (Appendix V2-D, p. V2-595, occasionally used).

Classical works

This chapter deals with the basic theory explaining the electronic structure of atoms and molecules. This is why we begin with Dimitryi Ivanovitch Mendeleev, who discovered in 1865, when writing his textbook "*Osnovy Khimii*" ("*Principles of Chemistry*"), St. Petersburg, Tovarishchestvo Obshchestvennaya Polza, 1869-1871, his famous periodic table of elements – one of the greatest human achievements. ★ Gilbert Newton Lewis in the paper "*The Atom and the Molecule*," published in the *Journal of the American Chemical Society*, 38(1916)762, and Walter Kossel in the article "*Über die Molekülbildung als Frage des Atombaus*," published in *Annalen der Physik*, 49(1916)229, introduced such important algorithmic theoretical tools as the octet rule and stressed the importance of the noble gas electronic configurations. ★ As soon as quantum mechanics was formulated in 1926, Douglas R. Hartree published several papers in the *Proceedings of the Cambridge Philosophical Society*, 24(1927)89, 24(1927)111, 26(1928)89, entitled "*The Wave Mechanics of an Atom with a Non-Coulomb Central Field*," containing the computations for large atoms such as Rb and Cl. These were *self-consistent ab initio*[2] computations, and the wave function was assumed to be the *product* of spin orbitals. ★ The LCAO approximation (for the solid state) was introduced by Felix Bloch in his PhD thesis "*Über die Quantenmechanik der Elektronen in Kristallgittern*," University of Leipzig, 1928, and three years later Erich Hückel used this method to describe the first molecule (benzene) in a publication "*Quantentheoretische Beitrage zum Benzolproblem. I. Die Elektronenkonfiguration des Benzols*," which appeared in *Zeitschrift für Physik*, 70(1931)203. ★ Vladimir

---

[1] In physics and chemistry we are seldom interested in the total energy. The energy differences of various states are of importance. Sometimes such precision is not enough, but the result speaks for itself.

[2] That is, derived from the first principles of (nonrelativistic) quantum mechanics! Note that these young people worked incredibly fast (no email, no PCs).

Fock introduced the antisymmetrization of the spin orbital product in his publication "*Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems*," in *Zeitschrift für Physik*, 61(1930)126 and *ibid*. 62(1930)795. ★ John Slater proposed the idea of the multiconfigurational wave function ("*Cohesion in Monovalent Metals*," *Physical Review*, 35(1930)509). ★ The Hartree–Fock method in the LCAO approximation was formulated by Clemens C.J. Roothaan in his work "*New Developments in Molecular Orbital Theory*," published in the *Reviews of Modern Physics*, 23(1951)69, and, independently, by George G. Hall in the paper "*The Molecular Orbital Theory of Chemical Valency*," in *Proceedings of the Royal Society (London) A*, 205(1951)541. ★ The physical interpretation of the orbital energies in the Hartree–Fock method was given by Tjalling C. Koopmans in his *only* quantum chemical paper "*On the Assignment of Wave Functions and Eigenvalues to the Individual Electron of an Atom*," published in *Physica*, 1(1933/1934)104. ★ The first localized orbitals (for the methane molecule) were computed by Charles A. Coulson despite the difficulties of war time (*Transactions of the Faraday Society*, 38(1942)433). ★ Hideo Fukutome, first in *Progress in Theoretical Physics*, 40(1968)998, and then in several following papers, analyzed general solutions for the Hartree–Fock equations from the symmetry viewpoint, and showed exactly eight classes of such solutions.

* * *

In the previous chapter the motion of nuclei was considered. In the Born–Oppenheimer approximation (Chapter 6) the motion of the nuclei takes place in the potential, which is the electronic energy of a system (being a function of the nuclei position, $\mathbf{R}$, in the configurational space). The electronic energy $E_k^0(\mathbf{R})$ is an eigenvalue given in Eq. (6.8) (adapted to the polyatomic case, hence $R \to \mathbf{R}$): $\hat{H}_0 \psi_k(\mathbf{r};\mathbf{R}) = E_k^0(\mathbf{R})\psi_k(\mathbf{r};\mathbf{R})$. We will now deal exclusively with this equation, i.e., we will consider the electronic motion at fixed positions of the nuclei (clamped nuclei). Thus, our goal is two-fold: we are interested in what the electronic structure looks like and in how the electronic energy depends on the positions of the nuclei.[3]

Any theoretical method applicable to molecules may be also used for atoms, although very accurate wave functions, even for simple atoms, are not easy to calculate.[4] In fact for atoms we know the solutions quite well only in the mean-field approximation, i.e., the atomic orbitals (AOs). Such orbitals play an important role as building blocks of many-electron wave functions.

---

[3] Due to tradition in the previous chapter the ground-state electronic energy $E_0^0(\mathbf{R})$ was denoted as $V(\mathbf{R})$.

[4] If an atom is considered in the Born–Oppenheimer approximation, the problem is even simpler, the electronic equation also holds; we can then take, e.g., $\mathbf{R} = \mathbf{0}$. People still try to compute correlated wave functions (i.e., beyond the mean-field approximation, see Chapter V2-2) for heavier atoms. Besides, relativistic effects (see Chapter 3) play increasingly important roles for such atoms. Starting with magnesium, they are larger than the correlation corrections. Fortunately, the relativistic corrections for atoms are largest for the inner electronic shells, which are least important for chemists.

8.1 Hartree–Fock method – a bird's eye view

Douglas R. Hartree (1897–1958) was born and died in Cambridge. He was a British mathematician and physicist, professor at Manchester University, and then professor of mathematical physics at Cambridge. Until 1921 his interest was in the development of numerical methods for antiaircraft artillery (he had some experience from the First World War), but a lecture by Niels Bohr completely changed his career. Hartree immediately started investigating atoms. He used the atomic wave function in the form of the spin orbital *product*. Hartree learned to use machines to solve differential equations while in Boston, and then he built one for himself at Cambridge. The machine was invented by Lord Kelvin, and constructed by Vannevar Bush in the USA. The machine integrated equations using a circle which rolled on a rotating disc. Later the first electronic computer, ENIAC was used, and Hartree was asked to come and help to compute missile trajectories.

An excerpt from "*Solid State and Molecular Theory*," Wiley, London, 1975 by John C. Slater: "*Douglas Hartree was very distinctly of the matter-of-fact habit of thought that I found most congenial. The hand-waving magical type of scientist regarded him as a "mere computer." Yet he made a much greater contribution to our knowledge of the behavior of real atoms than most of them did. And while he limited himself to atoms, his demonstration of the power of the self-consistent field (SCF) for atoms is what has led to the development of that method for molecules and solids as well.*"

Before introducing the detailed formalism of the Hartree–Fock method, let us first look at its principal features. It will help us to understand our mathematical goal.

Vladimir A. Fock (1898–1974), Russian physicist, professor at the Leningrad University (St. Petersburg), led investigations on quantum mechanics, gravity theory, and general relativity theory, and in 1930, while explaining atomic spectra, he invented the *antisymmetrization* of the spin orbitals product.

8.1.1 Spin orbitals as the one-electron building blocks

Although this comparison is not precise, the electronic wave function for a molecule is built of segments, as a house is constructed of bricks.

The electronic wave function of a molecule containing N electrons depends on $3N$ Cartesian coordinates of the electrons and on their N spin coordinates (for each electron, its $\sigma = \frac{1}{2}$ or $-\frac{1}{2}$). Thus, it is a function of position in $4N$-dimensional space. This function will be created out of simple "bricks," i.e., *molecular spin orbitals*. Each of those will be a function of the coordinates of *one* electron only: three Cartesian coordinates and one spin coordinate (cf. Chapter 1). A spin orbital is therefore a function of the coordinates in the four-dimensional space,[5] and in the most general case a normalized spin orbital reads as (Fig. 8.1)

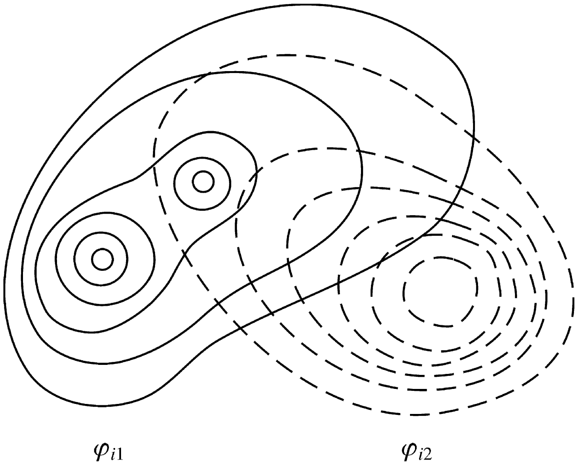


Fig. 8.1. According to Eq. (8.1) a spin orbital is a mixture of α and β orbital components: $\varphi_{i1}(\mathbf{r})$ and $\varphi_{i1}(\mathbf{r})$, respectively. The figure shows two sections of such a spin orbital (z denotes the Cartesian axis perpendicular to the plane of the page): section $z = 0, \sigma = \frac{1}{2}$ (solid isolines) and section $z = 0, \sigma = -\frac{1}{2}$ (dashed isolines). In practical applications most often a restricted form of spin orbitals is used: either $\varphi_{i1} = 0$ or $\varphi_{i2} = 0$, i.e., a spin orbital is taken as an orbital part times spin function α or β.

$$\phi_i(\mathbf{r}, \sigma) = \varphi_{i1}(\mathbf{r})\alpha(\sigma) + \varphi_{i2}(\mathbf{r})\beta(\sigma), \tag{8.1}$$

---

[5] The analogy of houses and bricks fails here, because both the house and the bricks come from the same three-dimensional space.

where the orbital components φ_{i1} and φ_{i2} (square integrable functions) that depend on position $\mathbf{r}$ of the electron can adopt *complex* values, while the spin functions α and β, which depend on the spin coordinate σ, are defined in Chapter 1, p. 33.

> In the vast majority of quantum mechanical calculations the spin orbital ϕ_i is a *real* function, and φ_{i1} and φ_{i2} are such that either $\varphi_{i1} = 0$ or $\varphi_{i2} = 0$.

Yet for the time being we do not introduce any significant[6] *restrictions for the spin orbitals.* Spin orbital ϕ_i will adopt different complex values for various spatial coordinates as well as for a given value[7] of the spin coordinate σ.

8.1.2 Variables

Thus the variables, which the wave function depends on, are as follows:

$x_1,\ y_1,\ z_1,\ \sigma_1$ or briefly 1,
$x_2,\ y_2,\ z_2,\ \sigma_2$ or briefly 2,
....
$x_N,\ y_N,\ z_N,\ \sigma_N$ or briefly N,

where x_i, y_i, z_i are the Cartesian coordinates and σ_i is the spin coordinate of electron i.

An exact wave function $\psi(x_1,\ y_1,\ z_1,\ \sigma_1, x_2,\ y_2,\ z_2,\ \sigma_2, ...) = \psi(1, 2, ..., N)$ belongs (see Fig. 8.2) to the set Ω which is the common part of the following sets:

- set Q of all square integrable functions,
- set $\Omega_{variables}$ of all the functions dependent on the abovementioned variables,
- set $\Omega_{antisym}$ of all the functions which are antisymmetric with respect to the mutual exchange of the coordinates of any two electrons (p. 40),

$$\psi \in \Omega = Q \cap \Omega_{variables} \cap \Omega_{antisym}.$$

8.1.3 Slater determinant

There should be something in the theory which ensures that if we renumber the electrons, no theoretical prediction will change. The postulate of the *antisymmetric* character of the wave

---

[6] The normalization condition does not reduce the generality of the approach.

[7] That is, we put $\sigma = \frac{1}{2}$ or $\sigma = -\frac{1}{2}$.

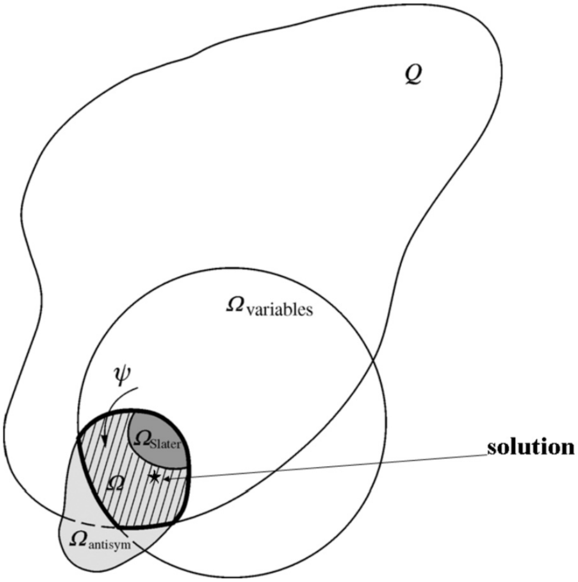


Fig. 8.2. Diagram of the sets, among which the solution ψ of the Schrödinger equation is sought. The Q set is the one of all square integrable functions, $\Omega_{variables}$ is the set of the functions with variables as those of the solution of Schrödinger equation, ψ, and $\Omega_{antisym}$ is the set of the functions which are antisymmetric with respect to the exchange of coordinates of any two electrons. The solutions of the Schrödinger equation, ψ, will be sought in the common part of these three sets: $\psi \in \Omega = Q \cap \Omega_{variables} \cap \Omega_{antisym}$. The Ω_{Slater} represents the set of single Slater determinants built of normalizable spin orbitals. The exact wave function ψ always belongs to $\Omega - \Omega_{Slater}$.

function with respect to the *exchange* of the coordinates of *any* two electrons certainly ensures this (Chapter 1, p. 33). The solution of the Schrödinger equation for a given stationary state of interest should be sought amongst *such* functions.

A Slater determinant is a function of the coordinates of N electrons, which *automatically* belongs to Ω:

$$\psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \dots & \dots & \dots & \dots \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}, \tag{8.2}$$

where ϕ_i are the orthonormal[8] one-electron functions, i.e., molecular spin orbitals.[9] The Slater determinants form a subset $\Omega_{Slater} \subset \Omega$.

John C. Slater (1901–1976), American physicist, for 30 years a professor and dean at the Physics Department of the Massachusetts Institute of Technology, then at the University of Florida Gainesville and the Quantum Theory Project at this school. His youth was in the stormy period of the intense development of quantum mechanics, and he participated vividly in it. For example, in 1926–1932 he published articles on the ground state of the helium atom, on the screening constants (Slater orbitals), on the antisymmetrization of the wave function (Slater determinant), and on the algorithm for calculating the integrals (the Slater–Condon rules).

A Slater determinant carries two important attributes of the exact wave function:

- Suppose we want to calculate the probability density that two electrons with the same spin coordinate σ are in the same place, i.e., that such two electrons have *all* their coordinates (spatial and spin ones) identical. If so, then the two columns of the abovementioned determinant are identical. And this means that the determinant becomes equal to zero.[10] From this and from the continuity of the wave function we may conclude that

  > electrons of the same spin cannot approach each other.

- Let us now imagine two electrons with opposite values of their spin coordinate σ. If these two electrons take the same position in space, the Slater determinant will not vanish, because in the general case there is nothing that forces $\phi_i(1)$ to be equal to $\phi_i(2)$, when

---

[8] It is *most often* so, and then the factor standing in front of the determinant ensures the normalization. The spin orbitals could be *nonnormalized* (but if they are to describe a stationary state, they should be square integrable). They also do not need to be mutually orthogonal, but certainly they *need to be linearly independent*. Any attempt to insert linearly dependent functions in the determinant will have a “tragic outcome” – we will get 0. This is due to the properties of the determinant (if a row is a linear combination of the others, the determinant is zero). It also follows that if we have a set of nonorthogonal spin orbitals in a Slater determinant, we could orthogonalize them by making the appropriate linear combinations. This would multiply the original Slater determinant by an irrelevant constant. This is why *it is no loss of generality to require the spin orbitals to be orthonormal*.

[9] In the theory of the atomic nucleus, the determinant wave function for the nucleons (fermions) is also used.

[10] Indeed, this is why we exist. Two objects built out of fermions (e.g., electrons) cannot occupy the same position in space. If it were not so, our bodies would sink in the ground.

$1 \equiv (\mathbf{r}_1, \sigma = \frac{1}{2})$ and $2 \equiv (\mathbf{r}_1, \sigma = -\frac{1}{2})$ for $i = 1, 2, ...$ From this and from the continuity of the wave function, we conclude that

electrons of opposite spins can approach each other.

8.1.4 What is the Hartree–Fock method all about?

The Hartree–Fock method is a *variational* one (p. 265), i.e., minimizes the mean value of the Hamiltonian $\varepsilon = \frac{\left\langle\psi|\hat{H}\psi\right\rangle}{\langle\psi|\psi\rangle}$ producing the Hartree–Fock energy $\varepsilon_{\min} = E_{HF} = \frac{\left\langle\psi_{HF}|\hat{H}\psi_{HF}\right\rangle}{\langle\psi_{HF}|\psi_{HF}\rangle}$ and uses the variational wave function ψ in the form of a *single Slater determinant*.

The Slater determinant is an antisymmetric function, but an antisymmetric function does not necessarily need to take the shape of a Slater determinant.

Taking the variational wave function in the form of one determinant means an automatic limitation to the subset Ω_{Slater} for searching for the optimum wave function. In fact, we should search the optimum wave function in the set Ω. Thus it is an *approximation* for the solution of the Schrödinger equation, with no chance of receiving the exact result.

Why are Slater determinants used so willingly? There are two reasons for this:

- A determinant is a kind of "template." Whatever you put inside, the result (if not zero) is antisymmetric *by definition*, i.e., it automatically satisfies one of the postulates of quantum mechanics.
- It is constructed of simple "bricks" – the one-electron functions (spin orbitals).

The Slater determinants built out of the complete set of spin orbitals do form the complete set.

Because of this, the true wave function can take the form of a linear combination of the determinants (we will discuss this in Chapter V2-2).[11]

---

[11] But never as one determinant, because the Hamiltonian of a molecule never represents the sum of the effective Hamiltonians for the individual electrons.

AN ANALOGY – THE COUPLED HARMONIC OSCILLATORS (to be continued, 1 of 5): The derivation of the Fock equation given below looks more complex than it really is. We decided to go in parallel and illustrate all steps of the Fock equation derivation in a much simpler case of the two coupled harmonic (bosonic, one-dimensional) oscillators. These illustrations will be given as the inserts like the present one (and in the same background gray color). Thus, reading only the inserts in the text the reader is able to catch the very essence of what I am doing in the main text.

The Hamiltonian for the two oscillators will be given by $\hat{H} = \hat{T} + \hat{V}$, where $\hat{T} = -\frac{\hbar^2}{2m_1}\frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m_2}\frac{\partial^2}{\partial x_2^2}$ and $V = \frac{1}{2}kx_1^2 + \frac{1}{2}kx_2^2 + \lambda x_1^4 x_2^4$, with $\lambda x_1^4 x_2^4$ as the coupling term. Considering the bosonic nature of the particles (the wave function has to be symmetric, see Chapter 1), we will use $\psi = \phi(1)\phi(2)$ as a variational function, where ϕ is a normalized spin orbital. This represents a restriction analogous to taking a single Slater determinant, because an exact wave function should not be $\psi = \phi(1)\phi(2)$, but rather $\Sigma_{ij}\phi_i(1)\phi_j(2)$.

8.2 Towards the optimal spin orbitals and the Fock equation

8.2.1 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} the distance between electrons 1 and 2).

For the one-electron integrals,

$$\left\langle i|\hat{h}|j\right\rangle \equiv \sum_{\sigma_1}\int dx_1 dy_1 dz_1 \phi_i^*(1)\hat{h}\phi_j(1), \tag{8.3}$$

and for the two-electron integrals,

$$\langle ij|kl\rangle \equiv \sum_{\sigma_1}\sum_{\sigma_2}\int dx_1 dy_1 dz_1 \int dx_2 dy_2 dz_2 \phi_i^*(1)\phi_j^*(2)\frac{1}{r_{12}}\phi_k(1)\phi_l(2). \tag{8.4}$$

The integrals over the spatial coordinates (only) will be denoted by round brackets (). For the one-electron integrals,

$$\left(i|\hat{h}|j\right) \equiv \int dx_1 dy_1 dz_1 \varphi_i^*(1)\hat{h}\,(1)\,\varphi_j(1), \tag{8.5}$$

and for the two-electron integrals,

$$(ij|kl) \equiv \int dx_1dy_1dz_1 \int dx_2dy_2dz_2\varphi_i^*(1)\varphi_j^*(2)\frac{1}{r_{12}}\varphi_k(1)\varphi_l(2). \tag{8.6}$$

This is called the Dirac notation (of the integrals).[12] See Fig. 8.3.

8.2.2 Energy functional to be minimized

Applying the first Slater–Condon rule[13] we get the following equation for *the mean value of the Hamiltonian* (without constant nuclear repulsion) calculated using the normalized *Slater one-determinant function* ψ, i.e., the energy functional $E\left[\psi\right]$,

$$E\left[\psi\right] = \langle\psi|\hat{H}|\psi\rangle = \sum_{i=1}^{N}\langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{N}(\langle ij|ij\rangle - \langle ij|ji\rangle), \tag{8.7}$$

where the indices symbolize the spin orbitals, and

$$\hat{h}(1) = -\frac{1}{2}\Delta_1 - \sum_{a=1}^{M}\frac{Z_a}{r_{a1}}, \tag{8.8}$$

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).

AN ANALOGY – THE COUPLED HARMONIC OSCILLATORS (continued, 2 of 5): The expression for the mean value of the Hamiltonian takes the form $E[\phi] = \left\langle\psi|\hat{H}\psi\right\rangle = \left\langle\phi(1)\phi(2)|\left(\hat{h}(1)+\hat{h}(2)\right)\phi(1)\phi(2)\right\rangle + \lambda\left\langle\phi(1)\phi(2)|x_1^4x_2^4\phi(1)\phi(2)\right\rangle = \left\langle\phi(1)\phi(2)|\hat{h}(1)\phi(1)\phi(2)\right\rangle + \left\langle\phi(1)\phi(2)|\hat{h}(2)\phi(1)\phi(2)\right\rangle + \lambda\left\langle\phi(1)|x_1^4\phi(1)\right\rangle\left\langle\phi(2)|x_2^4\phi(2)\right\rangle = \left\langle\phi(1)|\hat{h}(1)\phi(1)\right\rangle + \left\langle\phi(2)|\hat{h}(2)\phi(2)\right\rangle + \lambda\left\langle\phi(1)|x_1^4\phi(1)\right\rangle\left\langle\phi(2)|x_2^4\phi(2)\right\rangle = 2\left\langle\phi|\hat{h}\phi\right\rangle + \lambda\left\langle\phi|x^4\phi\right\rangle^2$, where the one-particle operator $\hat{h}\,(i) = -\frac{\hbar^2}{2m_i}\frac{\partial^2}{\partial x_i^2} + \frac{1}{2}kx_i^2$.

---

[12] Sometimes one uses the *Coulomb notation* $(ij|kl)_{Dirac} \equiv (ik|jl)_{Coulomb}$, also $\langle ij|kl\rangle_{Dirac} \equiv \langle ik|jl\rangle_{Coulomb}$. Coulomb notation emphasizes the physical interpretation of the two-electron integral, as the energy of the Coulombic interaction of two charge distributions $\varphi_i^*(1)\varphi_k(1)$ for electron 1 and $\varphi_j^*(2)\varphi_l(2)$ for electron 2. Dirac notation for the two-electron integrals emphasizes the two-electron functions "bra" and "ket" from the general Dirac notation (p. 23). *In the present book we will consequently use Dirac notation* (both for integrals using spin orbitals and for those using orbitals, the difference being emphasized by the type of bracket).

[13] Appendix N, p. 707; please take a look of this rule (you may leave out its derivation).

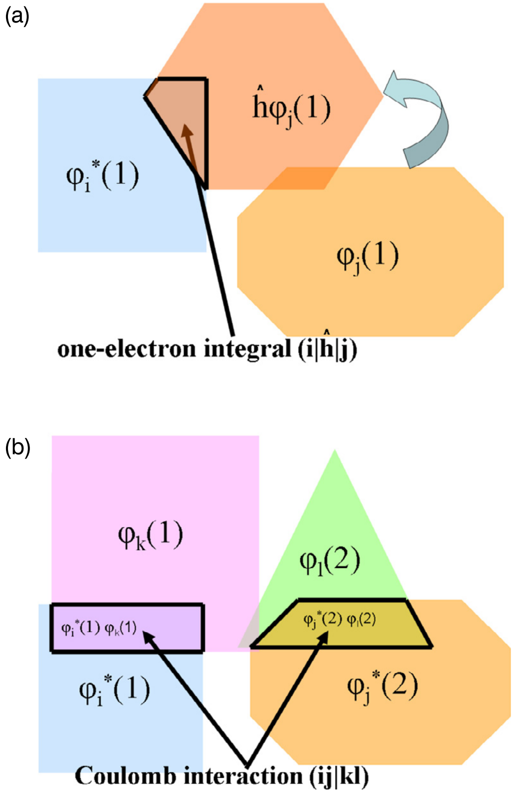


Fig. 8.3. A scheme showing physical interpretation of the one-electron and two-electron integrals (involving orbitals shown schematically in the form of some simple geometrical figures) – the Dirac notation. (a) The one-electron integral, Eq. (8.5), is equal to the overlap integral of the orbital $\varphi_i(1)$ and the function $\hat{h}(1)\varphi_j(1)$, i.e., $\varphi_j(1)$ transformed by the operator $\hat{h}$. (b) The two-electron integral $(ij|kl)$ of Eq. (8.6) represents the Coulombic interaction of two electronic clouds ("pseudocharges" here enclosed within the thick line borders): $\varphi_i^*(1)\varphi_k(1)$ of electron 1 and $\varphi_j^*(2)\varphi_l(2)$ of electron 2. These pseudocharges may have any sign that in addition may alternate in space. The latter property is a consequence of the sign alternation of the orbitals themselves (not shown in this schematic picture), necessary for their orthogonality.

8.2.3 *Energy minimization with constraints*

We would like to find such spin orbitals ("the best ones"), that *any* change in them leads to an *increase* in energy $E[\psi]$. But the changes of the spin orbitals need to be such that the above formula still holds, and it would hold only by assuming the *orthonormality* of the spin orbitals. This means that there are some constraints for the changed spin orbitals:

$$\langle i|j\rangle - \delta_{ij} = 0, \tag{8.9}$$

for $i, j = 1, 2, \ldots N$.

Thus we seek the *conditional* minimum. We will find it using the Lagrange multipliers method (Appendix O, p. 719). In this method the equations of the constraints multiplied by the Lagrange multipliers are added (or subtracted, does not matter) to the original function which is to be minimized. Then we minimize the function as if the constraints did not exist.

We do the same for the functionals. The necessary condition for the minimum is that the variation[14] of $E - \sum_{ij} L_{ij}(\langle i|j\rangle - \delta_{ij})$ is equal to zero (the numbers L_{ij} denote the Lagrange multipliers to be found).

The variation of a functional is defined as the *linear* part of the functional change coming from a change in the function which is its argument.

Variation is an analog of the differential (the differential is just the linear part of the function's change). Thus we calculate the linear part of a change (variation), i.e.,

$$\delta(E - \sum_{ij} L_{ij}\langle i|j\rangle) = 0, \tag{8.10}$$

using the (yet) undetermined Lagrange multipliers L_{ij}, and we set the variation equal to zero.[15]

AN ANALOGY – THE COUPLED HARMONIC OSCILLATORS (continued, 3 of 5): The change of E, because of the variation $\delta\phi^*$, is $E[\phi + \delta\phi] - E[\phi] = 2\left\langle\phi + \delta\phi|\hat{h}\phi\right\rangle + \lambda\left\langle\phi + \delta\phi|x^4\phi\right\rangle^2 - [2\left\langle\phi|\hat{h}\phi\right\rangle + \lambda\left\langle\phi|x^4\phi\right\rangle^2] = 2\left\langle\phi|\hat{h}\phi\right\rangle + 2\left\langle\delta\phi|\hat{h}\phi\right\rangle + \lambda\left\langle\phi|x^4\phi\right\rangle^2 + 2\lambda\left\langle\delta\phi|x^4\phi\right\rangle\left\langle\phi|x^4\phi\right\rangle + \lambda\left\langle\delta\phi|x^4\phi\right\rangle^2 - [2\left\langle\phi|\hat{h}\phi\right\rangle + \lambda\left\langle\phi|x^4\phi\right\rangle^2] = 2\left\langle\delta\phi|\hat{h}\phi\right\rangle + 2\lambda\left\langle\delta\phi|x^4\phi\right\rangle\left\langle\phi|x^4\phi\right\rangle + \lambda\left\langle\delta\phi|x^4\phi\right\rangle^2$. The linear part in $\delta\phi$ of the energy change, i.e., the variation, is, therefore, equal to $\delta E = 2\left\langle\delta\phi|\hat{h}\phi\right\rangle + 2\lambda\left\langle\delta\phi|x^4\phi\right\rangle\left\langle\phi|x^4\phi\right\rangle$. The variation $\delta\phi^*$ has, however, to ensure the normalization of ϕ, i.e., $\langle\phi|\phi\rangle = 1$. After multiplying by the unknown Lagrange multiplier chosen as 2ε, we get the extremum condition $\delta(E - 2\varepsilon\langle\phi|\phi\rangle) = 0$, i.e., $2\left\langle\delta\phi|\hat{h}\phi\right\rangle + 2\lambda\left\langle\delta\phi|x^4\phi\right\rangle\left\langle\phi|x^4\phi\right\rangle - 2\varepsilon\langle\delta\phi|\phi\rangle = 0$.

The stationarity condition for the energy functional

It is sufficient to vary *only the functions complex conjugate to the spin orbitals* or only the spin orbitals (cf. p. 266), yet the result is always the same. We decide the first.

---

[14] However, this is not a sufficient condition, because the vanishing of the differential for certain values of independent variables happens not only for minima, but also for maxima and saddle points (*stationary points*).

[15] Note that $\delta(\delta_{ij}) = 0$.

Substituting $\phi_i^* \to \phi_i^* + \delta\phi_i^*$ in (8.7) (and retaining only linear terms in $\delta\phi_i^*$ to be inserted into (8.10)) the variation takes the form (the symbols δi^* and δj^* mean $\delta\phi_i^*$ and $\delta\phi_j^*$)

$$\sum_{i=1}^{N}\Big(\langle\delta i|\hat{h}|i\rangle + \frac{1}{2}\sum_j\big(\langle\delta i, j|ij\rangle + \langle i, \delta j|ij\rangle - \langle\delta i, j|ji\rangle - \langle i, \delta j|ji\rangle - 2L_{ij}\langle\delta i|j\rangle\big)\Big) = 0. \tag{8.11}$$

Now we will try to express this in the form

$$\sum_i \langle\delta i|...\rangle = 0.$$

Since the δi^* may be arbitrary, the equation $|...\rangle = 0$ (called the Euler equation in variational calculus) results. This will be our next goal.

Noting that the sum indices and the numbering of electrons in the integrals are arbitrary we have the following equalities:

$$\sum_{ij}\langle i, \delta j|ij\rangle = \sum_{ij}\langle j, \delta i|ji\rangle = \sum_{ij}\langle\delta i, j|ij\rangle,$$
$$\sum_{ij}\langle i, \delta j|ji\rangle = \sum_{ij}\langle j, \delta i|ij\rangle = \sum_{ij}\langle\delta i, j|ji\rangle,$$

and after substitution in the expression for the variation, we get

$$\sum_i\Big(\langle\delta i|\hat{h}|i\rangle + \frac{1}{2}\sum_j\big(\langle\delta i, j|ij\rangle + \langle\delta i, j|ij\rangle - \langle\delta i, j|ji\rangle - \langle\delta i, j|ji\rangle - 2L_{ij}\langle\delta i|j\rangle\big)\Big) = 0. \tag{8.12}$$

Let us rewrite this equation in the following manner ($\langle\delta i| \equiv \langle\delta\phi_i(1)|$):

$$\sum_i\langle\delta\phi_i(1)|$$
$$\Big(\hat{h}\phi_i(1) + \sum_j\Big(\int \mathrm{d}\tau_2\frac{1}{r_{12}}\phi_j^*(2)\phi_j(2)\phi_i(1) - \int \mathrm{d}\tau_2\frac{1}{r_{12}}\phi_j^*(2)\phi_i(2)\phi_j(1) - L_{ij}\phi_j(1)\Big)\Big)\rangle_1 = 0, \tag{8.13}$$

where $\langle\delta\phi_i(1)|...\rangle_1$ means integration over spatial coordinates of electron 1 and summation over its spin coordinate, and $\mathrm{d}\tau_2$ refers to the spatial coordinate integration and spin coordinate summing for electron 2. The above must be true for *any* $\langle\delta\phi_i(1)| \equiv \delta\phi_i^*$, which means that *each individual term* in parentheses needs to be equal to zero, i.e.,

$$\hat{h}\phi_i(1) + \sum_j \left(\int d\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_j(2) \cdot \phi_i(1) - \int d\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_i(2) \cdot \phi_j(1) \right) = \sum_j L_{ij}\phi_j(1)). \tag{8.14}$$

The Coulombic and exchange operators

Let us introduce the following linear operators:

a) *two Coulombic operators: the total operator* $\hat{J}(1)$ and *the spin orbital operator* $\hat{J}_j(1)$, defined via their action on an arbitrary function $u(1)$ of the coordinates of electron 1,

$$\hat{J}(1)u(1) = \sum_j \hat{J}_j(1)u(1), \tag{8.15}$$

$$\hat{J}_j(1)u(1) = \int d\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_j(2)u(1), \tag{8.16}$$

b) and similarly, *two exchange operators: the total operator* $\hat{K}(1)$ and *the spin orbital operator* $\hat{K}_j(1)$,

$$\hat{K}(1)u(1) = \sum_j \hat{K}_j(1)u(1), \tag{8.17}$$

$$\hat{K}_j(1)u(1) = \int d\tau_2 \frac{1}{r_{12}} \phi_j^*(2)u(2)\phi_j(1). \tag{8.18}$$

Then Eq. (8.14) takes the form

$$\left(\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\right)\phi_i(1) = \sum_j L_{ij}\phi_j(1). \tag{8.19}$$

The equation is nice and concise except for one thing. It would be even nicer if the right-hand side were proportional to $\phi_i(1)$ instead of being a linear combination of all the spin orbitals. In such a case the equation would be similar to the eigenvalue problem and we would like it a lot. It would be similar but not identical, since the operators $\hat{J}$ and $\hat{K}$ include the *sought* spin orbitals ϕ_i. Because of this, the equation would be called the pseudoeigenvalue problem.

8.2.4 Slater determinant subject to a unitary transformation

How can we help? Let us note that we do not care too much about the spin orbitals themselves, because these are *by-products* of the method which is to give the optimum mean value of the Hamiltonian, and the corresponding N-electron wave function. We can choose some other spin orbitals, such that the mean value of the Hamiltonian as well as the wave function do not change and the Lagrange multipliers L_{ij} are diagonal. Is this at all possible? Let us see.

Let us imagine the linear transformation of spin orbitals ϕ_i, i.e., in matrix notation,

$$\boldsymbol{\phi}' = \mathbf{A}\boldsymbol{\phi}, \tag{8.20}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\phi}'$ are vertical vectors containing components ϕ_i. A vertical vector is uncomfortable for typography, in contrast to its transposition (a horizontal vector), and it is easier to write the transposed vector: $\boldsymbol{\phi}'^T = [\ \phi'_1,\ \ \phi'_2,\ \ \ldots,\ \ \phi'_N\ ]$ and $\boldsymbol{\phi}^T = [\ \phi_1,\ \ \phi_2,\ \ \ldots,\ \ \phi_N\ ]$. If we construct the determinant built of spin orbitals $\boldsymbol{\phi}'$ and not of $\boldsymbol{\phi}$, an interesting chain of transformations will result:

$$\frac{1}{\sqrt{N!}}\begin{vmatrix} \phi'_1(1) & \phi'_1(2) & \ldots & \phi'_1(N) \\ \phi'_2(1) & \phi'_2(2) & \ldots & \phi'_2(N) \\ \ldots & \ldots & \ldots & \ldots \\ \phi'_N(1) & \phi'_N(2) & \ldots & \phi'_N(N) \end{vmatrix} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \sum_i A_{1i}\phi_i(1) & \ldots & \sum_i A_{1i}\phi_i(N) \\ \sum_i A_{2i}\phi_i(1) & \ldots & \sum_i A_{2i}\phi_i(N) \\ \ldots & \ldots & \ldots \\ \sum_i A_{Ni}\phi_i(1) & \ldots & \sum_i A_{Ni}\phi_i(N) \end{vmatrix} \tag{8.21}$$

$$= \frac{1}{\sqrt{N!}}\det\left\{\mathbf{A}\begin{bmatrix} \phi_1(1) & \phi_1(2) & \ldots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \ldots & \phi_2(N) \\ \ldots & \ldots & \ldots & \ldots \\ \phi_N(1) & \phi_N(2) & \ldots & \phi_N(N) \end{bmatrix}\right\}$$

$$= \det\mathbf{A}\cdot\frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(1) & \phi_1(2) & \ldots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \ldots & \phi_2(N) \\ \ldots & \ldots & \ldots & \ldots \\ \phi_N(1) & \phi_N(2) & \ldots & \phi_N(N) \end{vmatrix}. \tag{8.22}$$

We have therefore obtained our initial Slater determinant multiplied by a *number:* det**A**. Thus, provided that det**A** is not zero,[16]

> the new wave function would provide the same mean value of the Hamiltonian.

The only problem from such a transformation is loss of the normalization of the wave function. Yet we may even preserve the normalization. Let us choose such a matrix **A** that $|\det\mathbf{A}| = 1$.

---

[16] The **A** transformation thus cannot be singular (see Appendix A, p. 589).

This condition will hold if $\mathbf{A} = \mathbf{U}$, where $\mathbf{U}$ is a *unitary* matrix (when $\mathbf{U}$ is real, we call $\mathbf{U}$ an *orthogonal* transformation).[17] This means that

> if a unitary transformation $\mathbf{U}$ is performed on the orthonormal spin orbitals, then the new spin orbitals ϕ' are also orthonormal.

This is why a unitary transformation is said to represent a rotation in the Hilbert space: the mutually orthogonal and perpendicular vectors do not lose these features upon rotation.[18] This can be verified by a direct calculation:

$$\left\langle \phi_i'(1)|\phi_j'(1)\right\rangle = \left\langle \sum_r U_{ir}\phi_r(1)|\sum_s U_{js}\phi_s(1)\right\rangle =$$
$$= \sum_{rs} U_{ir}^* U_{js} \left\langle \phi_r(1)|\phi_s(1)\right\rangle = \sum_{rs} U_{ir}^* U_{js}\delta_{rs}$$
$$= \sum_r U_{ir}^* U_{jr} = \delta_{ij}.$$

Thus, in the case of a unitary transformation even the normalization of the total one-determinant wave function is preserved; at worst the phase χ of this function will change (while $\exp(i\chi) = \det \mathbf{U}$), and this factor does not change either $|\psi|^2$ or the mean value of the operators.

8.2.5 The $\hat{J}$ and $\hat{K}$ operators are invariant

How does the Coulombic operator change upon a unitary transformation of the spin orbitals? Let us see. We have

$$\hat{J}(1)'\chi(1) = \int d\tau_2 \frac{1}{r_{12}} \sum_j \phi_j'^*(2)\phi_j'(2)\chi(1)$$
$$= \int d\tau_2 \frac{1}{r_{12}} \sum_j \sum_r U_{jr}^* \phi_r^*(2) \sum_s U_{js}\phi_s(2)\chi(1)$$

---

[17] For a unitary transformation $\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{1}$. The matrix $\mathbf{U}^\dagger$ arises from $\mathbf{U}$ via the exchange of rows and columns (this does not influence the value of the determinant), and via the complex conjugation of all elements (and this gives $\det\mathbf{U}^\dagger = (\det\mathbf{U})^*$). Finally, since $(\det\mathbf{U})(\det\mathbf{U}^\dagger) = 1$ we have $|\det\mathbf{U}| = 1$.

[18] Just as three fingers held at right angles do not cease to be of the same length (normalization) after rotation of your palm and continue to be orthogonal.

$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_{r,s} (\sum_j U_{js} U_{jr}^*) \phi_r^*(2)\phi_s(2)\chi(1)$$
$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_{r,s} (\sum_j U_{rj}^\dagger U_{js}) \phi_r^*(2)\phi_s(2)\chi(1)$$
$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_{r,s} \delta_{sr} \phi_r^*(2)\phi_s(2)\chi(1)$$
$$= \int \mathrm{d}\tau_2 \frac{1}{r_{12}} \sum_{r} \phi_r^*(2)\phi_r(2)\chi(1) = \hat{J}(1)\chi(1).$$

The operator $\hat{J}(1)'$ proves to be *identical* with the operator $\hat{J}(1)$. Similarly we may prove *the invariance of the operator K.*

The operators $\hat{J}$ and $\hat{K}$ are invariant with respect to any unitary transformation of the spin orbitals. In conclusion, while deriving the new spin orbitals from a unitary transformation of the old ones, we do not need to worry about $\hat{J}$ and $\hat{K}$ since they remain the same.

8.2.6 Diagonalization of the Lagrange multipliers

Eq. (8.19) may be written in matrix form, i.e.,

$$[\hat{h}(1) + \hat{J}(1) - \hat{K}(1)]\boldsymbol{\phi}(1) = \mathbf{L}\boldsymbol{\phi}(1), \tag{8.23}$$

where $\boldsymbol{\phi}$ is a column of spin orbitals. Transforming $\boldsymbol{\phi} = \mathbf{U}\boldsymbol{\phi}'$ and multiplying the Fock equation by $\mathbf{U}^\dagger$ (where $\mathbf{U}$ is a unitary matrix), we obtain

$$\mathbf{U}^\dagger[\hat{h}(1) + \hat{J}(1) - \hat{K}(1)]\mathbf{U}\boldsymbol{\phi}(1)' = \mathbf{U}^\dagger\mathbf{L}\mathbf{U}\boldsymbol{\phi}(1)', \tag{8.24}$$

because $\hat{J}$ and $\hat{K}$ did not change upon the transformation.

The $\mathbf{U}$ matrix *may* be chosen such that $\mathbf{U}^\dagger\mathbf{L}\mathbf{U}$ represents the *diagonal* matrix.

Its diagonal elements[19] will now be denoted as ε_i. Because $\hat{h}(1) + \hat{J}(1) - \hat{K}(1)$ is a linear operator we get equation

$$\mathbf{U}^\dagger \mathbf{U}\big(\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\big)\boldsymbol{\phi}(1)' = \mathbf{U}^\dagger \mathbf{L}\mathbf{U}\boldsymbol{\phi}(1)', \tag{8.25}$$

or alternatively

$$\big(\hat{h}(1) + \hat{J}(1) - \hat{K}(1)\big)\boldsymbol{\phi}(1)' = \boldsymbol{\varepsilon}\boldsymbol{\phi}(1)', \tag{8.26}$$

where $\varepsilon_{ij} = \varepsilon_i \delta_{ij}$.

8.2.7 Optimal spin orbitals are solutions of the Fock equation (general Hartree–Fock [GHF] method)

We leave out the "prime" to simplify the notation[20] and write *the Fock equation* for a single spin orbital:

THE FOCK EQUATION IN THE GENERAL HARTREE–FOCK (GHF) METHOD

$$\hat{F}(1)\phi_i(1) = \varepsilon_i \phi_i(1), \tag{8.27}$$

where the Fock operator $\hat{F}$ is

$$\hat{F}(1) = \hat{h}(1) + \hat{J}(1) - \hat{K}(1). \tag{8.28}$$

These ϕ_i are called *canonical spin orbitals*, and they are the solution of the Fock equation; ε_i is the *orbital energy* corresponding to the spin orbital ϕ_i. It is indicated in brackets that both the Fock operator and the molecular spin orbital depend on the coordinates of one electron only (represented by electron 1).

---

[19] Such diagonalization is possible because $\mathbf{L}$ is a Hermitian matrix (i.e., $\mathbf{L}^\dagger = \mathbf{L}$), and each Hermitian matrix may be diagonalized via the transformation $\mathbf{U}^\dagger\mathbf{L}\mathbf{U}$ with the unitary matrix $\mathbf{U}$. Matrix $\mathbf{L}$ is indeed Hermitian. It is clear when we write the complex conjugate of the variation $\delta(E - \sum_{ij} L_{ij}\langle i|j\rangle) = 0$. This gives $\delta(E - \sum_{ij} L_{ij}^*\langle j|i\rangle) = 0$, because E is real, and after the change of the summation indices $\delta(E - \sum_{ij} L_{ji}^*\langle i|j\rangle) = 0$. Thus, $L_{ij} = L_{ji}^*$, i.e., $\mathbf{L} = \mathbf{L}^\dagger$.

[20] This means that we finally forget about ϕ' (we pretend that they have never appeared), and we will deal only with such ϕ as corresponding to the *diagonal* matrix of the Lagrange multipliers.

AN ANALOGY – THE COUPLED HARMONIC OSCILLATORS (continued, 4 of 5): The equation for the conditional extremum may be rewritten as $2\left\langle\delta\phi|\left[\hat{h}+\lambda\bar{x}^4x^4-\varepsilon\right]\phi\right\rangle=0$, where $\bar{x}^4=\left\langle\phi|x^4\phi\right\rangle$ is a *number*. This gives (remember that $\delta\phi^*$ is arbitrary) the Euler equation $\left[\hat{h}+\lambda\bar{x}^4x^4-\varepsilon\right]\phi=0$, i.e., the analog of the Fock equation (8.28): $\hat{F}\phi=\varepsilon\phi$, with the operator $\hat{F}=\hat{h}+\lambda\bar{x}^4x^4$. Let us emphasize that the operator $\hat{F}$ is a one-particle operator, via the notation $\hat{F}\,(1)\,\phi\,(1)=\varepsilon\phi\,(1)$, while $\hat{F}\,(1)=\hat{h}\,(1)+\lambda\bar{x}^4x_1^4$. The term $\lambda x_1^4x_2^4$ is analogous to the term $\frac{1}{r_{12}}$, the term $\lambda\bar{x}^4x_1^4\phi(1)$ may be rewritten as $\int \mathrm{d}x_2\lambda x_1^4x_2^4\phi^*(2)\phi(2)\phi(1)$, which for the coupled oscillators might be written as $\hat{J}_{osc}(1)\phi(1)$, where the $\hat{J}_{osc}(1)=\int \mathrm{d}x_2\lambda x_1^4x_2^4\phi^*(2)\phi(2)$ operator (with the interaction operator $\lambda x_1^4x_2^4$ instead of $\frac{1}{r_{12}}$ for electrons) is analogous to $\hat{J}(1)$. There will be no term corresponding to $\hat{K}\,(1)$ since the latter results from the antisymmetry of the variational wave function (the Slater determinant), while the bosonic function is a symmetric product $\phi(1)\phi(2)$ of two "spin orbitals" ϕ and we do not have such an effect. Also, there is no unitary transformation appearing, because we have a single spin orbital here, not a set of N spin orbitals ϕ_i.

8.2.8 "Unrestricted" Hartree–Fock (UHF) method

One may limit the GHF method by forcing some restrictions on the form of the GHF spin orbitals. A special case of the GHF method is known in textbooks as the *unrestricted Hartree–Fock method* (UHF). Despite its name, UHF is not a fully unrestricted method (as the GHF is). In the UHF we assume (cf. Eq. (8.1)):

- orbital components φ_{i1} and φ_{i2} are *real* and
- *there is no mixing of the spin functions* α and β, i.e., either $\varphi_{i1}=0$ and $\varphi_{i2}\neq 0$ or $\varphi_{i1}\neq 0$ and $\varphi_{i2}=0$.

The UHF method is sometimes called the Different Orbitals for Different Spins (DODS) method. In the UHF (or DODS) method each spin orbital has its own orbital energy.

8.2.9 The closed shell systems and the restricted Hartree–Fock (RHF) method

Double occupation of the orbitals and the Pauli exclusion principle

When the number of electrons is even, the spin orbitals are usually formed out of orbitals in a very easy (and simplified with respect to Eq. (8.1)) manner,[21] by multiplication of each orbital by the spin functions α or β:

---

[21] It is not necessary, but quite comfortable.

$$\phi_{2i-1}(\mathbf{r},\sigma)=\varphi_i(\mathbf{r})\alpha(\sigma), \tag{8.29}$$

$$\phi_{2i}(\mathbf{r},\sigma)=\varphi_i(\mathbf{r})\beta(\sigma), \tag{8.30}$$

$$i=1,2,\ldots\frac{N}{2}, \tag{8.31}$$

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*[22]) (see Fig. 8.4). Thus we introduce an artificial *restriction* for spin orbitals (some of the consequences will be described on p. 510). This is why the method is called the restricted Hartree–Fock (RHF) method. Nothing forces us to do so, the criterion is simplicity. The problem of whether doing so we lose the Hartree–Fock energy will be discussed on p. 515.

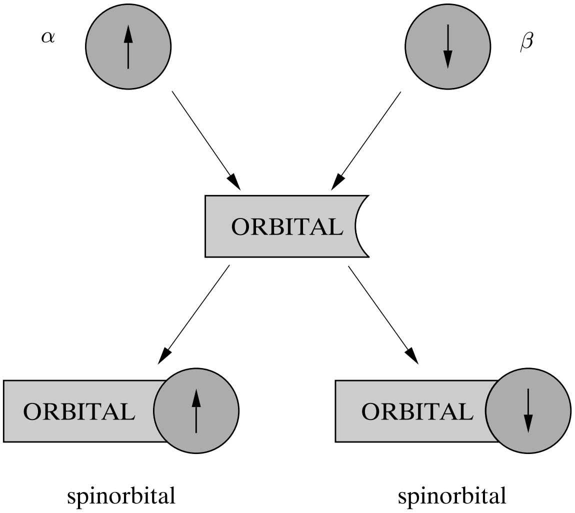


Fig. 8.4. Idea of a spin orbital in the RHF method (i.e., a function x, y, z, σ) as a *product* of an orbital (a function of x, y, z) and one of the two spin functions $\alpha(\sigma)$ or $\beta(\sigma)$.

There are as many spin orbitals as electrons, and there can be a maximum of two electrons per orbital.

---

[22] And only this. When the Slater determinant is written, the electrons loose their identity – they are not anymore distinguishable.

If we wished to occupy a given orbital with more than two electrons, we would need once again to use the spin function α or β when constructing the spin orbitals, i.e., repeating spin orbital. This would imply two identical rows in the Slater determinant, and the wave function would equal zero. This cannot be accepted. The above rule of maximum double occupation is quite often called the Pauli exclusion principle.[23] Such a formulation of the Pauli exclusion principle requires two concepts: the postulate of the antisymmetrization of the electronic wave function (p. 33), and double orbital occupancy. The first of these is of fundamental importance, the second is of a technical nature.[24]

We often assume the double occupancy of orbitals within what is called the *closed shell*. The latter term has an approximate character (Figs. 8.5 and 8.6). It means that for the studied system, there is a large energy difference between HOMO and LUMO orbital energies, where

HOMO is the Highest Occupied Molecular Orbital and LUMO is the Lowest Unoccupied Molecular Orbital. The unoccupied molecular orbitals are called *virtual orbitals*.

A CLOSED SHELL

A closed shell means that the HOMO is doubly occupied as are all the orbitals which are equal or lower in energy. The occupancy is such that the mathematical form of the Slater determinant does not depend on the spatial orientation of the x, y, z axes. Using group theory nomenclature (Appendix C), this function transforms according to the *fully symmetric irreducible representation* of the symmetry group of the electronic Hamiltonian.

---

[23] From "*Solid State and Molecular Theory*," Wiley, London, 1975 by John Slater:
"... I had a seminar about the work which I was doing over there – the only lecture of mine which happened to be in German. It has appeared that not only Heisenberg, Hund, Debye and young Hungarian PhD student Edward Teller were present, but also Wigner, Pauli, Rudolph Peierls and Fritz London, all of them on their way to winter holidays. Pauli, of course, behaved in agreement with the common opinion about him, and disturbed my lecture saying that "*he had not understood a single word out of it*," but Heisenberg has helped me to explain the problem. (...) Pauli was extremely bound to his own way of thinking, similar to Bohr, who did not believe in the existence of photons. Pauli was a warriorlike man, a kind of dictator... ".

[24] The concept of orbitals, occupied by electron pairs, exists only in the mean-field method. We will abandon this idea in the future, and the Pauli exclusion principle will be understood in its generic form as a postulate (see Chapter 1) of the antisymmetry of the electronic wave function.

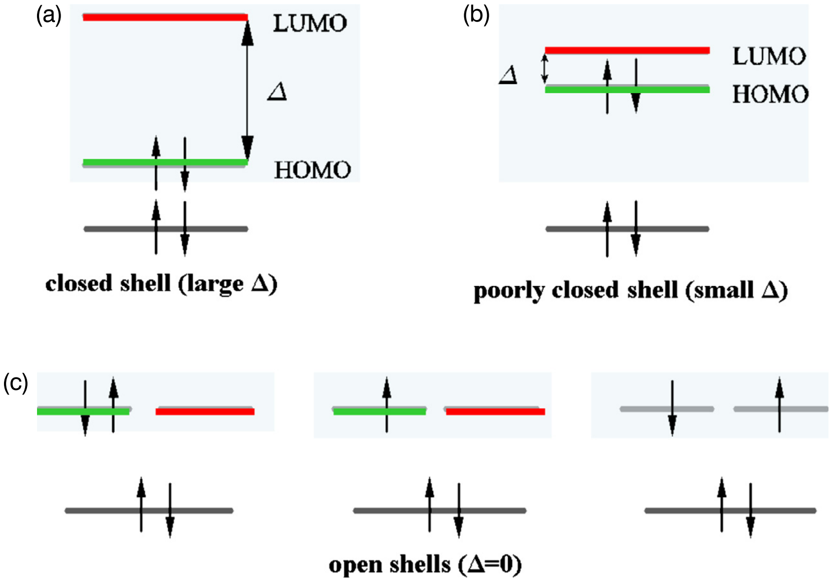


Fig. 8.5. Electron occupancy of orbitals and the concept of the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO). The LUMO-HOMO energy difference is shown as Δ. (a) If Δ is large we have the closed shell. (b) If Δ is small we have to do with a poorly closed shell. (c) $\Delta = 0$ means an open valence shell.

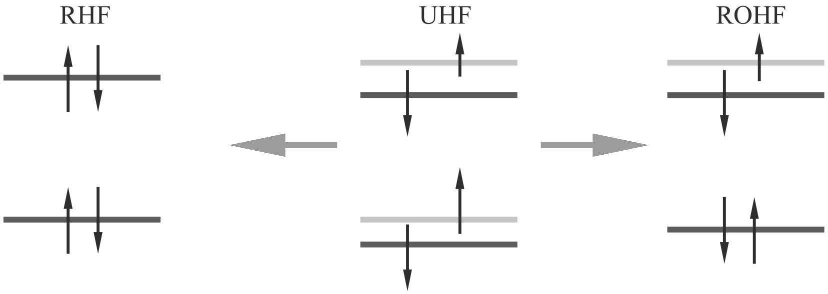


Fig. 8.6. Some restrictions imposed on the GHF method in computational practice: the RHF, UHF, and restricted open shell Hartree–Fock (ROHF) methods. (a) RHF: we force the same (real) orbitals for the electron pair (opposite spins) producing a pair of spin orbitals by using the same orbital. (b) UHF: we relieve this restriction for orbitals (still being real). (c) ROHF: we keep the double occupancy for inner shells as for the RHF method, while for the valence shell we use the UHF-type splitting of orbitals.

If a shell is not closed, it is called “open”.[25] We assume that there is a unique assignment for which molecular spin orbitals[26] within a closed shell are occupied in the ground state. The

---

[25] Sometimes we use the term semiclosed shell, if it is half-occupied by the electrons and we are interested in the state bearing maximum spin. In this case the Slater determinant is a very good approximation. The reason for

concept of the closed shell is approximate because it is not clear what it means when we say that the HOMO-LUMO energy distance[27] is large or small.[28]

We need to note that HOMO and LUMO have somewhat different meanings. As will be shown on p. 544, $-\varepsilon_{HOMO}$ represents an approximate ionization energy, i.e., binding energy of an electron interacting with the $(N-1)$-electron system, while $-\varepsilon_{LUMO}$ is an approximate electron affinity energy, i.e., energy of an electron interacting with the N-electron system.

The Fock equation for closed shells

The Fock equations for a closed shell (RHF method) can be derived in a very similar way as in the GHF method. This means the following steps:

- We write down the expression for the mean value of the Hamiltonian as a functional of the *orbitals*, the summation extends over all the *occupied orbitals* (there are $N/2$ of them, as will be recalled by the upper limit denoted by MO)[29] (see p. 490): $E = 2\sum_i^{\mathrm{MO}} \left(i|\hat{h}|i\right) + \sum_{i,j}^{\mathrm{MO}}[2(ij|ij) - (ij|ji)]$.
- We seek the conditional minimum of this functional (Lagrange multipliers method) allowing for the variation of the *orbitals* which takes their orthonormality into account, i.e., $\delta E = 2\sum_i^{\mathrm{MO}} \left(\delta i|\hat{h}|i\right) + \sum_{i,j}^{\mathrm{MO}}[2(\delta ij|ij) - (\delta ij|ji) + 2(i\delta j|ij) - (i\delta j|ji)] - \sum_{i,j}^{\mathrm{MO}} L'_{ij}(\delta i|j) = 0$.

---

this is, of course, the uniqueness of electron assignment to various spin orbitals. *If there is no uniqueness (as in the carbon atom), then the single-determinant approximation cannot be accepted.*

[26] The adjective "molecular" is suggested even for calculations for an atom. In a correct theory of electronic structure, the number of nuclei present in the system should not play any role. Thus, from the point of view of the computational machinery, an atom is just a molecule with one nucleus.

[27] The decision to occupy only the lowest-energy MOs (so-called *Aufbau Prinzip*; a name left over from the German origins of quantum mechanics) is accepted under the *assumption* that the total energy differences are sufficiently well approximated by the differences in the orbital energies.

[28] Unless the distance is zero. The helium atom, with the two electrons occupying the $1s$ orbital (HOMO), is a $1s^2$ shell of impressive "closure," because the HOMO-LUMO energy difference calculated in a high-quality basis set (6-31G**, see p. 504) of AOs is of the order of 62 eV. On the other hand, the HOMO-LUMO distance is zero for the carbon atom, because in the ground state six electrons occupy the $1s, 2s, 2p_x, 2p_y$, and $2p_z$ orbitals. There is room for 10 electrons, and we only have six. Hence, the occupation (configuration) in the ground state is $1s^2 2s^2 2p^2$. Thus, both HOMO and LUMO are the $2p$ orbital, with zero energy difference. If we asked for a single sentence describing why carbon compounds play a dominant role in living Nature, it should be emphasized that, for carbon atoms, the HOMO-LUMO distance is equal to zero (and that the orbital levels ε_{2s} and ε_{2p} are close in energy).
On the other hand, the beryllium atom is an example of a closed shell, which is not very tightly closed. Four electrons are in the lowest lying configuration $1s^2 2s^2$, but the orbital level $2p$ (LUMO) is relatively close to $2s$ (HOMO) (10 eV for the 6-31G** basis set is not a small gap, yet it amounts much less than that of the helium atom).

[29] And not spin orbitals; see Eqs. (N.18) and (N.19).

- we derive the Euler equation for this problem from $(\delta i|\ldots) = 0$. In fact it is the Fock equation expressed in *orbitals*,[30] i.e.,

$$\hat{\mathcal{F}}(1)\varphi_i(1) = \varepsilon_i \varphi_i(1), \tag{8.32}$$

where φ are the *orbitals*. The *Fock operator* is defined for the closed shell as

$$\hat{\mathcal{F}}(1) = \hat{h}(1) + 2\hat{\mathcal{J}}(1) - \hat{\mathcal{K}}(1), \tag{8.33}$$

where the first term ($\hat{h}$, see Eq. (8.8)) is the sum of the kinetic energy operator of electron 1 and the operator of the interaction of this electron with the nuclei in the molecule, and the next two terms, i.e., Coulombic $\hat{\mathcal{J}}$ 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 (slightly differently than before for $\hat{J}$ and $\hat{K}$ operators[31]) via the action on any function of the position of electron 1:

$$\begin{aligned} 2\hat{\mathcal{J}}(1)\chi(1) &= \sum_{i=1}^{\mathrm{MO}} 2\hat{\mathcal{J}}_i(1)\chi(1) \\ &= \sum_{i=1}^{\mathrm{MO}} 2\int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_i(2)\chi(1) \equiv 2\sum_i^{\mathrm{MO}} \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_i(2)\chi(1), \end{aligned} \tag{8.34}$$

$$\begin{aligned} \hat{\mathcal{K}}(1)\chi(1) &= \sum_{i=1}^{\mathrm{MO}} \hat{\mathcal{K}}_i(1)\chi(1) \\ &= \sum_{i=1}^{\mathrm{MO}} \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\chi(2)\varphi_i(1) \equiv \sum_i^{\mathrm{MO}} \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\chi(2)\varphi_i(1), \end{aligned} \tag{8.35}$$

where integration is now exclusively over the spatial coordinates[32] of electron 2. Multiplying the Coulombic operator by factor 2 results (as the reader presumably guessed) from the *double* occupation of the orbitals.

---

[30] After a suitable unitary transformation of orbitals, analogous to what we have done in the GHF case.

[31] Because we have orbitals here, and not spin orbitals.

[32] Simply, the summation over the spin coordinates has already been done when deriving the equation for the mean value of the Hamiltonian.

Interpretation of the Coulombic operator

The Coulombic operator is nothing else but calculation of the Coulombic potential (with the opposite sign as created by all the *electrons*, Fig. 8.7) at the position of electron 1. Indeed, such a potential coming from an electron occupying MO φ_i is equal to

$$\int \frac{\rho_i(2)}{r_{12}} \mathrm{d}V_2, \tag{8.36}$$

where $\rho_i(2) = \varphi_i(2)^* \varphi_i(2)$ is the probability density of finding electron 2 described by orbital φ_i. If we take into account that the orbital φ_i is occupied by *two* electrons and that the number of the doubly occupied MOs is $N/2$, then the electrostatic potential calculated at the position of electron 1 is

$$\int \frac{\sum_i^{\mathrm{MO}} 2\rho_i(2)}{r_{12}} \mathrm{d}V_2 = \sum_i^{MO} 2\hat{\mathcal{J}}_i = 2\hat{\mathcal{J}}(1).$$

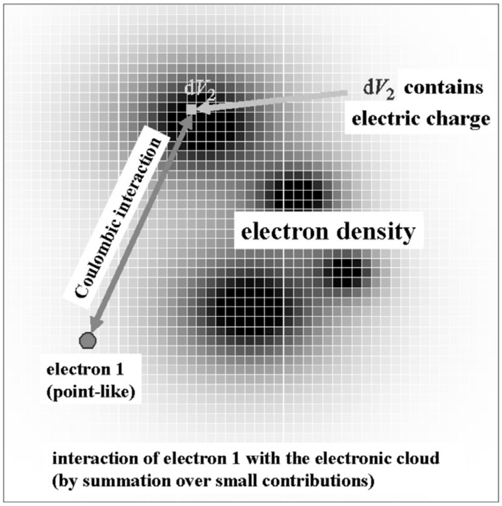


Fig. 8.7. Point-like electron 1 interacts with the total electron density (shown as an electron cloud with density $\rho = \sum_i^{\mathrm{MO}} 2\rho_i(2)$). To compute the interaction energy the total electron density is chopped into small cubes. The interaction energy of electron 1 with one of such cubes of volume $\mathrm{d}V_2$ containing charge $-\sum_i^{\mathrm{MO}} 2\rho_i(2)\mathrm{d}V_2$ is calculated according to the Coulomb law: charge × charge divided by their distance, i.e., $\frac{-1\times(-1)\sum_i^{\mathrm{MO}} 2\rho_i(2)\mathrm{d}V_2}{r_{12}}$, or, alternatively, as charge -1 times electric potential produced by a single cube at electron 1. The summation over all cubes gives $\int \frac{\sum_i^{\mathrm{MO}} 2\rho_i(2)}{r_{12}} \mathrm{d}V_2 = 2\hat{\mathcal{J}}(1)$.

The same expression also means the interaction energy (in a.u.) of two charges (Fig. 8.7): 1 (being the point-like elementary charge -1) and 2, the latter represented by a diffuse cloud of charge density distribution $\rho(2) = \sum_i^{\mathrm{MO}} 2\rho_i(2)$ carrying N elementary charges -1, because $-\int \rho(2) d\mathrm{V}_2 = -2\sum_i 1 = N(-1)$.

Integration in the formula for the operator $\hat{\mathcal{J}}$ is a consequence of the approximation of independent particles.

This approximation means that, in the Hartree–Fock method, the electrons do not move in the electric field of the other point-like electrons, but in the *mean static field of all the electrons* represented by electron cloud $\rho = \sum_i^{\mathrm{MO}} 2\rho_i(2)$; as if a driver (one of the electrons) in Paris did not use the position of other cars, but a map showing only the traffic intensity via the probability density cloud. The driver would then have a diffuse image of other vehicles, and could not satisfactorily optimize his car's position with respect other cars (it means higher energy for the molecule under study).

THE MEAN FIELD

This is typical for all the mean-field methods. In these methods, *instead of watching the motion of other objects in detail*, *we average these motions*, and the problem simplifies (obviously, we pay the price of lower quality).

This trick is general, ingenious, and worth remembering.[33]

Coulombic self-interaction

There is a problem with this. From what we have said, it follows that electron 1 uses the "maps" of *total* electron density, i.e., including its own contribution to the density.[34] This looks strange though. Let us take a closer look, maybe something has been missed in our reasoning. Note first of all that the repulsion of electron 1 (occupying, say, orbital φ_k) with the electrons, which is visible in the Fock operator, reads as $\left(\varphi_k \middle| \left(2\hat{\mathcal{J}} - \hat{\mathcal{K}}\right) \varphi_k\right)$ and not as $\left(\varphi_k \middle| \left(2\hat{\mathcal{J}}\right) \varphi_k\right)$. Let us write it down in more detail. Then we have

$$\left(\varphi_k \middle| \left(2\hat{\mathcal{J}} - \hat{\mathcal{K}}\right) \varphi_k\right) = \int \mathrm{d}V_1 \left|\varphi_k(1)\right|^2 \sum_{i=1}^{\mathrm{MO}} 2 \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_i(2) -$$

---

[33] We use it every day, although we do not call it a mean-field approach. Indeed, if we say: "*I will visit my aunt at noon, because it is easier to travel out of rush hours*," or "*I avoid driving through the center of town, because of the traffic jams*," in practice we are using the mean-field method. We average the motions of all citizens (including ourselves!) and we get a "map" (temporal or spatial), which allows us to optimize *our own* motion. The motion of our fellow citizens *disappears*, and we obtain a *one body* problem.

[34] Exactly as happens with real city traffic maps.

$$\sum_{i=1}^{\mathrm{MO}} \int \mathrm{d}V_1 \varphi_k(1)^* \varphi_i(1) \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_k(2) =$$

$$\int \mathrm{d}V_1 \left|\varphi_k(1)\right|^2 \sum_{i=1}^{\mathrm{MO}} 2 \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_i(2) - \int \mathrm{d}V_1 \varphi_k(1)^* \varphi_k(1) \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_k^*(2)\varphi_k(2) -$$

$$\sum_{i(\neq k)}^{\mathrm{MO}} \int \mathrm{d}V_1 \varphi_k(1)^* \varphi_i(1) \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i^*(2)\varphi_k(2) =$$

$$\int\int \mathrm{d}V_1 \mathrm{d}V_2 \frac{1}{r_{12}} \rho_k(1)[\rho(2) - \rho_k(2)] - \sum_{i(\neq k)}^{\mathrm{MO}} (ki|ik),$$

where $\rho_k = |\varphi_k(1)|^2$, i.e., the distribution of electron 1 interacts electrostatically with all the *other* electrons,[35] i.e., with the distribution $[\rho(2) - \rho_k(2)]$ with ρ denoting the total electron density $\rho = \sum_{i=1}^{\mathrm{MO}} 2|\varphi_i|^2$ and $-\rho_k$ excluding from it the self-interaction energy of the electron in question. Thus, the Coulombic and exchange operators together ensure that an electron interacts electrostatically with other electrons, not with itself.

AN ANALOGY – THE COUPLED HARMONIC OSCILLATORS (continued, 5 of 5):
The Fock operator is $\hat{F}(1) = \hat{h}(1) + \lambda\bar{x}^4 x_1^4$. It is now clear what the mean-field approximation really is: the two particle problem is reduced to a *single* particle one (denoted as number 1), and the influence of the second particle, which appears as a result of the coupling term $\lambda x_1^4 x_2^4$, is *averaged over its positions* $\bar{x}^4 = \langle\phi|x^4\phi\rangle = \langle\phi(2)|x_2^4\phi(2)\rangle$. We see a similar effect in the Hartree–Fock problem for molecules: a single electron denoted by 1 interacts with all other electrons through the operators $2\hat{\mathcal{J}} - \hat{\mathcal{K}}$, in which we have an integration (averaging) over positions of all these electrons. So, we have essentially the same mean-field picture for the electronic system and for the two coupled oscillators.

Here we end the illustration of the Hartree–Fock procedure by the two coupled harmonic oscillators.

Electrons with parallel spins repel less

There is also an exchange remainder $-\sum_{i(\neq k)}^{\mathrm{MO}} (ki|ik)$, which is just a by-product of the antisymmetrization of the wave function (i.e., the Slater determinant), which tells us that in the Hartree–Fock picture electrons of the same spin functions[36] repel less. What? As shown at the

---

[35] The fact that the integration variables pertain to electron 2 is meaningless; it is just a definite integration and the name of the variable does not count at all.

[36] When deriving the total energy expression (Appendix N), only those exchange terms survived which correspond to the parallel spins of the interacting electrons. Note also that for real orbitals (as in the RHF method), every exchange contribution $-(ki|ik) \equiv -\int \mathrm{d}V_1 \varphi_k(1)\varphi_i(1) \int \mathrm{d}V_2 \frac{1}{r_{12}} \varphi_i(2)\varphi_k(2)$ means a repulsion, because this is a self-interaction of the cloud $\varphi_k\varphi_i$.

beginning of the present chapter, two electrons of the same spin cannot occupy the same point in space, and therefore (from the continuity of the wave function) they avoid each other. It is as if they repelled each other, because of the Pauli exclusion principle, *in addition* to their Coulombic repulsion. Is there something wrong in our result then? No, everything is OK. The necessary antisymmetric character of the wave function says simply that the same spins should keep apart. However, when the electrons described by the same spin functions keep apart, this obviously means *their Coulombic repulsion is weaker than that of electrons of opposite spins*. This is what the negative term $-\sum_{i(\neq k)}^{\mathrm{MO}}(ki|ik)$ really means.

Hartree method

The exchange operator represents a (nonintuitive) result of the antisymmetrization postulate for the total wave function (Chapter 1) and it has no classical interpretation. If the variational wave function were the *product* of the spin orbitals[37] (Douglas Hartree did this in the beginning of quantum chemistry),

$$\phi_1(1)\,\phi_2(2)\,\phi_3(3)\ldots\phi_N(N),$$

then we would get the corresponding Euler equation, which in this case is called the Hartree equation, i.e.,

$$\hat{F}_{Hartree}(1)\phi_i(1) = \varepsilon_i \phi_i(1),$$

$$\hat{F}_{Hartree}(1) = \hat{h}(1) + \sum_{j(\neq i)}^{N} \hat{J}_j(1),$$

where $\hat{F}_{Hartree}$ corresponds to the Fock operator. Note that there is no self-interaction there.

8.2.10 Iterative solution: the Self-Consistent Field (SCF) method

The following is a typical technique for solving the Fock equation.

First, we meet the difficulty that in order to solve the Fock equation we should first ...know its solution. Indeed, the Fock equation is not an eigenvalue problem, but a pseudoeigenvalue problem, because the Fock operator depends on the *solutions* (obviously, unknown). So, in the Fock equation we do not know anything: all three quantities $\hat{\mathcal{F}}, \varphi_i, \varepsilon_i$ that constitute the equation are unknown. Regardless of how strange it might seem, we deal with this situation quite easily using an iterative approach, because of the structure of $\hat{\mathcal{F}}$. This is called the SCF method. In this method (Fig. 8.8) we:

---

[37] Such a function is not legal – it does not fulfill the antisymmetrization postulate. This illegal character (because the Pauli exclusion principle is not fulfilled) would sometimes give unfortunate consequences; for example, more than two electrons would occupy the 1s orbital, etc.

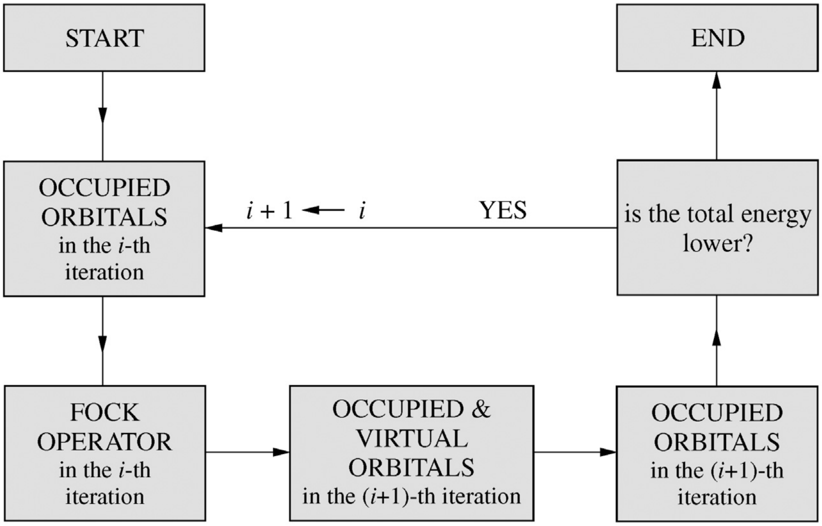


Fig. 8.8. Iterative solution of the Fock equation (the self-consistent field [SCF] method). We:
— start from any set of occupied orbitals (zeroth iteration);
— insert them to the Fock operator;
— solve the Fock equation;
— obtain the MOs of the first approximation;
— choose those of the lowest energy as the occupied ones; and
— if your criterion of the total energy is not satisfied, repeat the procedure.

- assume at the beginning (zeroth iteration) a certain shape of MOs[38];
- introduce these orbitals to the Fock operator, thus obtaining a sort of "caricature" of it (the zero-order Fock operator);
- solve the eigenvalue problem using the above "Fock operator" and get the MOs of the first iteration;
- repeat the process until the shape of the orbitals does not change in the next iteration, i.e., until the Fock equations are solved.[39]

---

[38] These are usually the any-sort "orbitals," although recently, because of the *direct SCF* idea (we calculate the integrals whenever they are needed, i.e., at each iteration), an effort is made to save computational time per iteration and therefore to provide as high-quality a starting function as possible. We may obtain it via an initial calculation with some two-electron integrals neglected.

[39] Using our driver analogy, we may say that at the beginning the driver has *false* maps of the probability density (thus the system energy is high, in our analogy the car repair costs are large). The next iterations (repair costs effectively teach all the drivers) improve the map, the final energy decreases, and at the very end we get the best map possible. The mean energy is the lowest possible (within the mean-field method). A further energy lowering is only possible beyond the Hartree–Fock approximation, i.e., outside of the mean-field method, which for the drivers means not believing maps, but their own eyes. A suspicious person (scientist) should be careful, because our solution *may depend on the starting point used*, i.e., from the initial, completely arbitrary orbitals.

8.3 Total energy in the Hartree–Fock method

In Appendix N (p. 707) we derived the following expressions for the mean value of the Hamiltonian using the normalized determinant (without a constant additive term for the nuclear repulsion energy V_{nn}; SMO means summation over the spin orbitals $i = 1, ..., N$; in the RHF method, the MO summation limit means summation over the orbitals $i = 1, ..., N/2$)

$$E'_{HF} = \sum_{i}^{\text{SMO}} \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{\text{SMO}} [\langle ij|ij\rangle - \langle ij|ji\rangle] \equiv \sum_{i}^{\text{SMO}} h_{ii} + \frac{1}{2}\sum_{i,j=1}^{\text{SMO}} [J_{ij} - K_{ij}]. \tag{8.37}$$

If double occupancy is assumed (i.e., the flexibility of the variational wave function is *restricted*) we may transform this expression in the following way:

$$E'_{RHF}(\text{double occupancy}) = \sum_{i}^{\text{MO}} \Big(\langle i\alpha|\hat{h}|i\alpha\rangle + \langle i\beta|\hat{h}|i\beta\rangle\Big) +$$

$$\frac{1}{2}\sum_{i}^{\text{MO}}\sum_{j}^{\text{SMO}} [\langle i\alpha, j|i\alpha, j\rangle - \langle i\alpha, j|j, i\alpha\rangle + \langle i\beta, j|i\beta, j\rangle - \langle i\beta, j|ji\beta\rangle] =$$

$$2\sum_{i}^{\text{MO}} (i|\hat{h}|i) + \frac{1}{2}\sum_{i}^{\text{MO}}\sum_{j}^{\text{MO}} [\langle i\alpha, j\alpha|i\alpha, j\alpha\rangle + \langle i\alpha, j\beta|i\alpha, j\beta\rangle - \langle i\alpha, j\alpha|j\alpha, i\alpha\rangle -$$

$$\langle i\alpha, j\beta|j\beta, i\alpha\rangle + \langle i\beta, j\alpha|i\beta, j\alpha\rangle + \langle i\beta, j\beta|i\beta, j\beta\rangle - \langle i\beta, j\alpha|j\alpha, i\beta\rangle - \langle i\beta, j\beta|j\beta, i\beta\rangle] =$$

$$2\sum_{i}^{\text{MO}} (i|\hat{h}|i) + \frac{1}{2}\sum_{i}^{\text{MO}}\sum_{j}^{\text{MO}} [4(ij|ij) - 2(ij|ji)] = 2\sum_{i}^{\text{MO}} (i|\hat{h}|i) + \sum_{i}^{\text{MO}}\sum_{j}^{\text{MO}} [2(ij|ij) - (ij|ji)].$$

This finally gives

$$E'_{RHF} = 2\sum_{i}^{\text{MO}} h_{ii} + \sum_{i,j}^{\text{MO}} [2\mathcal{J}_{ij} - \mathcal{K}_{ij}], \tag{8.38}$$

where $\mathcal{J}_{ij} \equiv (ij|ij)$, $\mathcal{K}_{ij} \equiv (ij|ji)$ (see p. 469).

Both Eqs. (8.37) and (8.38) may in general give different results, because in the first, no double occupancy is assumed (we will discuss this further on p. 514).

Besides, the iteration process does not necessarily need to be convergent. But it appears that the solutions in the Hartree–Fock method are usually independent of the zero-order MOs, and convergence problems are very rare. This is surprising. This situation is much worse for higher-quality computations, where the AOs of small exponents are included (*diffuse orbitals*). Then we truly meet the problem already described (p. 413) of searching for the global energy minimum among a multitude of local ones.

Given the equality $\langle i|\hat{h}|i\rangle = (i|\hat{h}|i)$, these integrals have been written here as h_{ii}. The *Coulombic and exchange* integrals involving spin orbitals are denoted J_{ij} and K_{ij} and involving orbitals as $\mathcal{J}_{ij}$ and $\mathcal{K}_{ij}$.

The additive constant corresponding to the internuclear repulsion (it is constant, since the nuclei positions are frozen),

$$V_{nn} = \sum_{a<b} \frac{Z_a Z_b}{R_{ab}}, \tag{8.39}$$

has not been introduced in the electronic Hamiltonian $\hat{H}_0$ and thus the full Hartree–Fock energy is

$$E_{RHF} = E'_{RHF} + V_{nn}. \tag{8.40}$$

Note that the mean value of the electronic repulsion energy in our system[40]

$$V_{ee} = \left\langle \psi_{HF} | \sum_{i<j} \frac{1}{r_{ij}} | \psi_{HF} \right\rangle = \sum_{i,j}^{\mathrm{MO}} [2(ij|ij) - (ij|ji)] = \sum_{i,j}^{\mathrm{MO}} [2\mathcal{J}_{ij} - \mathcal{K}_{ij}]. \tag{8.41}$$

It is desirable (interpretation purposes) to include the orbital energies in the formulae derived. Let us recall that the orbital energy ε_i is the mean value of the Fock operator for orbital i, i.e., the energy of an effective electron described by this orbital.[41] Based on Eqs. (8.33)–(8.35), this can be expressed as (i stands for the MO)

$$\varepsilon_i = h_{ii} + \sum_{j}^{\mathrm{MO}} [2\mathcal{J}_{ij} - \mathcal{K}_{ij}], \tag{8.42}$$

and this in turn gives an elegant expression for the Hartree–Fock electronic energy, i.e.,

$$E'_{RHF} = \sum_{i}^{\mathrm{MO}} [h_{ii} + \varepsilon_i]. \tag{8.43}$$

From Eqs. (8.38), (8.42), and (8.41), the total electronic energy may be expressed as

$$E'_{RHF} = \sum_{i=1}^{MO} 2\varepsilon_i - V_{ee}. \tag{8.44}$$

---

[40] Please recall that $\langle \psi_{RHF} | \hat{H} | \psi_{RHF} \rangle = E_{RHF}$ and V_{ee} is, therefore, the Coulombic interaction of electrons.

[41] $\left(\varphi_i | \hat{\mathcal{F}} \varphi_i \right) = \varepsilon_i$, because $\hat{\mathcal{F}} \varphi_i = \varepsilon_i \varphi_i$.

It can be seen that the total electronic energy (i.e., E'_{RHF}) *is not the sum of the orbital energies of electrons* $\sum_i 2\varepsilon_i$.

And we would already expect full additivity, since the electrons in the Hartree–Fock method are treated as independent. Yet "independent" does not mean "noninteracting." The reason for the nonadditivity is that for *each* electron we need to calculate its effective interaction with *all* the electrons, hence we would get too much repulsion.[42] Of course, the total energy, and not the sum of the orbital energies, is the most valuable. Yet in many quantum chemical problems we interpret orbital energy lowering as energetically profitable. And it turns out that such an interpretation has "practice motivated" justification. Works by Fraga, Politzer, and Ruedenberg[43] show that at the equilibrium geometry of a molecule, the formula

$$E_{RHF} = E'_{RHF} + V_{nn} \approx \frac{3}{2} \sum_{i=1}^{\mathrm{MO}} 2\varepsilon_i \tag{8.45}$$

works with 2%–4% precision, and even better results may be obtained by taking a factor of 1.55 instead of $\frac{3}{2}$.

8.4 Computational technique: atomic orbitals as building blocks of the molecular wave function

One of most powerful methods of computational analysis is to represent a function to be sought as a linear combination of some predefined set of functions (basis set). Fig. 8.9 shows the efficiency of such an idea in a step-by-step way (in case of a smooth function of a single variable).

ATOMIC ORBITAL (AO)

An "atomic orbital (AO)" means a function

$$g(\mathbf{r}) = f(x, y, z) \exp(-\zeta r^n),$$

where $f(x, y, z)$ is a polynomial and $n = 1, 2$. Such an AO is localized (centered) around $(0, 0, 0)$. The larger the exponent $\zeta > 0$, the more effective is this centering.

For $n = 1$, we have what is called *Slater-type orbitals (STOs)*, and for $n = 2$ we have *Gaussian-type orbitals (GTOs)*.

---

[42] For example, the interaction of electron 5 and electron 7 is calculated *twice*: as the interaction $\frac{1}{r_{57}}$ and as $\frac{1}{r_{75}}$.

[43] S. Fraga, *Theor. Chim. Acta*, 2(1964)406, P. Politzer, *J. Chem. Phys.*, 64(1976)4239, K. Ruedenberg, *J. Chem. Phys.*, 66(1977)375.

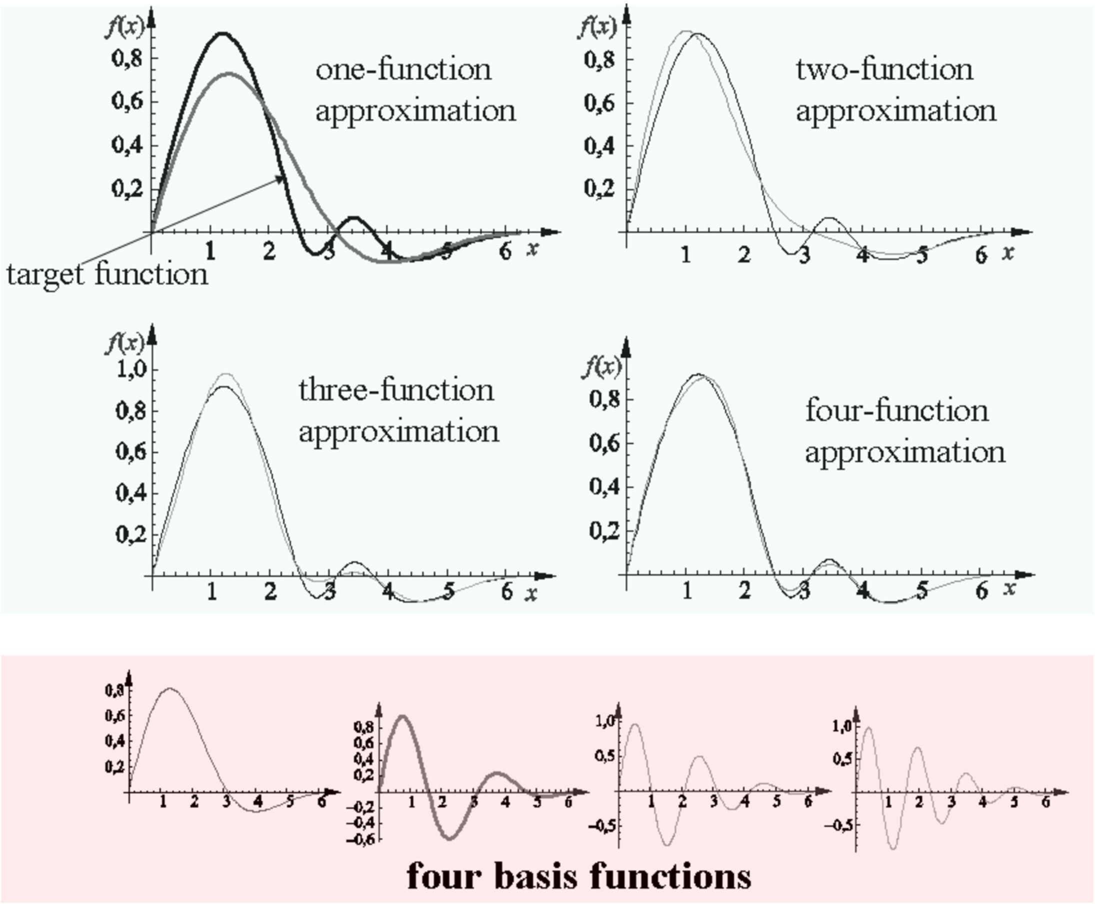


Fig. 8.9. An approximation of a function as a linear combination of some basis functions (one variable case, Mathematica software was used). The first (larger) figures show how a function (thick solid line) is approximated by, consecutively: one, two, three, and four basis functions (shown individually as small inserts at the bottom). As one can see, the approximation gets better and better with an increasing number of expansion functions. In an ideal situation (never reached in practice) this number would be infinite.

We have to be careful, because the term “atomic orbital” is used in quantum chemistry with a double meaning. These are: (i) the Hartree–Fock orbitals for a particular atom, or (ii) functions localized in the space about a given center.[44] The role of the AOs is to provide a complete set of functions, i.e., a suitable LCAO should in principle be able to approximate any continuous and square integrable function.

---

[44] AOs (the first meaning) may be thought of as expressed through LCAOs in the second meaning. The AOs may be centered on the nuclei (common practice), but can also be centered off the nuclei.

8.4.1 Centering of the atomic orbital

If a complete set of the orbitals were at our disposal, then all the AOs might be centered in a *single* point.

> It is more economic, however, to allow using the incomplete set and the possibility of the AOs being centered in various points of space.

An AO $g(\mathbf{r})$ may be shifted by a vector $\mathbf{A}$ in space (translation operation $\hat{\mathcal{U}}(\mathbf{A})$, see Chapter 2) to result in the new function $\hat{\mathcal{U}}(\mathbf{A})g(\mathbf{r}) = g(\hat{T}^{-1}(\mathbf{A})\,\mathbf{r}) = g(\hat{T}(-\mathbf{A})\,\mathbf{r}) = g(\mathbf{r}-\mathbf{A})$, because $\hat{T}^{-1}(\mathbf{A}) = \hat{T}(-\mathbf{A})$. Hence the orbital centered at a given point (indicated by a vector $\mathbf{A}$) is (Fig. 8.10)

$$g(\mathbf{r}-\mathbf{A}) = f(x-A_x, y-A_y, z-A_z)\exp[-\zeta\,|\mathbf{r}-\mathbf{A}|^n]. \tag{8.46}$$

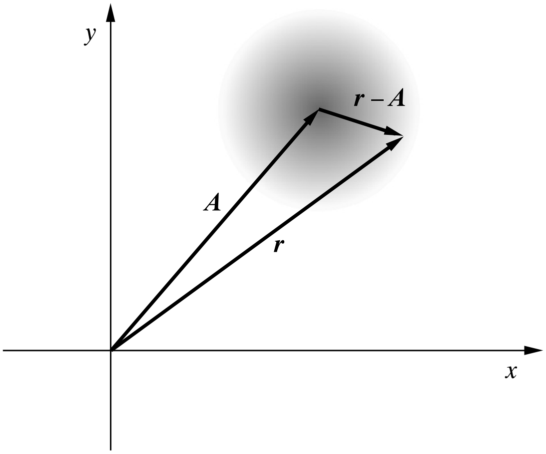


Fig. 8.10. The AO $g(\mathbf{r})$ centered at the point shown by vector $\mathbf{A}$ means creation of the orbital $g(\mathbf{r}-\mathbf{A})$. A linear combination of such orbitals can describe any smooth function of the position in space, of any degree of complexity.

8.4.2 Slater-type orbitals (STOs)

The STOs[45]

$$\chi_{STO,nlm}(r,\theta,\phi) = N r^{n-1}\exp(-\zeta r)\, Y_l^m(\theta,\phi) \tag{8.47}$$

---

[45] We will distinguish two similar terms here: STOs and Slater orbitals. The latter is reserved for special STOs, in which the exponent is easily computed by considering the effect of the screening of nucleus by the internal electronic shells. The screening coefficient is calculated according to the Slater rules (see p. 528).

(N stands for a suitable normalization constant) differ from the atomic orbitals of the hydrogen atom (see p. 232). The first difference is that the radial part is simplified in the STOs (the radial part of an STO has no nodes) (cf. p. 232). The second difference is in the orbital exponent, which has no constraint except[46] $\zeta > 0$.

The STOs have a great advantage: they decay with distance from the center in a similar way to the "true" orbitals – let us recall the exponential vanishing of the hydrogen atom orbitals (see Chapter 4).[47] STOs would be fine, but finally we have to compute a large number of the integrals needed.[48] And here is a real problem. Since the Hamiltonian contains the electron–electron interactions, integrals appear with, in general, four AOs (of different centers). These integrals are difficult to calculate, and are therefore excessively computer time consuming.

8.4.3 Gaussian-type orbitals (GTOs)

If the exponent in Eq. (8.46) is equal to $n = 2$, we are dealing with GTOs (N' is a normalization constant), i.e.,

$$\chi_{GTO,nlm}\left(r,\theta,\phi\right) = N' r^{n-1} \exp\left(-\zeta r^2\right) Y_l^m\left(\theta,\phi\right). \tag{8.48}$$

The most important among them are $1s$-type orbitals, given for an arbitrary center $\mathbf{R}_p$:

$$\chi_p \equiv G_p(\mathbf{r};\alpha_p,\mathbf{R}_p) = \left(\frac{2\alpha_p}{\pi}\right)^{\frac{3}{4}} \exp(-\alpha_p|\mathbf{r}-\mathbf{R}_p|^2), \tag{8.49}$$

where α_p is the orbital exponent. Why are $1s$-type orbitals so important? Because we may construct "everything" (even s, p, d-like orbitals) out of them using proper linear combinations. For example, the difference of two $1s$ orbitals, centered at $(a,0,0)$ and $(-a,0,0)$, is similar to the $2p_x$ orbital (Fig. 8.11).

---

[46] Otherwise the orbital would not be square integrable.

[47] It has been proved that *each* of the Hartree–Fock orbitals has *the same* asymptotic dependence on the distance from the molecule (N.C. Handy, M.T. Marron, H.J. Silverstone, *Phys. Rev.*, 180(1969)45), i.e., $const \cdot \exp(-\sqrt{-2\varepsilon_{\max}}r)$, where $\varepsilon_{\max}$ is the orbital energy of HOMO. Earlier, people thought the orbitals decay as $\exp(-\sqrt{-2\varepsilon_i}r)$, where ε_i is the orbital energy expressed in a.u. The last formula, as is easy to prove, holds for the AOs of hydrogen atoms (see p. 232). R. Ahrichs, M. Hoffmann-Ostenhoff, T. Hoffmann-Ostenhoff, J.D. Morgan III, *Phys. Rev. A*, 23(1981)2106 have shown that at a long distance r from an atom or a molecule, the square root of the *ideal* electron density satisfies the inequality $\sqrt{\rho} \le C\left(1+r\right)^{\frac{(Z-N+1)}{\sqrt{2\varepsilon}}-1} \exp\left[-\left(2\varepsilon r\right)\right]$, where ε is the first *ionization potential*, Z is the sum of the nuclear charges, N is the number of electrons, and $C > 0$ is a constant.

[48] The number of necessary integrals may reach billions.

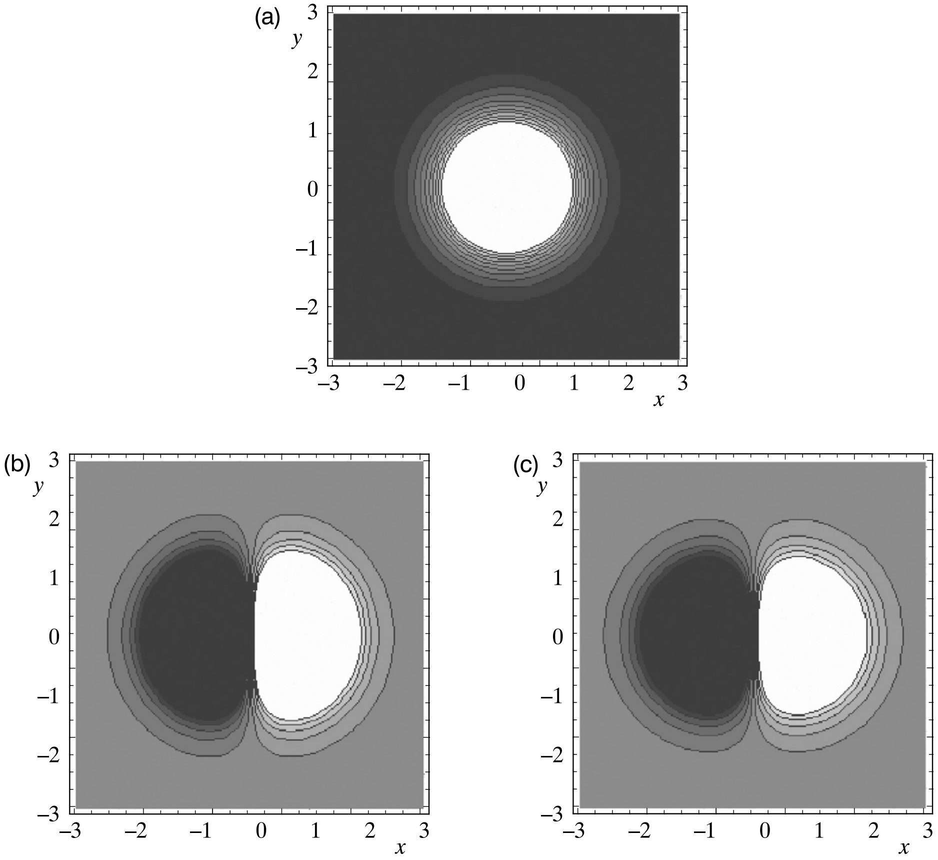


Fig. 8.11. Two spherically symmetric GTOs of the “$1s$ type” $G(\mathbf{r}; 1, \mathbf{0})$ (a) are used to form the difference orbital (b): $G(\mathbf{r}; 1, -0.5\mathbf{i}) - G(\mathbf{r}; 1, +0.5\mathbf{i})$, where $\mathbf{i}$ is the unit vector along the x axis. For comparison (c) the Gaussian-type p_x orbital is shown: $xG(\mathbf{r}; 1, \mathbf{0})$. It can be seen that the spherical orbitals may indeed simulate the $2p$ ones. Similarly, they can model the spatial functions of arbitrary complexity.

The most important reason for the great progress of quantum chemistry in recent years is replacing the formerly used STOs by GTOs as the expansion functions.

Orbital size

Each orbital extends to infinity and it is impossible to measure its extension using a ruler. Still, the α_p coefficient may allow comparison of the sizes of various orbitals. And the quantity

$$\rho_p = (\alpha_p)^{-\frac{1}{2}} \tag{8.50}$$

may be called (which is certainly an exaggeration) the *orbital radius* of the orbital χ_p, because[49]

$$\int_0^{\rho_p}\int_0^{\pi}\int_0^{2\pi} \chi_p^2 dV = 4\pi \int_0^{\rho_p} \chi_p^2 r^2 dr = 0.74, \tag{8.51}$$

where the integration over r goes through the inside of a sphere of radius ρ_p. This gives us an idea about the part of space in which the orbital has an important amplitude. For example, the $1s$ hydrogen atom orbital can be approximated as a linear combination of three $1s$ GTOs (here centered on the origin of the coordinate system; such a popular approximation is abbreviated to STO-3G)[50]:

$$1s \approx 0.64767\, G_1(\mathbf{r}; 0.151374, \mathbf{0}) + 0.40789\, G_2(\mathbf{r}; 0.681277, \mathbf{0}) + 0.07048\, G_3(\mathbf{r}; 4.50038, \mathbf{0}), \tag{8.52}$$

which corresponds to the following radii ρ of the three GTOs: 2.57, 1.21, and 0.47 a.u.

Product of GTOs

The GTOs have an outstanding feature (along with the square dependence in the exponent), which decides about their importance in quantum chemistry.

> The product of two Gaussian-type $1s$ orbitals (even if they have different centers) is a single Gaussian-type $1s$ orbital.

The case of GTOs other than $1s$ does not give any trouble,[51] but the result is slightly different. The polynomials of x, y, z standing in both GTOs multiplied by each other (recall the

---

[49] See, e.g., I.S. Gradshteyn, I.M. Ryzhik, "*Table of Integrals, Series, and Products*," Academic Press, Orlando, 1980, formula (3.381).

[50] S. Huzinaga, *J. Chem. Phys.*, 42(1965)1293.

[51] Let us take two (not normalized) GTOs $1s$: $\exp(-a(\mathbf{r}-\mathbf{A})^2)$ and $\exp(-b(\mathbf{r}-\mathbf{B})^2)$, the first centered on the point shown by vector $\mathbf{A}$, the second on that by vector $\mathbf{B}$. It will be shown that their product is the GTO

$$\exp(-a(\mathbf{r}-\mathbf{A})^2)\exp(-b(\mathbf{r}-\mathbf{B})^2) = N\exp(-c(\mathbf{r}-\mathbf{C})^2),$$

$$\text{with parameters}\qquad c = a+b\,,$$

$$\mathbf{C} = (a\mathbf{A}+b\mathbf{B})/(a+b),$$

$$N = \exp[-\frac{ab}{a+b}(\mathbf{A}-\mathbf{B})^2].$$

dependence of the polynomial on the orbital centering, Eq. (8.46)) can always be presented as a certain polynomial of x', y', z' taken versus the new center **C**. Hence, in the general case,

the product of any two GTOs is a linear combination of GTOs.

Vector **C** shows the center of the new GTO. It is identical to the center of mass position, where the role of mass is played by the orbital exponents a and b.
Here is the proof:

$$\begin{aligned}&\underline{\text{Left side}} =\\ &\quad= \exp(-ar^2 + 2a\mathbf{r}\mathbf{A} - aA^2 - br^2 + 2b\mathbf{r}\mathbf{B} - bB^2)\\ &\quad= \exp(-(a+b)r^2 + 2\mathbf{r}(a\mathbf{A} + b\mathbf{B}))\exp[-(aA^2 + bB^2)]\\ &\quad= \exp(-cr^2 + 2c\mathbf{C}\mathbf{r})\exp[-(aA^2 + bB^2)].\end{aligned}$$

$$\begin{aligned}&\underline{\text{Right side}} =\\ &\quad= N\exp(-c(\mathbf{r} - \mathbf{C})^2) = N\exp[-c(r^2 - 2\mathbf{C}\mathbf{r} + C^2)]\\ &\quad= \text{left side},\end{aligned}$$

if $N = \exp(cC^2 - aA^2 - bB^2)$.
It is instructive to transform the expression for N, which is a kind of amplitude of the GTO originating from the multiplication of two GTOs. So,

$$\begin{aligned}N &= \exp[(a+b)C^2 - aA^2 - bB^2] = \exp\left(\frac{(a^2A^2 + b^2B^2 + 2ab\mathbf{A}\mathbf{B})}{(a+b)} - aA^2 - bB^2\right)\\ &= \exp\left(\frac{1}{a+b}(a^2A^2 + b^2B^2 + 2ab\mathbf{A}\mathbf{B} - a^2A^2 - abA^2 - b^2B^2 - abB^2)\right)\\ &= \exp\left(\frac{1}{a+b}(2ab\mathbf{A}\mathbf{B} - abA^2 - abB^2)\right)\\ &= \exp\left(\frac{ab}{a+b}(2\mathbf{A}\mathbf{B} - A^2 - B^2)\right) = \exp\left(\frac{-ab}{a+b}(\mathbf{A} - \mathbf{B})^2\right).\end{aligned}$$

This is what we wanted to show.
It is seen that if $\mathbf{A} = \mathbf{B}$, then amplitude N is equal to 1 and the GTO with the $a + b$ exponent results (as it should). The amplitude N strongly depends on the distance $|\mathbf{A} - \mathbf{B}|$ between two centers. If the distance is large, N is very small, which gives the product of two distant GTOs as practically zero (in agreement with common sense). It is also clear that if we multiply two strongly contracted GTOs ($a, b \gg 1$) of different centers, the "GTO product" is again small. Indeed, let us take, e.g., $a = b$. We get $N = \exp\{[-a/2][\mathbf{A} - \mathbf{B}]^2\}$.

Integrals

If somebody wanted to perform alone[52] quantum chemical calculations, they would immediately face integrals to compute, the simplest among them being the $1s$-type. Expressions for these integrals are given in Appendix Q on p. 729.

8.4.4 Linear combination of atomic orbitals (LCAO) method

Algebraic approximation

Usually we apply the SCF approach with the LCAO method; this is then the SCF LCAO MO.[53] In the SCF LCAO MO method, each MO is presented as an LCAO χ_s,

$$\boxed{\varphi_i(1) = \sum_s^M c_{si}\chi_s(1), \qquad (8.53)}$$

where the symbol (1) emphasizes that each of the AOs and the resulting MO depend on the spatial coordinates of *one* electron only (say, electron 1). The coefficients c_{si} are called the LCAO coefficients. The STOs and GTOs are important only in the context of the LCAO.

The approximation in which the MOs are expressed as linear combinations of the AOs is also called the *algebraic approximation*.[54]

---

[52] That is, independent of existing commercial programs, which only require the knowledge of how to push a few buttons.

[53] This English abbreviation turned out to be helpful for Polish quantum chemists in communist times (as "specialists in MO methods," MO standing for the mighty "citizen police," which included the secret police). It was independently used by Professors Wiktor Kemula (University of Warsaw) and Kazimierz Gumiński (Jagiellonian University). A young coworker of Professor Gumiński complained that despite much effort he still could not get the official registered address in Cracow, required for employment at the university. The Professor wrote a letter to the officials, and asked his coworker to deliver it in person. The reaction was immediate: "*Why didn't you show this to us earlier?!*"

[54] It was introduced in solid state theory by Felix Bloch (his biography can be found on p. V2-10) and used in chemistry for the first time by Hückel.

Erich Hückel (1896–1980), German physicist, professor at the universities in Stuttgart and Marburg, student of Bohr and Debye. Erich Hückel, presumably inspired by his brother Walter, an eminent organic chemist, created a simplified version of the Hartree–Fock method, which played a major role in linking quantum theory with chemistry. Even today, although this tool is extremely simplistic and has been superceded by numerous and much better computational schemes, Hückel theory is valued as an initial insight into the electronic structure of some categories of molecules and solids.

Curiosity: these people liked to amuse themselves with little rhymes. Felix Bloch has translated a poem by Walter Hückel from German to English. It does not look like great poetry, but deals with the famous Erwin (Schrödinger) and his mysterious function ψ:

"*Erwin with his ψ can do*
Calculations quite a few.
But one thing has not been seen,
Just what does ψ really mean."

Why is it so useful? Imagine we do not have such a tool at our disposal. Then we are confronted with defining a function that depends on the position in three-dimensional space and has a quite complex shape. If we want to do it accurately, we should provide the function values at many points in space, say, for a large lattice with a huge number of nodes, and the memory of our PC will not stand it. Also, in such an approach one would not make use of the fact that the function is smooth. We find our way through by using AOs. For example, even if we wrote that an MO is in fact *a single* AO (we can determine the latter by giving *only four* numbers: three coordinates of the orbital center and the orbital exponent), although very primitive, this would carry a lot of physical intuition (truth): (i) the spatial distribution of the probability of finding the electron is concentrated in some small region of space, (ii) the function decays exponentially when we go away from this region, etc.

"Blocks" of MOs φ_i are constructed out of "primary building blocks" – the one-electron functions χ_r (in jargon called *atomic orbitals*), which are required to fill two basic conditions, i.e.,

- they need to be square integrable and
- they need to form the complete set, i.e., "everything" can be constructed from this set (any smooth square integrable function of x, y, z),

and several *practical* conditions:

- they should be *effective*, i.e., each single function should include a part of the physics of the problem (position in space, decay rate while going to ∞, etc.),

- they should be "flexible," i.e., their parameters should influence their shape to a large extent,
- the resulting *integrals should be easily computable* (numerically and/or analytically) (see p. 499).

In *computational practice*, unfortunately, we fulfill the second set of conditions only to some extent: the set of orbitals taken into calculations (i.e., *the basis set*) is always limited, because computing time requires money, etc. In some calculations for crystals, we also remove the first set of conditions (we often use *plane waves*, $\exp(i\mathbf{k}\cdot\mathbf{r})$, and these are not square integrable).

We could construct an MO of any complexity *exclusively* using a linear combination of the orbitals $g(\mathbf{r}) = \exp(-\zeta\,|\mathbf{r}-\mathbf{A}|^n)$ with different ζ and $\mathbf{A}$, i.e., the $f(x,y,z) = const$, known as the $1s$ orbitals. We could do it also even for a Beethoven bust in a kind of "hole repairing" (plastering-like) procedure by building Beethoven from a mist of very many $1s$ orbitals similarly as a sculptor makes it from clay.[55] But why do we not do it like this in practice? The reason is simple: the number of such AOs that we would have to include in the calculations would be too large. Instead, chemists allow for higher-order polynomials $f(x,y,z)$. This makes for more efficient "plastering," because, instead of spherically symmetric objects ($1s$), we can use orbitals $g(\mathbf{r})$ of virtually any shape (via an admixture of the $p, d, f, \ldots$ functions that introduce more and more complicated patterns of wave functions increasing or decreasing). For example, how a rugby ball-shaped orbital can be achieved is shown in Fig. 8.12.

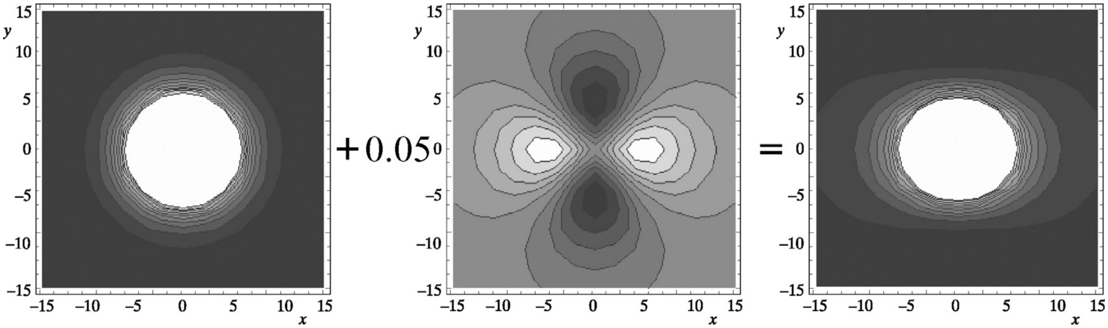


Fig. 8.12. An example of function modeling by a linear combination of AOs. If a tiny admixture of the $3d_{x^2-y^2}$ function is added to the spherically symmetric $1s$ orbital (a football, both with 0.5 orbital exponent). We will get shrinking in one direction, and elongation in the other (the dimension in the third direction is unchanged), i.e., a flattened rugby ball. In our case the tiny admixture means 0.05. If the admixture were of the $2p$ type, the ball would look more like an egg. As we see, nearly everything can be simulated like this. This is the essence of the LCAO method.

---

[55] Frost even derived the method of *Floating Spherical Gaussian Orbitals (FSGO),* A.A. Frost, *J. Chem. Phys.,* 47(1967)3707), i.e., GTOs of variationally chosen positions. Their number is truly minimal – equal to the number of occupied MOs.

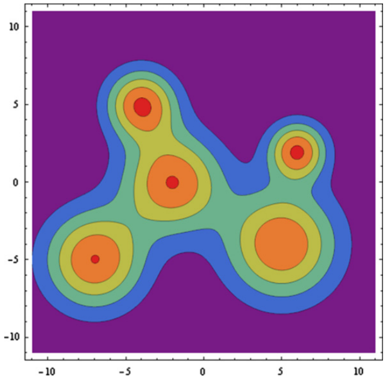


Fig. 8.13. The concept of an MO as an LCAO, a section view. From the point of view of mathematics, it is an expansion in a series of a complete set of functions. From the viewpoint of physics, it is just recognizing that when an electron is close to nucleus a, it should behave in a similar way as that required by the AO of atom a. From the point of view of a bricklayer, it represents the construction of a large building from soft and mutually interpenetrating bricks.

If in Fig. 8.13 we take the linear combination of five AOs and provide a reasonable choice of their centers, the exponents, and the weights of the functions, we will get quite a good approximation of the ideal orbital. We account for the advantages as follows: instead of providing a huge number of function values at the grid nodes, we master the function using only $5 \times 5 = 25$ numbers.[56]

Interpretation of LCAO

The idea of LCAO MO is motivated by the fact that the MO should consist of spatial sections (AOs), because in a molecule in the vicinity of a given atom, an electron should be described by an AO of *this* atom. The essence of the LCAO approach is just the connection (unification) of such sections. But only some AOs are important in practice. This means that the main effort of constructing MOs is connected to precise shaping and polishing, by the inclusion of more and more of the necessary AOs.[57]

Effectiveness of AO mixing

When could we expect that two normalized AOs will have comparable LCAO coefficients in a low-energy MO? Two rules hold (both can be deduced immediately from Eq. (D.1)) for the

---

[56] Three coordinates of the center, the exponent, and the coefficient standing at AO altogether give five parameters per AO.

[57] Which plays the role of the filling mass, because we aim for a beautiful (i.e., ideal from the point of view of the variational method) shape for the MOs.

mixing effectiveness of the AOs, obtained from numerical experience:

EFFECTIVENESS OF AO MIXING
– AOs must correspond to comparable energies (in the meaning of the mean value of the Fock operator);
– AOs must have large overlap integral.

Let us see what we obtain as the orbital energies[58] (in a.u.) for several important atoms:

	1s	**2s**	**2p**	**3s**	**3p**
H	**−0.5**	–	–	–	–
C	−11.34	**−0.71**	**−0.41**	–	–
N	−15.67	−0.96	**−0.51**	–	–
O	−20.68	−1.25	**−0.62**	–	–
F	−26.38	−1.57	**−0.73**	–	–
Cl	−104.88	−10.61	−8.07	−1.07	**−0.51**

Now, which orbitals will mix effectively when forming methane? The hydrogen atom offers the $1s$ orbital with energy -0.5. As we can see from the table, there is no possibility of effectively mixing the carbon $1s$ orbital, while the $2s$ and $2p$ are very good candidates. Note that

the orbital energies of all the outermost (the so-called valence) orbitals are similar for all the elements (highlighted as bold in the table), and therefore they are able to mix effectively, i.e., to lower energy by forming chemical bonds. This is why chemistry is mainly the science of outer shell orbitals.

The mathematical meaning of LCAO

From a mathematical point of view, Eq. (8.53) represents an expansion of an unknown function φ_i in a series of the known functions χ_r, which belong to a certain complete set, thus M should be equal ∞. In real life, we need to truncate this series, i.e., use some limited M.

---

[58] J.B. Mann, "*Atomic Structure Calculations. I. Hartree-Fock Energy Results for the Elements H through Lr*," Report LA-3690 (Los Alamos National Laboratory, 1967).

8.4.5 Basis sets of atomic orbitals

BASIS SET

The set of the AOs $\{\chi_r\}$ used in the LCAO expansion is called an atomic basis set.

The choice of the basis set functions χ (the incomplete set) is one of the most important *practical* (numerical) problems of quantum chemistry. Yet, because it is of a technical character, we will just limit ourselves to a few remarks.

Although atomic functions do not need to be *atomic* orbitals (e.g., they may be placed *in-between* nuclei), in most cases they are centered directly on the nuclei of the atoms belonging to the molecule under consideration. If M is small (in the less precise calculations), the Slater AOs discussed above are often used as the expansion functions χ_r; for larger M (in more accurate calculations), the relation between χ_r and the orbitals of the isolated atoms is lost, and χ_r are chosen based on the numerical experience gathered from the literature.[59]

8.4.6 The Hartree–Fock–Roothaan method (SCF LCAO MO)

Clemens C.J. Roothaan (1918–2019), American physicist, professor at the University of Chicago. He became interested in this topic after recognizing that in the literature people write about the effective one-electron operator, but he could not find its mathematical expression.

The Hartree–Fock equations are nonlinear differential-integral equations, which can be solved by appropriate numerical methods. For example, in the case of atoms and diatomics the orbitals may be obtained in a numerical form.[60] High accuracy at long distances from the nuclei is their great advantage. However, the method is prohibitively difficult to apply for larger systems.

---

[59] For those who love such problems, we may recommend the article by S. Wilson "*Basis Sets*" in the book "*Ab initio Methods in Quantum Chemistry*," ed. K.P. Lawley, 1987, p. 439. In fact this knowledge is a little magic. Certain notations describing the quality of basis sets are in common use. For example, the symbol 6-31G* means that the basis set uses GTOs (G), the hyphen divides two electronic shells (here K and L, see p. 522). The K shell is described by a single AO, which is a certain linear combination (a "contracted orbital") of six GTOs of the 1s type, and the two digits, 31, pertain to the L shell and denote two contracted orbitals for each valence orbital (2s,2p$_x$,2p$_y$,2p$_z$), one of these contains three GTOs, the other one GTO (the latter is called "contracted," with a bit of exaggeration). The starlet corresponds to d functions used additionally in the description of the L shell (called polarization functions).

[60] J. Kobus, *Adv. Quantum Chem.*, 28(1997)1.

A solution is the use of the LCAO MO method (algebraization of the Fock equations). It leads to simplification of the computational scheme of the Hartree–Fock method.[61] In the SCF LCAO MO method, the Fock equations (complicated differential-integral equations) are solved in a very simple way. From (8.53) and (8.32) we have

George G. Hall (1925–2018), Irish physicist, professor of Mathematics at the University of Nottingham. His scientific achievements are connected to localized orbitals, ionization potentials, perturbation theory, solvation, and chemical reactions.

$$\hat{\mathcal{F}}\sum_s c_{si}\chi_s = \varepsilon_i \sum_s c_{si}\chi_s. \tag{8.54}$$

Making the scalar product with χ_r for $r = 1, 2, ..., M$ we obtain (the symbols h_{rs} and $\mathcal{F}_{rs}$ are the matrix elements of the corresponding operators in the AO basis set)

$$\sum_s (\mathcal{F}_{rs} - \varepsilon_i S_{rs})c_{si} = 0. \tag{8.55}$$

This is equivalent to the *Roothaan matrix equation*[62]:

$$\boldsymbol{\mathcal{F}}\mathbf{c} = \mathbf{S}\mathbf{c}\boldsymbol{\varepsilon}, \tag{8.56}$$

where $\mathbf{S}$ is the matrix of the overlap integrals $\langle \chi_r|\chi_s\rangle$ involving the AOs, $\boldsymbol{\varepsilon}$ is the diagonal matrix of the orbital energies[63] ε_i, and $\boldsymbol{\mathcal{F}}$ is the Fock operator matrix. Each of these matrices is square (of the order M); $\boldsymbol{\mathcal{F}}$ depends on $\mathbf{c}$ (and this is why it is a *pseudoeigenvalue* equation).

If the LCAO expansion is introduced to the expression for the total energy, Eq. (8.43) gives ($\varepsilon_i = (i|\hat{\mathcal{F}}|i)$)

$$E'_{HF} = \sum_i [h_{ii} + (i|\hat{\mathcal{F}}|i)] = \sum_{i=1}^{MO}\sum_{rs} c^*_{ri}c_{si}[(r|\hat{h}|s) + (r|\hat{\mathcal{F}}|s)] \equiv \frac{1}{2}\sum_{rs} P_{sr}[h_{rs} + \mathcal{F}_{rs}], \tag{8.57}$$

---

[61] The LCAO approximation was introduced to the Hartree–Fock method, independently, by C.C.J. Roothaan and G.G. Hall.

[62] Left-hand side: $L = \sum_s \mathcal{F}_{rs}c_{si}$, right-hand side: $P = \sum_{s,l} S_{rs}c_{sl}\varepsilon_{li} = \sum_{s,l} S_{rs}c_{sl}\delta_{li}\varepsilon_i = \sum_s S_{rs}c_{si}\varepsilon_i$. Comparison of both sides of the equation gives the desired result.

[63] In fact some approximations to them. Their values approach the orbital energies when the basis set of AOs gets closer to the complete basis set.

where **P** in the RHF method is called the *bond-order matrix*,

$$P_{sr} = 2\sum_{j}^{\mathrm{MO}} c_{rj}^{*} c_{sj},$$

and the summation goes over all occupied MOs. In consequence, a useful expression for the total energy in the Hartree–Fock method may be written as

$$E_{RHF} = \frac{1}{2}\sum_{rs}^{\mathrm{AO}} P_{sr}(h_{rs} + \mathcal{F}_{rs}) + \sum_{a<b} \frac{Z_a Z_b}{R_{ab}}, \tag{8.58}$$

where the first summation goes over the AOs. For completeness, we also give the expression for $\mathcal{F}_{rs}$,

$$\mathcal{F}_{rs} = (r|\hat{h} + 2\hat{\mathcal{J}} - \hat{\mathcal{K}}|s) = \tag{8.59}$$

$$h_{rs} + \sum_{i}^{\mathrm{MO}} [2(ri|si) - (ri|is)] = h_{rs} + \sum_{i}^{\mathrm{MO}} \sum_{pq}^{\mathrm{AO}} c_{pi}^{*} c_{qi} [2(rp|sq) - (rp|qs)] = \tag{8.60}$$

$$= h_{rs} + \sum_{pq}^{\mathrm{AO}} P_{qp}[(rp|sq) - \frac{1}{2}(rp|qs)], \tag{8.61}$$

where i is the index of an MO, and r and s denote the AOs.

The self-consistent field solution

The Hartree–Fock–Roothaan matrix equation is solved iteratively:

a) we assume an initial **c** matrix (i.e., also an initial **P** matrix; often in the zeroth iteration we put $\mathbf{P} = \mathbf{0}$, as if there were no electron repulsion);
b) we find the $\mathcal{F}$ matrix using matrix **P**;
c) we solve the Hartree–Fock–Roothaan equation (see Appendix M, p. 705) and obtain the M MOs, we choose the $N/2$ occupied orbitals (those of the lowest energy);
d) we obtain a new **c** matrix, and then a new **P**, etc.;
e) we go back to **a**).

The iterations are terminated when the total RHF energy (more liberal approach) or the coefficients **c** (less liberal one) change less than the assumed threshold values. Both these criteria (ideally fulfilled) may be considered as a sign that the output orbitals are already *self-consistent*.

Practically, these are never the exact solutions of the Fock equations, because a *limited* number of AOs was used, while expansion to the complete set requires the use of an infinite number of AOs (the total energy in such a case would be called the *Hartree–Fock limit energy*).

After finding the MOs (hence, also the Hartree–Fock function) in the SCF LCAO MO approximation, we may calculate the total energy of the molecule as the mean value of its Hamiltonian. We need only the *occupied* orbitals, and not the *virtual* ones for this calculation.

The Hartree–Fock method only takes care of the total energy *and completely ignores the virtual orbitals*, which may be considered as a kind of by-product.

8.4.7 Some practical problems

Size of the AO basis set

NUMBER OF MOs

The number of MOs obtained from the SCF procedure is always equal to the number of the AOs used. Each MO consists of various contributions of the same basis set of AOs. The apparent exception is when *due to symmetry*, the coefficients at some AOs are equal to zero).

For double occupancy, M needs to be larger than or equal to $N/2$. Typically we are forced to use large basis sets ($M \gg N/2$), and then along the occupied orbitals we get $M - N/2$ unoccupied orbitals (*virtual orbitals*). Of course, we should aim at the highest-*quality* MOs (i.e., how close they are to the solutions of the Fock equations), and avoiding large M (computational effort is proportional to M^4), but in practice a better basis set often means a larger M. The variational principle implies the ordering of the total energy values obtained in different approximations (Fig. 8.14).

It is required that a basis set as large as possible is used (mathematics: we approach the complete set), but we may also ask if a basis set dimension may be *decreased* freely (economy!). Of course, the answer is no! The absolute limit M is equal to half the number of the electrons, because only then can we create M spin orbitals and write the Slater determinant. However, in quantum chemistry, rather misleadingly, we call the minimum basis set the basis set resulting from inner shell and valence orbitals in the corresponding atoms. For example, the minimum basis set for a water molecule is $1s$, $2s$, and three $2p$ orbitals of oxygen and two $1s$ orbitals of hydrogen atoms, seven AOs in total (while the truly minimal basis would contain only $10/2 = 5$ AOs).

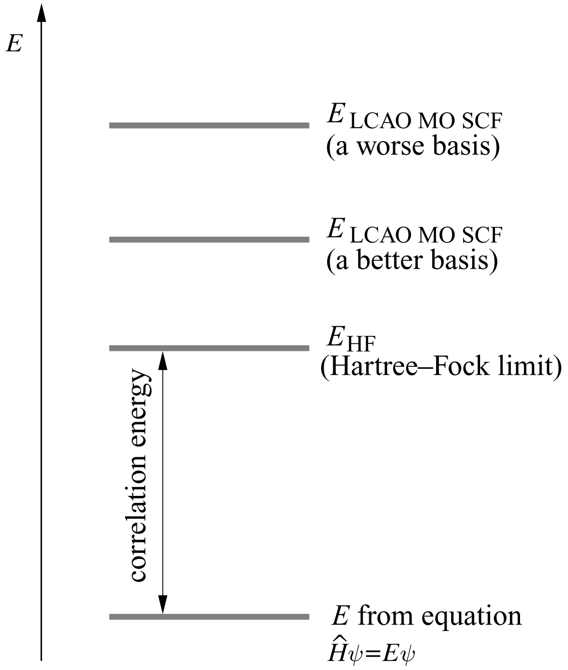


Fig. 8.14. The Hartree–Fock method is variational. The better the wave function, the lower the mean value of the Hamiltonian. An extension of the AO basis set (i.e., adding new AOs) *has to* lower the energy, and the ideal solution of the Fock equations gives the "Hartree–Fock limit." The ground-state eigenvalue of the Hamiltonian is *always* lower than the Hartree–Fock limit, because the Hartree–Fock method is able to produce only an approximation to the solution of the Schrödinger equation.

"Flip-flop"

The M MOs result from each iteration. We order them using the increasing orbital energy ε criterion, and then we use the $N/2$ orbitals of the lowest orbital energy in the Slater determinant – we call it the occupation of MOs by electrons. We might ask, why do we make the lowest-lying MOs occupied? *The variational principle does not hold for orbital energies.* And yet we do so (not trying all possible occupations), and only very rarely we get into trouble. The most frequent trouble is that the criterion of orbital energy leads to the occupation of one set of MOs in odd iterations, and another set of MOs in even ones (typically both sets differ by including/excluding one of the two MOs that are neighbors on the energy scale) and the energy resulting from the odd iterations is different from that of the even ones. Such behavior of the Hartree–Fock method is indeed annoying[64] (Fig. 8.15).

---

[64] There are methods for mastering this circus by using the matrix **P** in the k-th iteration, not taken from the previous iteration (as usual), but as a certain linear combination of **P** from the $k-1$ and $k-2$ iterations. When

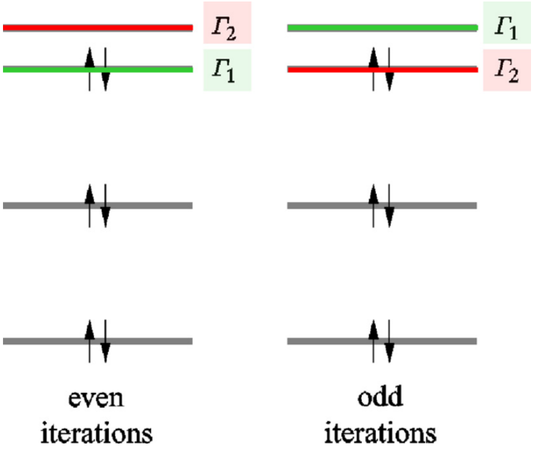


Fig. 8.15. A difficult case for the SCF method ("flip-flop"). We are sure that the orbitals exchange in subsequent iterations, because they differ in symmetry (Γ_1, Γ_2).

Dilemmas of AO centering

Returning to the total energy issue, we should recall that in order to decrease the total energy, we may move the nuclei (so far frozen during the Hartree–Fock procedure). This is called geometry optimization. Practically all calculations should be repeated for each nuclear geometry during such optimization.[65] And there is one more subtlety. As said before, the AOs are most often centered on the nuclei. When the nuclei are moved, the question arises whether a nucleus or not should pull its AOs to a new place.[66] If *not*, then this "slipping off" the nuclei will significantly increase the energy (independent of whether the geometry is improved or not). If *so*, then in fact we use different basis sets for each geometry, hence in each case we search for the solution in a slightly different space (because it is spanned by another basis set). People use the second approach. It is worth noting that the problem would disappear if the basis set of AOs were complete.

The problem of AO centering is a bit shameful in quantum chemistry. Let us consider the LCAO approximation and a real molecule such as Na_2CO_3. As mentioned above, the LCAO functions

---

the contribution of **P** from the $k-2$ iteration is large, in comparison with that from the $k-1$ iteration, it corresponds to a gentle attempt at quietening the nervous stallion.

[65] Let us take the example of CH_4. First, we set any starting geometry, say, a square-like planar. Now, we try to change the configuration to make it out-of-plane (the energy goes down). Taking the H–C–H angles as all equal (tetrahedral configuration) once more lowers the total energy computed. Putting all the C–H bonds of equal length gives even lower energy. Finally, by trying different C–H bond lengths we arrive at the optimum geometry (for a given AO basis set). In practice, such geometry changes are made automatically by computing the gradient of total energy.

[66] Even if the AOs were off the nuclei, we would have the same dilemma.

have to form a complete set. But which functions? Since they have to form a complete set, they may be chosen as the eigenfunctions of a certain Hermitian operator (e.g., the energy operator for the three-dimensional harmonic oscillator or the energy operator for the hydrogen atom or the Fock operator for uranium atom). We decide, and we are free to choose. In addition to this freedom, we add another freedom, i.e., that of the centering. Where should the eigenfunctions (of the oscillator or the hydrogen or uranium atom) of the complete set be centered, i.e., positioned in space? Since it is the complete set, *each way of centering is good by definition*. It really looks like this if we hold to the principles.

But in practical calculations, we never have the complete set at our disposal. We always need to limit it to a certain finite number of functions, and it does not represent any complete set. We usually try to squeeze the best results from our time and money. How do we do this? We apply our physical intuition to the problem, believing that it will pay off. First of all, intuition suggests the use of functions for some atom which is present in the molecule, and not those of the harmonic oscillator, or the hydrogen or uranium atom, which are absent from our molecule. And here we meet another problem. *Which* atom, because we have Na, C, and O in Na_2CO_3. It appears that

> the solution close to optimum is to take as a basis set *the beginnings* of several complete sets – each of them centered on one of the atoms.

So, we could center the $1s$, $2s$, $2p$, $3s$ orbitals on both Na atoms, and the $1s$, $2s$, $2p$ set on the C and O atoms.[67]

8.5 Back to foundations

8.5.1 When does the RHF method fail?

The reason for any Hartree–Fock method failure can be only one thing: the wave function is approximated as a *single* determinant. All possible catastrophes come from this, and we might even deduce when the Hartree–Fock method is not appropriate for description of a particular real system. First, let us ask when a single determinant would be OK. Well, if out of all Slater

---

[67] This is nearly everything, except for a small paradox, that if we are moderately poor (reasonable but not extensive basis sets), then our results will be good, but if we became rich (and performed high-quality computations using very large basis sets for each atom) then we would get into trouble. This would come from the fact that our basis set starts to look like six distinct complete sets. Well, that looks too good, doesn't it? We have an *overcomplete* set, and trouble must come. The overcompleteness means that any orbital from one set is already representable as a linear combination of another complete set. You would see strange things when trying to diagonalize the Fock matrix. No way! You can be sure that you would be begging to be less rich.

determinants which could be constructed from a certain spin orbital basis set, only its energy (i.e., the mean value of Hamiltonian for this determinant) were close to the true energy of the molecule. In such a case, only this determinant would matter in the linear combination of determinants,[68] and the others would have negligible coefficients. It could be so,[69] if the energies of the occupied orbitals were much lower than those of the virtual ones ("*Aufbau Prinzip*," p. 522). Indeed, various electronic states of different total energies may be approximately formed while the orbitals scheme is occupied by electrons, and if the virtual levels are at high energies, the total energy calculated from the "excited determinant" (replacement: occupied spin orbital → virtual spin orbital) would also be high (Fig. 8.16).

In other words, the danger for the RHF method is when the energy difference between HOMO and LUMO is small. For example, RHF completely fails to describe metals properly, because $\Delta = 0$ there.[70] Always, when the HOMO-LUMO gap is small, expect bad results.

Incorrect description of dissociation by the RHF method

Another example is provided by the H_2 molecule at long internuclear distances.

In the simplest LCAO MO approach, two electrons are described by the *bonding orbital* (χ_a and χ_b are $1s$ orbitals centered on the H nuclei, a and b, respectively; both obtained by using symmetry requirements)

$$\varphi_{bond} = \frac{1}{\sqrt{2(1+S)}}(\chi_a + \chi_b), \tag{8.62}$$

but there is another orbital, an *antibonding* one,

$$\varphi_{antibond} = \frac{1}{\sqrt{2(1-S)}}(\chi_a - \chi_b). \tag{8.63}$$

These names stem from the respective energies which are obtained if we accept that the MOs satisfy a sort of "Schrödinger equation" using an effective Hamiltonian (say, an analog of the Fock operator), $\hat{H}_{ef}\varphi = E\varphi$, and after introducing notation, the overlap integral $S = (\chi_a|\chi_b)$, $H_{aa} = (\chi_a|\hat{H}_{ef}\chi_a)$, *the resonance integral*[71] $H_{ab} = H_{ba} = (\chi_a|\hat{H}_{ef}\chi_b) < 0$. For the bonding

---

[68] The Slater determinants form the complete set, p. 468. In the configuration interaction method (which will be described in Chapter V2-2) the electronic wave function is expanded using Slater determinants.

[69] We shift here from the total energy to the one-electron energy, i.e., to the orbital picture.

[70] It shows up as strange behavior of the total energy per metal atom, which exhibits poorly decaying oscillations with an increasing of numbers of atoms. In addition, the exchange interactions notorious for fast (exponential) decay as calculated by the Hartree–Fock method are of a long-range character (see Chapter V2-1).

[71] This integral is negative. It is its sign which decides the energy effect of the chemical bond formation (because H_{aa} is nearly equal to the energy of an electron in the H atom, i.e., $-\frac{1}{2}$ a.u.).

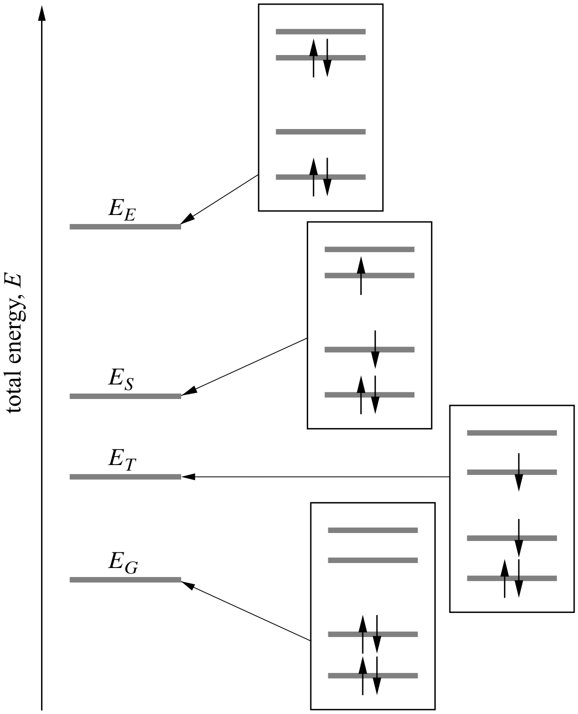


Fig. 8.16. In exact theory there is no such a thing as MOs. In such a theory we would only deal with the many-electron states and the corresponding energies of the molecule (left-hand side). If, nevertheless, we decided to *stick to the one-electron approximation*, we would have the MOs and the corresponding orbital energies (right-hand side). These one-electron energy levels can be occupied by electrons (0, 1, or 2) in various ways (the meaning of the occupation is given on p. 480), and a many-electron wave function (a Slater determinant) corresponds to each occupation. This function gives a certain mean value of the Hamiltonian, i.e., the total energy of the molecule. *In this way (in many cases) one value of the total energy of the molecule corresponds to a diagram of orbital occupation*. The case of the S and T states is somewhat more complex than the one shown here, and we will come back to it on p. 538.

orbital we have $E_{bond} = \left(\varphi_{bond}|\hat{H}_{ef}\varphi_{bond}\right) = \frac{1}{2(1+S)}\left[H_{aa} + H_{bb} + 2H_{ab}\right]$, which gives

$$E_{bond} = \frac{H_{aa} + H_{ab}}{1 + S} < H_{aa},$$

and for the antibonding orbital (similar derivation),

$$E_{antibond} = \frac{H_{aa} - H_{ab}}{1 - S} > H_{aa}.$$

The resonance integral H_{ab} and the overlap integral S decay exponentially when the internuclear distance R increases.

INCORRECT DISSOCIATION LIMIT OF THE HYDROGEN MOLECULE

Thus we have obtained the *quasidegeneracy* (a near-degeneracy of two orbitals) for long distances, while we need to occupy *only one* of these orbitals (bonding one) in the RHF method. The antibonding orbital is treated as virtual and, as such, is *completely ignored*. However, as a matter of fact, for long distances R, it corresponds to the same energy as the bonding energy.

We have to pay for such a huge drawback. And the RHF method pays, for its result significantly deviates (Fig. 8.17) from the exact energy for large R values (tending to the energy of the two isolated hydrogen atoms). This effect is known as an "incorrect dissociation of a molecule" in the RHF method (here exemplified by the hydrogen molecule). The failure may be explained in several ways and we have presented one point of view above.

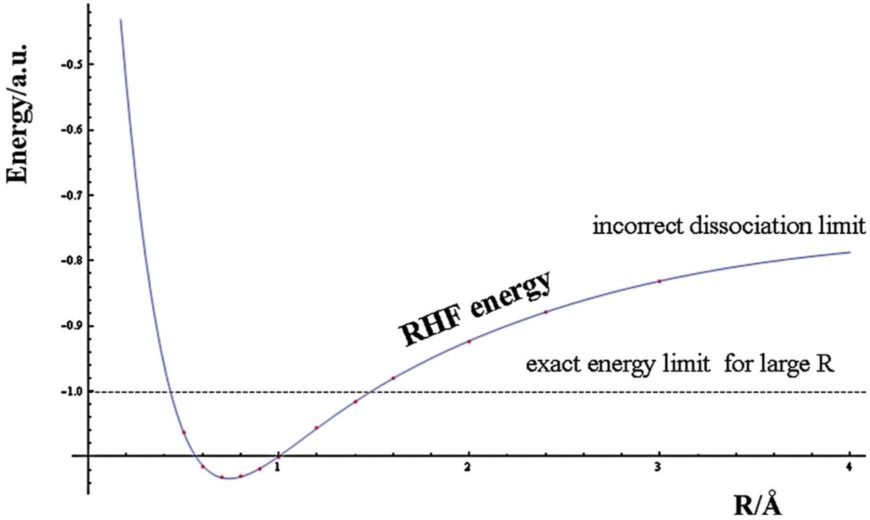


Fig. 8.17. Incorrect dissociation of H_2 in the molecular orbital (RHF) method. The wave function in the form of one Slater determinant leads to dissociation products, which are neither atoms, nor ions (while they should be two ground-state hydrogen atoms with energy $2E_H = -1$ a.u.). The Hartree–Fock computations have been carried out (program Gaussian) for the internuclear distances shown as points by using the atomic basis set known as 6-31G(d,p). The energy of an isolated hydrogen atom (E_H) has been calculated in the corresponding single-atom basis set. The curve displayed has been created as a multiexponential approximation by using the least mean square method. The minimum of the curve computed in such a way corresponds to the internuclear distance (R) of 1.385 a.u. (the Hartree–Fock limit is 1.370 a.u., while the solution of the Schrödinger equation in the Born–Oppenheimer approximation gives 1.401 a.u.). The corresponding energy equals −1.131 a.u., while the Hartree–Fock limit is −1.134 a.u., and the solution of the Schrödinger equation within the Born–Oppenheimer approximation gives −1.174 a.u.

8.5.2 Fukutome classes

Symmetry dilemmas and the Fock operator

We have derived the general Hartree–Fock method (GHF, p. 478) providing completely free variations for the spin orbitals taken from Eq. (8.1). As a result, the Fock equation of the form (8.27) was derived.

We then decided to limit the spin orbital variations via *our own* condition of the double occupancy of the MOs as the real functions. This has led to the RHF method and to the Fock equation in the form (8.32).

The Hartree–Fock method is a complex (nonlinear) machinery. Do the Hartree–Fock solutions have any symmetry features as compared to the Hamiltonian ones? This question may be asked both for the GHF method and for any *spin orbital constraints* (e.g., the RHF constraints). The following problems may be addressed:

- Do the output orbitals belong to the irreducible representations of the symmetry group (Appendix C on p. 605) of the Hamiltonian? Or, if we set the nuclei in the configuration corresponding to symmetry group G, will the canonical orbitals transform according to some irreducible representations of the G group? Or, still in other words, does the Fock operator exhibit the symmetry typical of the G group[72]?
- Does the same apply to electron density?
- Is the Hartree–Fock determinant an eigenfunction of the $\hat{S}^2$ and $\hat{S}_z$ operators?
- Is the probability density of finding a $\sigma = \frac{1}{2}$ electron equal to the probability density of finding a $\sigma = -\frac{1}{2}$ electron at any point of space?

Instabilities in the Hartree–Fock method

The above questions are connected to the *stability of the solutions*. The Hartree–Fock solution is stable if any change of the spin orbitals leads to a *higher* energy. We may put certain conditions for spin orbital changes. *Relaxing the condition of double occupancy may take various forms*, e.g., the paired orbitals may be equal but *complex*, or all orbitals may be *different* real functions, or we may admit them as different complex functions, etc. Could the energy increase along with this gradual orbital constraints removal? No, an energy increase is, of course, impossible, because of the variational principle; the energy might, however, remain constant or decrease.

---

[72] It has been shown that the translational symmetry in the RHF solution for polymers is broken and that the symmetry of the electron density distribution in polymers exhibits a unit cell twice as long as that of the nuclear pattern (Bond-Order Alternating Solution [BOAS], J. Paldus, J. Čižek, *J. Polym. Sci., Part C*, 29(1970)199; also J.-M. André, J. Delhalle, J.G. Fripiat, G. Hennico, J.-L. Calais, L. Piela, *J. Mol. Struct. (Theochem)*, 179(1988)393). The BOAS represents a feature related to the Jahn–Teller effect in molecules and to the Peierls effect in the solid state (see Chapter V2-1).

The general answer to this question (the character of the energy change) cannot be given since it depends on the molecule under study, interatomic distances, the AO basis set, etc. However, as shown by Fukutome[73] by a group theory analysis, there are exactly eight situations which *may* occur. Each of these leads to a characteristic shape of the set of occupied orbitals, which is given in Table 8.1. We may pass the borders between these eight classes of GHF method solutions while changing various parameters.

Table 8.1. Fukutome classes (for φ_{i1} and φ_{i2} see Eq. (8.1)).

Class	**Orbital components** $\begin{bmatrix} \varphi_{11} & \varphi_{21} & \cdots & \varphi_{N1} \\ \varphi_{12} & \varphi_{22} & \cdots & \varphi_{N2} \end{bmatrix}$	**Remarks**	**Eigenfunction of**
1	$\begin{bmatrix} \varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0 \\ 0 & \varphi_1 & 0 & \varphi_2 & \ldots & 0 & \varphi_{N/2} \end{bmatrix}$	φ_i real	RHF, $\hat{S}^2$,$\hat{S}_z$
2	$\begin{bmatrix} \varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0 \\ 0 & \varphi_1 & 0 & \varphi_2 & \ldots & 0 & \varphi_{N/2} \end{bmatrix}$	φ_i complex	$\hat{S}^2$,$\hat{S}_z$
3	$\begin{bmatrix} \varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0 \\ 0 & \varphi_1^* & 0 & \varphi_2^* & \ldots & 0 & \varphi_{N/2}^* \end{bmatrix}$	φ_i complex	$\hat{S}_z$
4	$\begin{bmatrix} \varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0 \\ 0 & \chi_1 & 0 & \chi_2 & \ldots & 0 & \chi_{N/2} \end{bmatrix}$	φ, χ real	UHF, $\hat{S}_z$
5	$\begin{bmatrix} \varphi_1 & 0 & \varphi_2 & 0 & \ldots & \varphi_{N/2} & 0 \\ 0 & \chi_1 & 0 & \chi_2 & \ldots & 0 & \chi_{N/2} \end{bmatrix}$	φ, χ complex	$\hat{S}_z$
6	$\begin{bmatrix} \varphi_1 & \chi_1 & \varphi_2 & \chi_2 & \cdots & \varphi_{N/2} & \chi_{N/2} \\ -\chi_1^* & \varphi_1^* & -\chi_2^* & \varphi_2^* & \cdots & -\chi_{N/2}^* & \varphi_{N/2}^* \end{bmatrix}$	φ, χ complex	
7	$\begin{bmatrix} \varphi_1 & \chi_1 & \varphi_2 & \chi_2 & \cdots & \varphi_{N/2} & \chi_{N/2} \\ \tau_1 & \kappa_1 & \tau_2 & \kappa_2 & \ldots & \tau_{N/2} & \kappa_{N/2} \end{bmatrix}$	$\varphi, \chi, \tau, \kappa$ real	
8	$\begin{bmatrix} \varphi_1 & \chi_1 & \varphi_2 & \chi_2 & \cdots & \varphi_{N/2} & \chi_{N/2} \\ \tau_1 & \kappa_1 & \tau_2 & \kappa_2 & \ldots & \tau_{N/2} & \kappa_{N/2} \end{bmatrix}$	$\varphi, \chi, \tau, \kappa$ complex	

RHF/UHF triplet instability

May the UHF method give lower energy for the hydrogen molecule than the RHF procedure?

Let us take the RHF function

$$\psi_{RHF} = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1(1) & \phi_1(2) \\ \phi_2(1) & \phi_2(2) \end{vmatrix}.$$

---

[73] A series of papers by H. Fukutome starts with the article in *Prog. Theor. Phys.*, 40(1968)998 and the review article *Int. J. Quantum Chem.*, 20(1981)955. I recommend a beautiful paper by J.-L. Calais, *Adv. Quantum Chem.*, 17(1985)225.

Both spin orbitals have a *common real orbital part* φ: $\phi_1 = \varphi\alpha$, $\phi_2 = \varphi\beta$.

Now we allow for a *diversification* of the orbital part (keeping the functions *real*, i.e., staying within the Fukutome class 4, usually called UHF in quantum chemistry) for both spin orbitals. We proceed slowly from the closed shell situation, using as the orthonormal spin orbitals

$$
\begin{aligned}
\phi_1' &= N_-(\varphi - \delta)\alpha, \\
\phi_2' &= N_+(\varphi + \delta)\beta,
\end{aligned}
$$

where δ is a small real correction to the φ function and N_+ and N_- are the normalization factors.[74] The electrons hate each other (Coulomb law) and may thank us for giving them separate apartments[75]: $\varphi + \delta$ and $\varphi - \delta$. We will worry about the particular mathematical shape of δ in a minute. For the time being let us see what happens to the UHF function. We have

$$
\begin{aligned}
\psi_{UHF} = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_1'(1) & \phi_1'(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} &= \frac{1}{\sqrt{2}} N_- \begin{vmatrix} [\varphi(1) - \delta(1)]\alpha(1) & [\varphi(2) - \delta(2)]\alpha(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} = \\
\frac{1}{\sqrt{2}} N_- &\left\{ \begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} - \begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \phi_2'(1) & \phi_2'(2) \end{vmatrix} \right\} = \\
\frac{1}{\sqrt{2}} N_+ N_- &\left\{ \begin{array}{l} \begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ [\varphi(1) + \delta(1)]\beta(1) & [\varphi(2) + \delta(2)]\beta(2) \end{vmatrix} - \\ \begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ [\varphi(1) + \delta(1)]\beta(1) & [\varphi(2) + \delta(2)]\beta(2) \end{vmatrix} \end{array} \right\} = \\
\frac{1}{\sqrt{2}} N_+ N_- &\left\{ \begin{array}{l} \begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ \varphi(1)\beta(1) & \varphi(2)\beta(2) \end{vmatrix} + \begin{vmatrix} \varphi(1)\alpha(1) & \varphi(2)\alpha(2) \\ \delta(1)\beta(1) & \delta(2)\beta(2) \end{vmatrix} - \\ \begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \varphi(1)\beta(1) & \varphi(2)\beta(2) \end{vmatrix} - \begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \delta(1)\beta(1) & \delta(2)\beta(2) \end{vmatrix} \end{array} \right\} = \\
N_+ N_- \psi_{RHF} + \frac{1}{\sqrt{2}} N_+ N_- &\left\{ [\varphi(1)\delta(2) - \varphi(2)\delta(1)] \left[\alpha(1)\beta(2) + \alpha(2)\beta(1) \right] \right\} - \\
\frac{1}{\sqrt{2}} N_+ N_- &\begin{vmatrix} \delta(1)\alpha(1) & \delta(2)\alpha(2) \\ \delta(1)\beta(1) & \delta(2)\beta(2) \end{vmatrix}.
\end{aligned}
$$

The first and last functions are singlets ($S_z = 0$, $S = 0$), while the second function represents a *triplet state* ($S_z = 0$, $S = 1$) (Appendix R on p. 731). Thus a small diversification of the orbital functions leads to some *triplet (second term) and singlet (third term) admixtures* to the

---

[74] Such a form is not fully equivalent to the UHF method, in which a general form of real orbitals is allowed, not just $\varphi \pm \delta$.

[75] Well, not quite, because the apartments overlap.

original singlet function $N_+N_-\psi_{RHF}$ (called *triplet contamination*, generally *spin contamination*). The former is proportional to δ and the latter to δ^2. Now the total wave function is no longer an eigenfunction of the $\hat{S}^2$ operator. How is this possible? If one electron has a spin coordinate of $\frac{1}{2}$ and the second one of $-\frac{1}{2}$, are they not paired? Well, not necessarily, because one of the triplet functions (which describes the *parallel* configuration of both spins[76]) is $\left[\alpha(1)\beta(2)+\alpha(2)\beta(1)\right]$.

Is the resulting UHF energy (calculated for such a function) lower than the corresponding RHF energy (calculated for ψ_{RHF}), i.e., is the RHF solution unstable towards UHF-type spin orbital changes (No. 4 in the table of Fukutome classes)?

It depends on a particular situation. A while before, we have promised to consider what the δ function should look like for the hydrogen molecule. In the RHF method, both electrons occupy the same molecular orbital φ. If we ensured within the UHF method that whenever one electron is close to nucleus a, the second one prefers to be closer to b; this would happily be accepted by the electrons, since they repel each other (the mean value of the Hamiltonian would decrease, this is welcome). Taking the $\delta=\varepsilon\tilde{\varphi}$ function (where $\tilde{\varphi}$ is the antibonding orbital and $\varepsilon>0$ is a small coefficient) would have such consequences. Indeed, the sum $\varphi+\delta=\varphi+\varepsilon\tilde{\varphi}$ takes a larger absolute value preferentially at one of the nuclei[77] (Fig. 8.18). Since both orbitals correspond to electrons with opposite spins, there will be some net spin on each of the nuclei.

A similar reasoning pertaining to function $\varphi-\delta=\varphi-\varepsilon\tilde{\varphi}$ results in opposite preferences for the nuclei. Such a particular UHF method, which uses virtual orbitals $\tilde{\varphi}$ to change RHF orbitals, carries the friendly name of the AMO approach.[78]

Per-Olov Löwdin (1916–2000), Swedish chemist and physicist, professor at the University of Uppsala (Sweden), founder and professor of the Quantum Theory Project at Gainesville University (Florida, USA), very active in organizing the scientific life of the international quantum chemistry community.

---

[76] To call them parallel is an exaggeration, since they form an angle of 70.5° (see Chapter 1, p. 33), but this is customary in physics and chemistry.

[77] In our example, the approximate bonding orbital is $\varphi=\frac{1}{\sqrt{2}}(1s_a+1s_b)$ and $\tilde{\varphi}=\frac{1}{\sqrt{2}}(1s_a-1s_b)$, hence $\varphi+\varepsilon\tilde{\varphi}=\frac{1}{\sqrt{2}}[(1+\varepsilon)\,1s_a+(1-\varepsilon)\,1s_b]$, while $\varphi-\varepsilon\tilde{\varphi}=\frac{1}{\sqrt{2}}[(1-\varepsilon)\,1s_a+(1+\varepsilon)\,1s_b]$. Thus one of the new orbitals has a larger amplitude at nucleus a, and the other one on nucleus b (as we had initially planned).

[78] P.-O. Löwdin, "*Alternant Molecular Orbitals*," *Symp. Mol. Phys., Nikko (Tokyo Maruzen)*, (1954), p. 13; also R. Pauncz, "*Alternant Molecular Orbitals*," Saunders, Philadelphia, 1967.

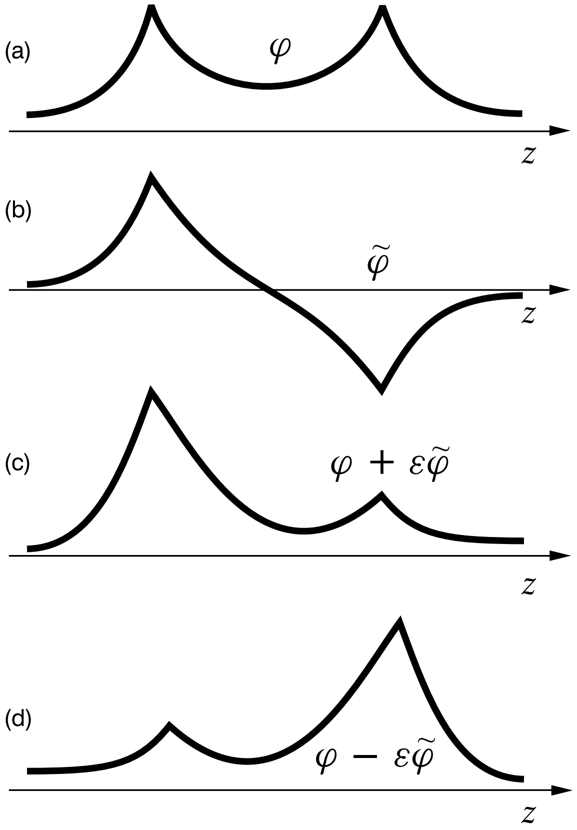


Fig. 8.18. The effect of mixing the bonding orbital φ (a) with the antibonding orbital $\tilde{\varphi}$ (b). A small admixture (c) of $\tilde{\varphi}$ to the orbital φ leads to an increase of the probability amplitude of the resulting orbital at the left nucleus, while a subtraction of $\tilde{\varphi}$ (d) leads to a larger probability amplitude of the resulting orbital at the right nucleus. Thus it results in partial separation of the spins $\frac{1}{2}$ and $-\frac{1}{2}$.

Now,

$$\psi_{UHF} = N_+N_-\psi_{RHF} + \frac{1}{\sqrt{2}}N_+N_-\varepsilon\left\{\left[\varphi(1)\,\tilde{\varphi}(2) - \varphi(2)\,\tilde{\varphi}(1)\right]\left[\alpha(1)\beta(2) + \alpha(2)\beta(1)\right]\right\} -$$

$$\frac{1}{\sqrt{2}}N_+N_-\varepsilon^2\begin{vmatrix} \tilde{\varphi}(1)\,\alpha(1) & \tilde{\varphi}(2)\,\alpha(2) \\ \tilde{\varphi}(1)\,\beta(1) & \tilde{\varphi}(2)\,\beta(2) \end{vmatrix} = N_+N_-\left[\psi_{RHF} + \varepsilon\sqrt{2}\psi_{T''} - \varepsilon^2\psi_E\right],$$

where the following notation is used for normalized functions: ψ_{RHF} for the ground state of the energy E_{RHF}, $\psi_{T''}$ for the triplet state of the energy E_T, and ψ_E for the singlet state with a doubly occupied antibonding orbital that corresponds to the energy E_E ("doubly excited").

Let us calculate the mean value of the Hamiltonian using the ψ_{UHF} function. Because of the orthogonality of the spin functions (remember that the Hamiltonian is independent of spin) we have $\left\langle\psi_{RHF}|\hat{H}\psi_{T''}\right\rangle = \langle\psi_{RHF}|\psi_{T''}\rangle = 0$, and we obtain (with accuracy up to ε^2 terms)

$$\bar{E}_{UHF} \approx \frac{\left\langle\psi_{RHF}|\hat{H}\psi_{RHF}\right\rangle + 2\varepsilon^2\left\langle\psi_{T''}|\hat{H}\psi_{T''}\right\rangle - 2\varepsilon^2\left\langle\psi_{RHF}|\hat{H}\psi_E\right\rangle}{\langle\psi_{RHF}|\psi_{RHF}\rangle + 2\varepsilon^2\langle\psi_{T''}|\psi_{T''}\rangle} =$$

$$\frac{E_{RHF} + 2\varepsilon^2 E_T - 2\varepsilon^2(\varphi\varphi|\tilde{\varphi}\tilde{\varphi})}{1+2\varepsilon^2} \approx E_{RHF} + 2\varepsilon^2\left[(E_T - E_{RHF}) - (\varphi\varphi|\tilde{\varphi}\tilde{\varphi})\right],$$

where the Taylor expansion and Slater–Condon' rule III have been used (p. 707): $\left\langle\psi_{RHF}|\hat{H}\psi_E\right\rangle = (\varphi\varphi|\tilde{\varphi}\tilde{\varphi}) > 0$. The last integral is greater than zero, because it corresponds to the Coulombic self-repulsion of a certain charge distribution.

It is now clear that if the singly excited triplet state $\psi_{T''}$ is of high energy as compared to the ground state ψ_{RHF}, then the spatial diversification of the opposite spin electrons (connected with the stabilization of $-2\varepsilon^2(\varphi\varphi|\tilde{\varphi}\tilde{\varphi})$) will not pay. But if E_T is close to the ground-state energy, then the *total energy will decrease upon the addition of the triplet state*, i.e., the RHF solutions will be unstable towards the UHF (or AMO)-type change of the orbitals.

This is the picture we obtain in numerical calculations for the hydrogen molecule (Fig. 8.19). At short distances between the atoms (up to 2.30 a.u.) the interaction is strong and the triplet state is of high energy. Then the variational principle does not allow the triplet state to contribute to the ground state and the UHF and the RHF give the same result. But beyond the 2.30 a.u. internuclear distance, the triplet admixture results in a small stabilization of the ground state and the UHF energy is lower than the RHF. For very long distances (when the energy difference between the singlet and triplet states is very small), the energy gain associated with the triplet component is very large.

We can see from Fig. 8.19b the drama occurring at $R = 2.30$ a.u. for the mean value of the $\hat{S}^2$ operator. For $R < 2.30$ a.u. the wave function preserves the singlet character, for larger R the triplet addition increases fast, and at $R = \infty$ the mean value of the square of the total spin $\hat{S}^2$ is equal to 1, i.e., half-way between the $S(S+1) = 0$ result for the singlet ($S = 0$) and the $S(S+1) = 2$ result for the triplet ($S = 1$), since the UHF determinant is exactly a

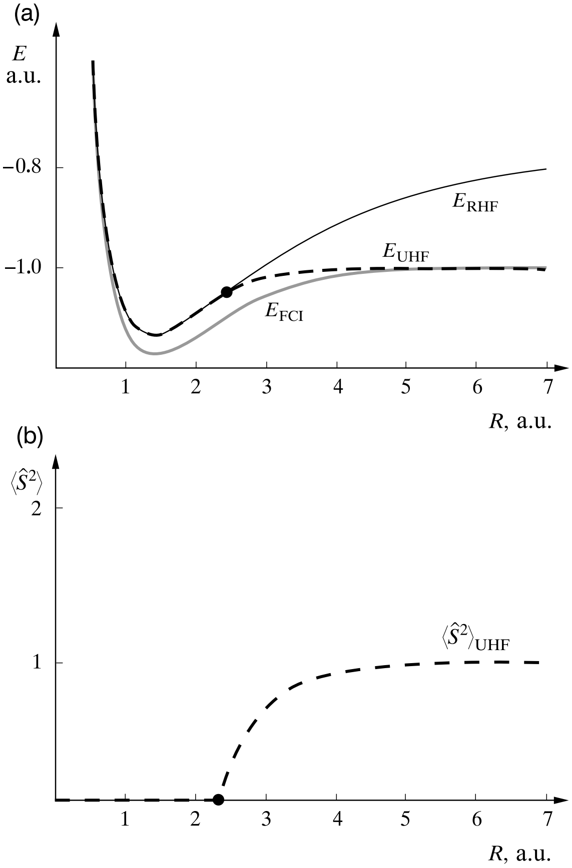


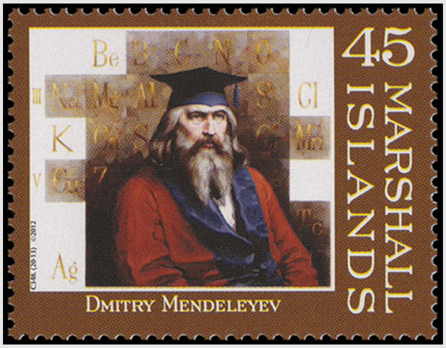
Fig. 8.19. (a) The mean value of Hamiltonian (E) calculated by the RHF and UHF methods. The lowest curve (E_{FCI}) corresponds to the accurate result (called the full configuration interaction method, see Chapter V2-2). (b) The mean value of the $\hat{S}^2$ operator calculated by the RHF and UHF methods. The energies $E_{RHF}(R)$ and $E_{UHF}(R)$ are identical for internuclear distances $R < 2.30$ a.u. For larger R values the two curves separate, and the RHF method gives an incorrect description of the dissociation limit, while the UHF method still gives a correct dissociation limit. For $R < 2.30$ a.u., the RHF and UHF wave functions are identical, and they correspond to a singlet, while for $R > 2.30$ the UHF wave function has a triplet contamination. T. Helgaker, P. Jørgensen, J. Olsen, "*Molecular Electronic Structure Theory*," Wiley, Chichester, 2000, reproduced with permission of John Wiley and Sons Ltd.

50%:50% singlet:triplet mixture. Thus, *one determinant (UHF) is able to describe properly the dissociation of the hydrogen molecule in its ground state (singlet), but at the expense of a large spin contamination (triplet admixture).*

RESULTS OF THE HARTREE–FOCK METHOD

8.6 Mendeleev periodic table

8.6.1 All atoms are similar to the hydrogen atom – the orbital model of an atom



Dimitryi Ivanovitch Mendeleev (1834–1907), Russian chemist, professor at the University in Petersburg, and later controller of the Russian Standards Bureau of Weights and Measures (after he was expelled from the University by the tsarist powers for supporting a student protest). He was born in Tobolsk, as the youngest of 14 children of a headmaster. In 1859 young Mendeleev – thanks to a tsarist scholarship – went to Paris and Heidelberg, where he worked with Robert Bunsen and Gustav Kirchhoff. After getting his PhD in 1865, he became professor of Chemistry at the age of 32 at the University in St. Petersburg. Since he had no good textbook, he started to write his own ("Principles of chemistry"). This is when he discovered one of the major human generalizations (1869): the periodicity law of chemical elements.

In 1905 he was nominated for the Nobel Prize, but lost by one vote to Henri Moissan, the discoverer of fluorine. The Swedish Royal Academy thus lost *its* chance, because in a year or so Mendeleev was dead. Many scientists had a similar intuition as Mendeleev had, but it was Mendeleev who completed the project, who organized the known elements in the table, and who predicted the existence of unknown elements. The following example shows how difficult it was for science to accept the periodic table. In 1864 John Newlands presented to The Royal Society in London his work showing similarities of the light elements, occurring for each eighth element with increasing atomic mass. The President of the meeting, quite amused by these considerations, suggested: "*haven't you tried to organize them according to the alphabetic order of their names?*

The Hartree–Fock method gives an approximate wave function for the atom of *any* chemical element from the Mendeleev periodic table (*orbital picture*). The Hartree–Fock method stands behind the *orbital model of atoms*. The model says essentially that a single Slater determinant can describe the atom to an accuracy that in most cases satisfies chemists. To tell the truth, the orbital model is in principle false,[79] but it is remarkable that nevertheless the conclusions drawn from it agree with experiment, at least qualitatively. It is quite exciting that

---

[79] Because the contributions of other Slater determinants (configurations) are not negligible (see Chapter V2-2).

the electronic structure of all elements can be generated to a reasonable accuracy using the *Aufbau Prinzip*, i.e., a certain scheme of filling the AOs of the hydrogen atom.

Thus, the simple and robust orbital model serves chemistry as a “work horse.” Let us take some examples. All the atoms are built on a similar principle. A nodeless spherically symmetric AO, of the lowest orbital energy, is called $1s$; next, the second-lowest (and also the spherically symmetric, one radial node) is called $2s$; etc. Therefore, when filling orbital energy states by electrons some electronic shells are formed: K ($1s^2$), L ($2s^2 2p^6$), ..., where the maximum for shell orbital occupation by electrons is shown.

The very foundations of a richness around us (its basic building blocks being atoms in the Mendeleev periodic table) result from a very simple idea: the proton and electron form a stable system called the hydrogen atom.

8.6.2 Shells and subshells

The larger the atomic number, the more complex the electronic structure. For neutral atoms the following occupation scheme applies (the “$<$” sign relates to the orbital energy):

Aufbau Prinzip

The *Aufbau Prinzip* relies on the following scheme of orbital energies (in ascending order):

	Orbital energy **Direction of occupying subshells $\Longrightarrow\Downarrow$**			**[Noble gas atoms]**
n				
1	$1s$			$[\mathrm{He}(2)]=1s^2$
2	$2s$		$2p$	$[\mathrm{Ne}(10)]=[\mathrm{He}(2)]2s^2 2p^6$
3	$3s$		$3p$	$[\mathrm{Ar}(18)]=[\mathrm{Ne}(10)]3s^2 3p^6$
4	$4s$	$3d$	$4p$	$[\mathrm{Kr}(36)]=[\mathrm{Ar}(18)]4s^2 3d^{10} 4p^6$
5	$5s$	$4d$	$5p$	$[\mathrm{Xe}(54)]=[\mathrm{Kr}(36)]5s^2 4d^{10} 5p^6$
6	$6s$ $4f$	$5d$	$6p$	$[\mathrm{Rn}(86)]=[\mathrm{Xe}(54)]6s^2 4f^{14} 5d^{10} 6p^6$
7	$7s$ $5f$	$6d$	$7p$	$[\mathrm{Uuo}(118)]=[\mathrm{Rn}(86)]7s^2 5f^{14} 6d^{10} 7p^6$

This sequence of the orbital energies may be viewed as a result of the restricted open shell Hartree–Fock (ROHF) calculations for atoms. As one can see the sequence differs from that for the hydrogen atom orbital energies. The main reason for this is that *unlike for the hydrogen atom*, for other atoms for a given principal quantum number n (counting atomic *shells*: K for $n = 1$, L for $n = 2$, M for $n = 3$, etc.), the orbital energies increase with the quantum number l, which define *subshells*: s, p, d, ... for $l = 0, 1, 2, ...$. This in turn reflects the fact that an electron

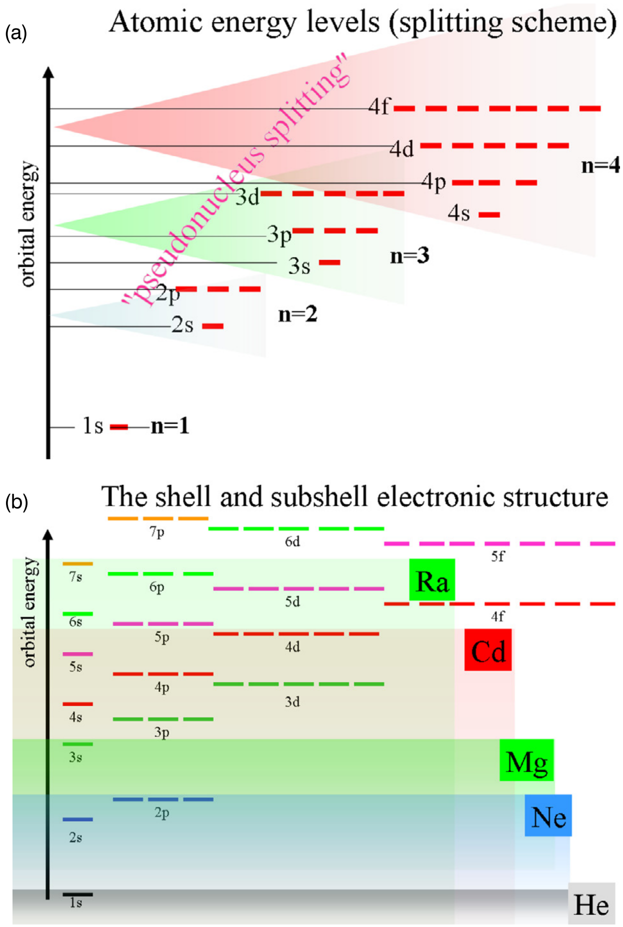
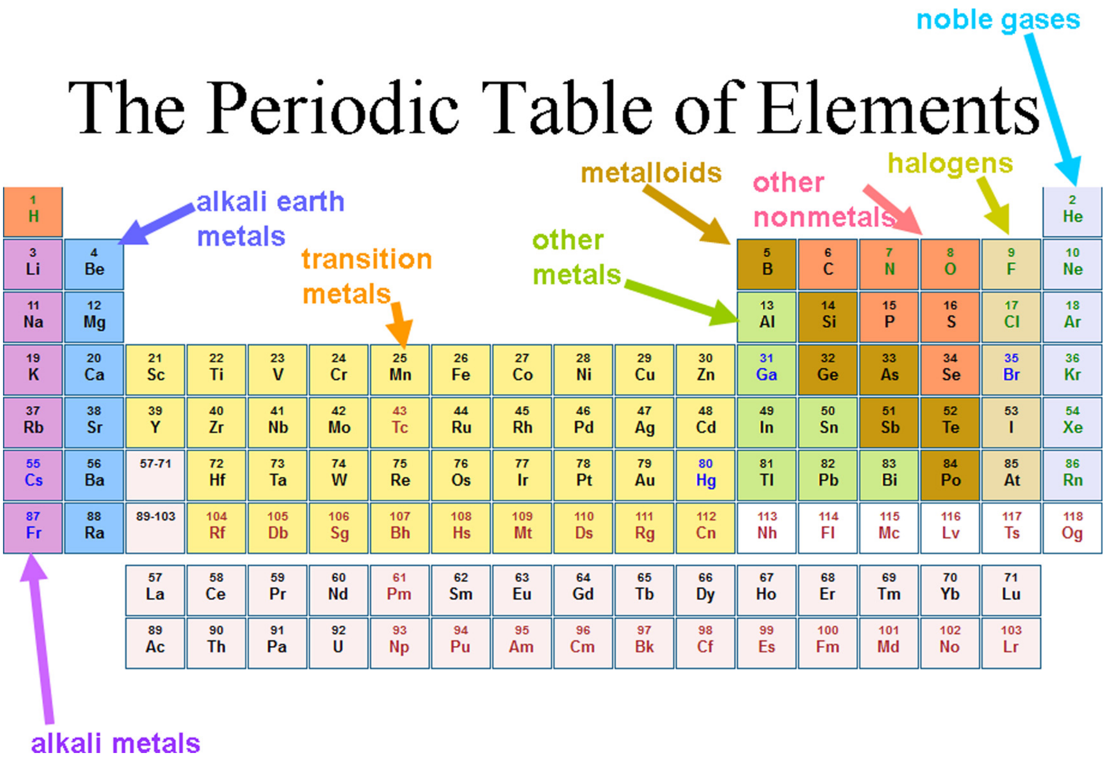


Fig. 8.20. The electronic shell structure for an atom. (a) A nonpoint-like "pseudonucleus" causes that, contrary to the hydrogen atom case, the energy of an electron depends not only on the principal quantum number n, but also on the azimuthal quantum number l (s,p,d,f... mean $l = 0, 1, 2, 3...$). This dependence is highlighted by the triangular backgrounds for the principal quantum numbers: $n = 1, 2, ...4$. (b) Quite fortunately it turns out that the sequence of the energy levels shown in the figure is preserved throughout the whole Mendeleev table (the energy scale is arbitrary). This makes it possible to predict the electronic shell structure for any atom just by adding electrons (the Pauli exclusion principle satisfied) consecutively from the bottom of the figure up to a desired level (shown by the corresponding rectangular background). Several examples are shown: He, Ne, Mg, Cd, Ra, with the number of electrons equal to 2, 10, 12, 48, 88, respectively.

in such an atom does not see a point-like nucleus. Instead it sees the point-like nucleus but surrounded by an electron cloud (of other electrons), i.e., with the nuclear charge screened by the electron cloud, *as if it were a spatially extended, huge "pseudonucleus."* This "pseudonucleus" controls now the electron motion and leads to such an "anomaly" that the $3d$ orbital energy is higher than that of $4s$, etc. (see Fig. 8.20a).

We cannot however expect that all nuances of atomic stabilities and of the ions corresponding to them might be deduced from a single simple rule like the *Aufbau Prinzip* that would replace the hard work of solving the Schrödinger equation (plus also the relativistic effects, Chapter 3) individually for each particular system.

> Increasing the nuclear charge of an atom (together with adding a proper number of its electrons) and *consecutively* occupying *the same sequence* of the electronic energy levels lead in a natural way to electronic shells and subshells (Fig. 8.20b). This produces a quasiperiodicity (sometimes called in chemistry the periodicity) of the valence shells, and in consequence a quasiperiodicity of all chemical and physical properties of the elements reflected in the Mendeleev periodic table (see Fig. 8.21).



1 H																	2 He
3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	57-71	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 Ra	89-103	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og
		57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu	
		89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr	

Fig. 8.21. The periodic table of elements.

This kind of "universality of the sequence" represents rather a fortunate (and to some extent accidental) didactic help that makes chemistry simpler to understand.

8.6.2.1 Universal language

There is no such thing as an orbital. This concept appears only after *we* decide to take a single Slater determinant as an *approximate* wave function. This is the orbital approximation introduced in the present chapter. Therefore, strictly speaking there is no such thing as an electronic energy level for an electron (see Fig. 8.20). However, it will turn out in the chapters to come that, quite fortunately, in most cases we encounter in chemistry, a single Slater determinant of Eq. (8.2) represents a rather satisfactory description of molecules. In other words, the overlap integral $\langle \psi_{\mathrm{HF}} | \psi_{\mathrm{exact}} \rangle$ in most cases is close to 1. It turns out that numerous and important properties and phenomena can be described in terms of orbitals, their occupancy by electrons, changes in this occupancy, etc. To give some examples: color of substances (the UV-VIS absorption and emission, Fig. 8.22a,b), X-ray production (Fig. 8.22c), the richness of chemistry (Fig. 8.22d) – they can all be explained using the energy level diagrams. Any science needs its special vocabulary that is able to express what happens and to explain it. The "orbital language" covers almost all we need in theory and practice of chemistry.

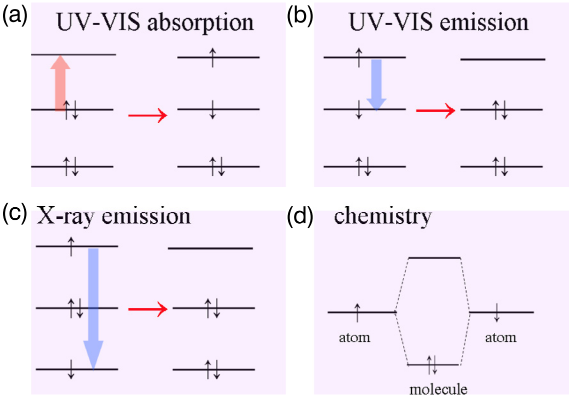


Fig. 8.22. This picture highlights one of the most important messages in chemistry: the orbital model creates the language of chemistry. Whatever chemical phenomenon you think of, you probably get its very essence by considering what electrons are doing in the corresponding orbital level diagrams. (a, b) The UV-VIS excitations explain the color of substances. (c) Filling with an electron a deep-lying hole in the electronic diagram leads to emission of very high-energy radiation, which are X-rays. (d) The chemical bond formation between any atoms requires their valence orbital energies to be of a similar value. The resulting bonding and antibonding MO energies are shown in the center (note using the same language for MOs).

Moreover,

> even if we soon will go beyond the orbital approximation, the orbital language will be still useful *even in this new situation*. Thus, the orbital language enables chemists to communicate about chemistry, to think about the essence of their discoveries in a common and simple realm of ideas. To tell the truth, we do not have any other language in chemistry.

These ideas fit the Mendeleev table, which follows from them and represents a kind of systematic metric not only of atoms, but also of their compounds and properties. Let us see how it works in chemical practice.

Example 1 (Noble gases). The atoms He, Ne, Ar, Kr, Xe, and Rn have a remarkable feature, in that they all exhibit the full occupancy of the electronic shells. According to the discussion on p. 503, what chemistry is all about is the outermost occupied orbitals (constituting the *valence shell*) that participate in forming chemical bonds. To have a chemical bond an atom has to offer or receive an electron. However,

> the noble gases have the highest ionization potential among all chemical elements and zero electron affinity. This confirms the common chemical knowledge that the noble gases do not form chemical bonds.

One has to remember, however, that even the closed shells of the noble gases can be opened either in extreme physical conditions (like pressure) or by using aggressive compounds that are able to detach an electron from them.

Example 2 (Alkali metals). The atoms Li, Na, K, Rb, Cs, and Fr have the following dominant electronic configurations (the inner shells have been abbreviated by reporting the corresponding noble gas atom configuration):

	Inner shell	Valence configuration
Li	[He]	$2s^1$
Na	[Ne]	$3s^1$
K	[Ar]	$4s^1$
Rb	[Kr]	$5s^1$
Cs	[Xe]	$6s^1$
Fr	[Rn]	$7s^1$

Since the valence shell decides about chemistry, no wonder the elements Li, Na, K, Rb, Cs, and Fr exhibit similar chemical and physical properties. Let us take any property we want, e.g., what we will get if the element is thrown into water. Lithium is a metal that reacts slowly with water, producing a colorless basic solution and hydrogen gas. Sodium is a metallic substance, and with water produces a very dangerous spectacle (wild dancing flames): it reacts rapidly with water to form a colorless basic solution and hydrogen gas. The other alkali metals are even more dangerous. Potassium is a metal as well, and reacts very rapidly with water, giving a colorless basic solution and hydrogen gas. Rubidium is a metal which reacts very rapidly with water, producing a colorless basic solution and hydrogen gas. Cesium metal reacts rapidly with water. The result is a colorless solution and hydrogen gas. Francium is very scarce and expensive, and probably no one has tried its reaction with water. *We may however expect, with very high probability, that if the reaction were allowed to take place, it would be faster than that with cesium and that a basic solution would be produced.*

However, maybe all elements react rapidly with water to form a colorless basic solution and hydrogen gas? Well, this is not true. The noble gases do not. They only dissolve in water, without any accompanying chemical reaction. They seem to be for the water structure just inert balls of increasing size. No wonder then that (at 293 K) one obtains the following monotonic sequence of their solubilities: 8.61 (He) $<$ 10.5 (Ne) $<$ 33.6 (Ar) $<$ 59.4 (Kr) $<$ 108.1 (Xe) $<$ 230 (Rn) $\mathrm{cm}^3/\mathrm{kg}$.

Example 3 (Halogens). Let us see whether there are other families. Let us concentrate on atoms which have p^5 as the outermost configuration. Using our scheme of orbital energies we produce the following configurations with this property: $[\mathrm{He}]2s^2 2p^5$ with nine electrons, i.e., F, $[\mathrm{Ne}]3s^2 3p^5$ with 17 electrons, i.e., Cl, $[\mathrm{Ar}]4s^2 3d^{10} 4p^5$ with 35 electrons, i.e., Br, $[\mathrm{Kr}]4d^{10} 5p^5$ with 53 electrons, i.e., I, $[\mathrm{Xe}]6s^2 4f^{14} 5d^{10} 6p^5$ with 85 electrons, i.e., At. Are these elements similar? What happens to halogens in contact with water? Maybe they react very rapidly with water, producing a colorless basic solution and hydrogen gas like the alkali metals, or do they just dissolve in water like the noble gases? Let us see.

Fluorine reacts with water to produce oxygen (O_2) and ozone (O_3). This is strange in comparison with alkali metals. Next, chlorine reacts with water to produce hypochlorite (OCl^-). Bromine and iodine do a similar thing, producing hypobromite (OBr^-) and hypoiodite (OI^-). Nothing is known about the reaction of astatine with water. Apart from the exceptional behavior of fluorine,[80] there is no doubt we have a family of elements. This family is different from the noble gases and from the alkali metals.

---

[80] For light elements the details of the electronic configuration play a more important role. For example, hydrogen may also be treated as an alkali metal, but its properties differ widely from the properties of the other members of this family.

Thus, the families show evidence that elements differ widely among families, but much less within a family, with rather small (and often monotonic) changes within it. This is what (quasi)periodicity of the Mendeleev periodic table is all about. The families are called *groups* (usually columns) in the Mendeleev table.

The Mendeleev table represents more than just a grid of information; it is a kind of compass in chemistry. Instead of having a sort of wilderness, where all the elements exhibit their unique physical and chemical properties as *deus ex machina*, we obtain *understanding* that the animals are in a zoo, and are not unrelated, that there are some families which exhibit a similar structure and occupancy of the outer electronic shells. Moreover, it became clear for Mendeleev that there were cages in the zoo with animals yet waiting to be discovered. The animals could have been described in detail *before they were actually found by experiment*. This periodicity pertains not only to the chemical and physical properties of elements, but also to all parameters that appear in theory and are related to atoms, molecules, and crystals.

8.6.3 Educated guess of atomic orbitals – the Slater rules

John Slater, by analyzing what his young coworkers were bringing from the computer room (as he wrote: "*when the boys were computing*"), noticed that he can quite easily *predict* the approximate shape of the atomic orbitals without any calculation. According to his rules it is enough to introduce a screening in the STO exponent in order to get a rough idea of realistic AOs for a particular atom. Slater proposed the STOs $\chi_{STO,nlm}(r,\theta,\phi) = N_{nlm} r^{n-1} \exp(-\zeta r)\, Y_l^m(\theta,\phi)$ with $\zeta = \frac{Z-\sigma}{n}$, where Z stands for the nuclear charge, σ tells us how other electrons screen (i.e., effectively diminish) the charge of the nucleus for an electron "sitting on" the analyzed STO, and n is the principal quantum number, the same as that in the *Aufbau Prinzip*.[81]

We focus on the electron occupying the orbital in question (that for which we are going to find ζ), and we try to estimate "what the electron sees." The electron sees that the nucleus charge is screened by its fellow electrons. The Slater rules for screening σ are the following:

- Write down the electronic configuration of the atom by grouping its orbitals in the following way: $[1s][2s2p][3s3p][3d]$.... The screening is determined by the group the designed AO belongs to.
- Electrons from the groups to the right of this sequence give zero contribution.
- The electrons in the same group contribute 0.35 each, except the $[1s]$ fellow electron (if we consider the $1s$ orbital), which contributes 0.30.
- For an electron in an $[ns\,np]$ group each electron in the $n-1$ group contributes 0.85, for lower groups (more on the left-hand side) each contributes 1.0, and for the $[nd]$ and $[nf]$ groups, all electrons in the groups to the left contribute 1.0.

---

[81] For $n \geq 4$ some modifications have been designed.

Example 4 (Carbon atom). Configuration in groups: $[1s^2][2s^2 2p^2]$. There will be two σ's: $\sigma_{1s} = 0.30, \sigma_{2s} = \sigma_{2p} = 3 \cdot 0.35 + 2 \cdot 0.85 = 2.75$. Hence, $\zeta_{1s} = \frac{6-0.30}{1} = 5.70, \zeta_{2s} = \zeta_{2p} = \frac{6-2.75}{2} = 1.625$, which means the following AOs: $1s_C = N_{1s} \exp(-5.70r)$, $2s_C = N_{2s} \exp(-1.625r)$, $2p_{x,C} = N_{2p} x \exp(-1.625r)$, $2p_{y,C} = N_{2p} y \exp(-1.625r)$, $2p_{z,C} = N_{2p} z \exp(-1.625r)$. The advantage of such estimations is that prior to any computation we have already an idea what kind of AOs we should expect from more accurate approaches.

8.6.4 Atomic radii

The Slater rules enable us to look into the problem of atoms' sizes. First, since every orbital extends to infinity, we are certainly in a delicate situation, because it is not evident what one should call an atomic radius. On the other hand the atoms compose crystals and molecules and two atoms do not occupy the same space; roughly speaking they behave like two soft balls. It is reasonable (and first of all practical) to agree about some radii of these balls. There are several concepts about possible criteria. We may agree about an atom radius:

- by assuming (arbitrarily) a certain electron frontier density threshold;
- we may judge about the atomic radii from the chemical bond lengths assuming that they are simply a sum of the two atomic *covalent* radii;
- for ionic crystals, we may introduce in a similar way a concept of the *ionic* radii;
- from the van der Waals intermolecular or interatomic complexes we may have a concept of the *van der Waals atomic radii*.

Although all these choices will give different sets of the atomic radii, still the resulting numbers are useful in the specific situations of molecules, ionic crystals, van der Waals complexes, etc. For example, they may be used to predict the interatomic distances in some new compounds or crystals.

In a naive approach one may suppose that the larger the atomic number Z (i.e., the number of electrons) the larger should be the atom. Instead, the atomic radii (whatever the consensus used) show a spectacular dependence on Z (Fig. 8.23).

Note however that what we see in Fig. 8.23 may be predicted from the Slater rules. What counts for the atomic radius are the most extended STO orbitals, i.e., the contribution from the last group of orbitals (with the lowest orbital exponent $\zeta = \frac{Z-\sigma}{n}$) occupied by valence electrons. Three important features of the STO formula are in play when we keep increasing the atomic number:

- The most important contribution for the frontier density comes from the last group, i.e., the valence electrons (corresponding for a given atom to its *largest* n). When one begins to

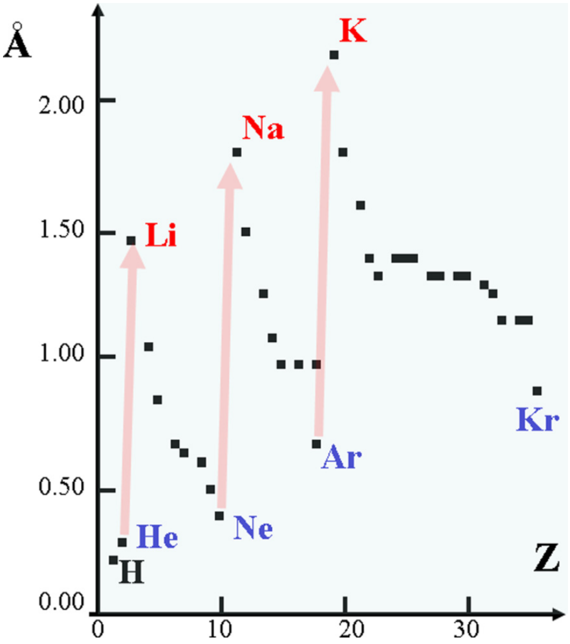


Fig. 8.23. The covalent atomic radii in Å as a function of the atomic number Z. The arrows show a giant increasing of the atom size when a new electron shell appears. Then the atoms shrink considerably when this shell is filled by electrons one by one until the noble gas electron configuration is reached.

create this group, which is the case for alkali metals, Li, Na, K, etc., then ζ abruptly drops down, which means a *giant increasing in atomic radius for alkali metals* when compared to lighter atoms (see Fig. 8.23).

- This increasing of the radius is magnified for larger n, because of the larger and larger prefactor r^{n-1} in the STO (and $r^{2(n-1)}$ in the electron density), which has to be damped down by the weaker and weaker exponential decaying.
- What happens next is maybe unexpected. When we keep increasing the atomic number for this group occupancy, we increase both Z and σ. However, Z increases much faster, since the fellow electrons from the same group contribute only 0.35 each for the screening coefficient σ, but 1 for Z. This, however, means increasing $\zeta = \frac{Z-\sigma}{n}$ (with n kept constant). Therefore, for increasing occupancy of this group this results in *lowering* of the atomic radius! This is what we see in Fig. 8.23 as a very large effect (for example, from Li to Ne, from Na to Ar, etc.) as the radii steeply go down.

All this reflects in this particular case the famous (quasi)periodic character of the Mendeleev table resulting from the shell-like electronic structure.

8.7 The nature of the chemical bond – quantum makes a difference

As shown on p. 511, the MO method explains the nature of the chemical bond via the argument that the orbital energy in the molecule is lower than that in the isolated atom. But why is this so? Which interactions decide bond formation? Do chemical bonds have their origin in quantum or in classical mechanics?

To answer these questions, we will analyze the simplest case: chemical bonding in a molecular ion H_2^+. It seems that quantum mechanics is not required here: we deal with one repulsion and two attractions. No wonder there is bonding, since the net effect is one attraction. But the same applies, however, to the dissociated system (the hydrogen atom and the proton). Thus, the story is becoming more subtle.

8.7.1 The simplest chemical bond: H_2^+ in the MO picture

Let us analyze chemical bonding as viewed by the poor version of the MO method (only two $1s$ hydrogen atom orbitals are used in the LCAO expansion, see Appendix S on p. 735). Much can be seen thanks to such a poor version. The mean kinetic energy of the (only) electron of H_2^+, residing on the bonding MO $\varphi = [2(1+S)]^{-1/2}(a+b)$, is given as (for simplicity a and b denote the atomic $1s$ orbitals centered, respectively, on the a and b nuclei)

$$\bar{T} \equiv (\varphi|\hat{T}\varphi) = \frac{T_{aa}+T_{ab}}{1+S}, \tag{8.64}$$

where S is the overlap integral $S=(a|b)$, and

$$T_{aa} = (a|-\frac{1}{2}\Delta|a) = T_{bb},$$
$$T_{ab} = (a|-\frac{1}{2}\Delta|b) = T_{ba}.$$

The noninteracting hydrogen atom and the proton have the mean kinetic energy of the electron equal to T_{aa}. The kinetic energy change is thus

$$\Delta T = \bar{T} - T_{aa} = \frac{T_{ab} - ST_{aa}}{1+S}. \tag{8.65}$$

Let us note (please recall the a and b functions are the eigenfunctions of the hydrogen atom Hamiltonian) that $T_{ab} = E_H S - V_{ab,b}$ and $T_{aa} = E_H - V_{aa,a}$, where E_H is the ground-state energy of the H atom,[82] and

$$V_{ab,b} = V_{ab,a} = -(a|\frac{1}{r_b}|b),$$

---

[82] For example, $T_{ab} = (a|-\frac{1}{2}\Delta|b) = (a|-\frac{1}{2}\Delta - \frac{1}{r_b} + \frac{1}{r_b}|b) = E_H S + (a|\frac{1}{r_b}|b) = E_H S - V_{ab,b}$.

$$V_{aa,a} = -(a|\frac{1}{r_a}|a).$$

Now, ΔT can be presented as

$$\Delta T = -\frac{V_{ab,a} - SV_{aa,a}}{1+S}, \tag{8.66}$$

because the terms with E_H cancel each other out. In this way the change in kinetic energy of the electron when a molecule is formed may be *formally* presented as the integrals describing the potential energy. Using the results of Appendix S (p. 735) we have $-V_{ab,a} + SV_{aa,a} = (1+R)\exp(-R) - S = S - \frac{R^2}{3}\exp(-R) - S = -\frac{R^2}{3}\exp(-R) \leq 0$. This means $\Delta T \leq 0$, i.e.,

> the kinetic energy change stabilizes the molecule.

This agrees with our intuition that an electron in the molecule has more space ("larger box," see p. 189), and the energy levels in the box (potential energy is zero in the box, therefore we mean kinetic energy here) decrease when the box dimension increases. This example shows that some abstract problems which can be solved exactly (here the particle in the box) serve as a beacon for more complex problems.

Now let us calculate the change in the mean potential energy. The mean potential energy of the electron (the nucleus–nucleus interaction will be added later) equals

$$\bar{V} = (\varphi|V|\varphi) = (\varphi| - \frac{1}{r_a} - \frac{1}{r_b}|\varphi) = \frac{(V_{aa,a} + V_{aa,b} + 2V_{ab,a})}{1+S}, \tag{8.67}$$

while in the hydrogen atom it is equal to $V_{aa,a}$. The difference, ΔV, is

$$\Delta V = \frac{(-SV_{aa,a} + 2V_{ab,a} + V_{aa,b})}{1+S}. \tag{8.68}$$

We can see that when the change in *total* electronic energy $\Delta E_{el} = \Delta T + \Delta V$ is calculated, some kinetic energy terms will cancel out the corresponding potential energy terms, and *potential energy will dominate during bond formation*; we have

$$\Delta E_{el} = \frac{V_{ab,a} + V_{aa,b}}{1+S}. \tag{8.69}$$

To obtain the change, ΔE, in the total energy of the system during bond formation, we have to add the term $1/R$ describing the nuclear repulsion, i.e.,

$$\Delta E = \frac{V_{ab,a}}{1+S} + \frac{V_{aa,b}}{1+S} + \frac{1}{R}. \tag{8.70}$$

This formula is identical (because $V_{ab,a} = V_{ab,b}$) to the difference in orbital energies in the molecule H_2^+ and in the hydrogen atom, as given in Appendix S on p. 735.

Eq. (8.70) can be easily interpreted. Let us first consider the electron density described by the φ orbital: $\varphi^2 = [2(1+S)]^{-1}(a^2 + b^2 + 2ab)$. Let us note that the density can be divided into the part ρ_a close to nucleus a, ρ_b close to nucleus b, and ρ_{ab} concentrated in the bonding region, i.e.,

$$\varphi^2 = \rho_a + \rho_b + \rho_{ab}, \tag{8.71}$$

where $\rho_a = [2(1+S)]^{-1}a^2$, $\rho_b = [2(1+S)]^{-1}b^2$, $\rho_{ab} = [(1+S)]^{-1}ab$. It can be seen[83] that the charge associated with ρ_a is $[-2(1+S)]^{-1}$, the charge on b is the same, and the overlap charge ρ_{ab} is $-S/(1+S)$. Their sum gives $-2/[2(1+S)] - 2S/[2(1+S)] = -1$ (the unit electronic charge). The formula for ΔE may also be written as (we use symmetry: the nuclei are identical, and the a and b orbitals differ only in their centers)

$$\Delta E = \frac{V_{ab,a}}{[2(1+S)]} + \frac{V_{ab,b}}{[2(1+S)]} + \frac{V_{aa,b}}{[2(1+S)]} + \frac{V_{bb,a}}{[2(1+S)]} + \frac{1}{R}. \tag{8.72}$$

Now it is clear that this formula describes the Coulombic interaction (Fig. 8.24a,b):

- of the electron cloud from the a atom (with density $\frac{1}{2}\rho_{ab}$) with the b nucleus and *vice versa* (the first two terms of the expression),
- of the electron cloud of density ρ_a with the b nucleus (third term),
- of the electron cloud of density ρ_b with the a nucleus (fourth term),
- of the a and b nuclei (fifth term).

If we consider *classically* a proton approaching a hydrogen atom, the only terms for the total interaction energy are (Fig. 8.24c)

$$\Delta E_{class} = V_{aa,b} + \frac{1}{R}. \tag{8.73}$$

The difference between ΔE and ΔE_{class} only originates from the difference in electron density calculated quantum mechanically and classically (cf. Fig. 8.24). The ΔE_{class} is a *weak* interaction (especially for long distances) and tends to $+\infty$ for small R, because[84] of the $1/R$ term. This can be understood because ΔE_{class} is the difference between two Coulombic interactions: of a point charge with a spherical charge cloud, and of the respective two point charges (the difference is called *penetration energy*); ΔE contains two more terms in comparison with ΔE_{class}: $V_{ab,a}/[2(1+S)]$ and $V_{ab,b}/[2(1+S)]$, and both decrease exponentially to $V_{aa,a} = -1$ a.u. when R decreases to zero. Thus these terms are not important for long distances, stabilize

---

[83] After integrating of ρ_a.

[84] $V_{aa,a}$ is finite.

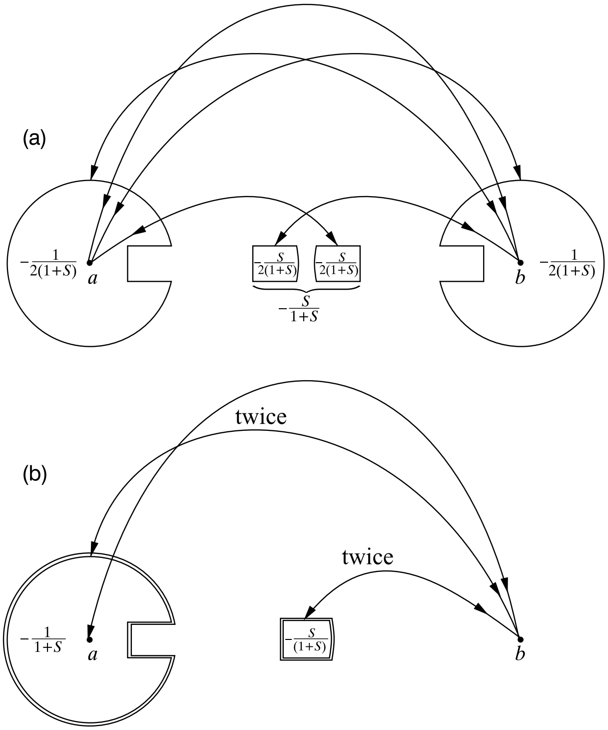


Fig. 8.24. The nature of the chemical bond in the H_2^+ molecule (schematic interpretation according to Eq. (8.72)):

— (a) *The quantum picture of the interaction*. The total electron density $\varphi^2 = \rho_a + \rho_b + \rho_{ab}$ consists of three electronic clouds: $\rho_a = [2(1+S)]^{-1}a^2$, bearing the $-\frac{1}{2(1+S)}$ charge (in a.u.) concentrated close to the a nucleus, a similar cloud $\rho_b = [2(1+S)]^{-1}b^2$, concentrated close to the b nucleus, and the rest (the total charge is -1) $\rho_{ab} = [(1+S)]^{-1}ab$, bearing the charge of $-2\frac{S}{2(1+S)}$, concentrated in the middle of the bond. The losses of the charge on the a and b atoms have been shown schematically, since the charge in the middle of the bond originates from these losses. The interactions have been denoted by arrows: there are all kinds of interactions of the fragments of one atom with the fragments of the second one.

— (b) *The quantum picture – summary* (we will need it in just a moment). This scheme is similar to (a), but it has been emphasized that the attraction of ρ_a by nucleus b is the same as the attraction of ρ_b by nucleus a, hence they were both presented as one interaction of nucleus b with charge of $-2\rho_a$ at a (hence the double contour line in the figure). In this way two of the interaction arrows have disappeared as compared to (a).

— (c) *The classical picture of the interaction between the hydrogen atom and a proton*. The proton (nucleus b) interacts with the electron of the a atom, bearing the charge of $-1 = -2\frac{1}{2(1+S)} - 2\frac{S}{2(1+S)}$ and with nucleus a. Such division of the electronic charge indicates that it consists of two fragments ρ_a (as in (b)) and of two fragments of the $-\frac{S}{2(1+S)}$ charge (i.e., similar to (b), but *centered in another way*). A major difference as compared to (b) is that in the classical picture nucleus b interacts with two *quite distant* electronic charges (put in the vicinity of nucleus a), while in the quantum picture (schemes (a) and (b)) the same charges attract at short distance.

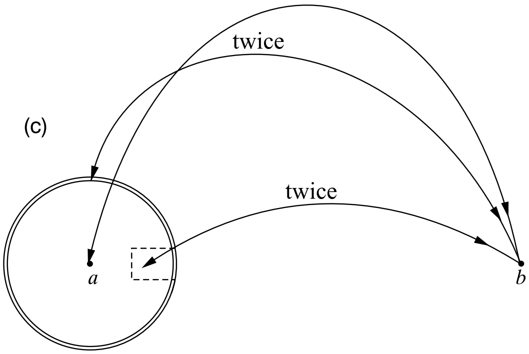


Fig. 8.24. (*continued*)

the molecule for intermediate distances (and provide the main contribution to the chemical bond energy), and are dominated by the $1/R$ repulsion for small distances.

In the quantum case, for the electron charge cloud connected with the a nucleus, a^2 is decreased by a charge of $S/(1+S)$, *which shifts to the halfway point towards nucleus* b. In the classical case, there is no charge shift – the whole charge is close to a. In both cases there is the nucleus–nucleus and the nucleus–electron interaction. The first is identical, but the latter is completely different in both cases. Yet even in the latter interaction, there is something in common: the interaction of the nucleus with the major part of the electron cloud, with charge $-[1-S/(1+S)]=-1/(1+S)$. The difference in the cases is the interaction with the remaining part of the electron cloud,[85] the charge $S/(1+S)$.

> In the classical view this cloud is located close to distant nucleus a, in the quantum view it is in the middle of the bond. The latter is much better for bonding. This interaction, of the (negative) electron cloud ρ_{ab} in the middle of the bond with the positive nuclei, stabilizes the chemical bond.

Fig. 8.24 shows an idea of the quantum mechanical nature of chemical bonds by using some particular *schematically drawn* electronic clouds, their interaction favoring the quantum mechanical picture over the classical one. However, there is nothing that prevents us *to show the same clouds in a realistic way* (Fig. 8.25).

---

[85] This simple interpretation gets more complex when further effects are considered, such as contributions to energy due to the polarization of the spherically symmetric AOs or the exponent dependence of the $1s$ orbitals (i.e., the dimensions of these orbitals) on the internuclear distance. When there are several factors at play (some positive, some negative) and when the final result is of the order of a single component, then *we* decide which component carries responsibility for the outcome. The situation is similar to that in Parliament, when two MPs from a small party are blamed for the result of a vote (the party may be called the balancing party) while perhaps 200 others who also voted in a similar manner are left in peace.

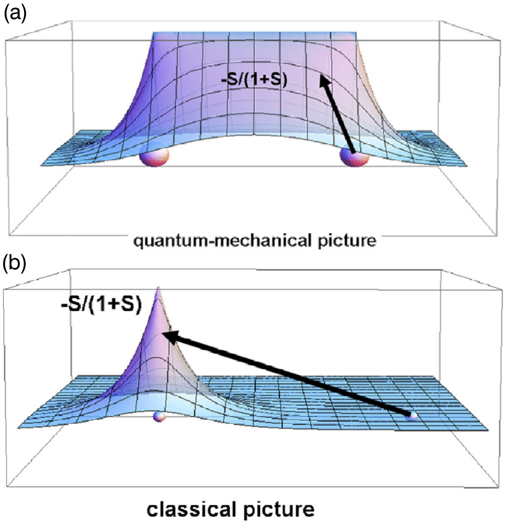


Fig. 8.25. The same reasoning as in Fig. 8.24, but this time the key charge distributions are drawn in a realistic way instead of a schematic diagram. The sections $z = 0$ of the two crucial electronic charge distributions are drawn. (a) In the quantum mechanical picture what decides about chemical bonding is an electron cloud $-\frac{1}{1+S} 1s_a \cdot 1s_b \equiv -\frac{1}{1+S} ab$ that contains the charge $-\frac{S}{1+S}$ (S stands for the overlap integral $(a|b)$) and has a uniform density along the bond. This is why its attraction with the nucleus b (of charge $+1$), which is in the immediate neighborhood, is strong. (b) In the classical picture the cloud $-\frac{S}{1+S}(1s_a)^2$, corresponding to the same total charge $-\frac{S}{1+S}$ as before, is close to the nucleus a and therefore its attraction with the nucleus b is much weaker. Thus, the atoms bind so strongly due to the quantum nature of the electron involved.

Note that the critically important charge density distribution ρ_{ab} has a very unusual shape. It represents something similar to an electron density rod (the figure shows its section only) that connects the two nuclei, its density being the largest and *constant* (!) within the section of the straight line connecting the two nuclei (an extremely although understandable striking shape[86]) and decaying very fast beyond.

8.7.2 Can we see a chemical bond?

If a substance forms crystals, it may be subjected to X-ray analysis. Such an analysis is quite exceptional, since it is one of very few techniques (which also include neutronography and

---

[86] For points along the section ($0 \le x \le R$) one has $\rho_{ab}(x,0,0) = \frac{1}{1+S} a(x,0,0) b(x,0,0) = \frac{1}{1+S}\frac{1}{\pi}\exp(-|x|)\exp(-|x-R|) = \frac{1}{1+S}\frac{1}{\pi}\exp(-x)\exp(x-R) = \frac{1}{1+S}\frac{1}{\pi}\exp(-x+x-R) = \frac{1}{1+S}\frac{1}{\pi}\exp(-R)$, which is a number that is independent of x! In Chapter V2-3 we will detect a chemical bond as a "rope" formed by the total electron density distribution (the Bader analysis), but the rod we find here as distinguishing the classical and the quantum situations represents only a part of this "rope" from Chapter V2-3.

nuclear magnetic resonance spectroscopy) which can show atomic positions in space. More precisely, X-ray analysis shows electronic density maps, because the radiation sees electrons, not nuclei. The inverse is true in neutronography. If we have the results of X-ray and neutron scattering, we can subtract the electron density of atoms (positions shown by neutron scattering) from the electron density of the molecular crystal (shown by X-ray scattering). This difference would be a consequence of the chemical bonding (and to a smaller extent of the intermolecular interactions). This method is called X−N, or X-Ray minus Neutron Diffraction.[87] Hence differential maps of the crystal are possible, where we can see the shape of the "additional" electron density at the chemical bond, or the shape of the electron deficit (negative density) in places where the interaction is antibonding.[88]

From the differential maps we can estimate (by comparison with standard substances) the following:

1) the strength of a chemical bond via the value of the positive electron density at the bond,
2) the deviation of the bond electron density (perpendicular intersection) from the cylindrical symmetry, which gives information on the π character of the chemical bond,
3) the shift of the maximum electron density towards one of the atoms which indicates the *polarization* of the bond, and
4) the shift of the maximum electron density away from the straight line connecting the two nuclei, which indicates bent (banana-like) bonding.

This opens up new possibilities for comparing theoretical calculations with experimental data.

8.8 Excitation energy, ionization potential, and electron affinity (RHF approach)

8.8.1 Approximate energies of electronic states

Let us consider (within the RHF scheme) the simplest two-electron closed shell system with both electrons occupying the same orbital φ_1. The Slater determinant, called ψ_G (G from the *ground state*) is built from two spin orbitals $\phi_1 = \varphi_1\alpha$ and $\phi_2 = \varphi_1\beta$. We also have the virtual orbital φ_2, corresponding to orbital energy ε_2, and we may form two other spin orbitals from it. We are now interested in the energies of *all the possible excited states* which can be formed from this pair of orbitals. These states will be represented as Slater determinants, built from φ_1 and φ_2 orbitals with the appropriate electron occupancy. We will also assume that excitations

---

[87] There is also a pure X-ray version of this method. It uses the fact that the X-ray reflections obtained at large scattering angles see only the spherically symmetric part of the atomic electron density, similarly to that which we obtain from neutron scattering.

[88] R. Boese, *Chemie in unserer Zeit*, 23(1989)77; D. Cremer, E. Kraka, *Angew. Chem.*, 96(1984)612.

do not deform the φ orbitals (which is, of course, only partially true). Now all possible states may be listed by occupation of the ε_1 and ε_2 orbital levels (see Table 8.2).

Table 8.2. All possible occupations of levels ε_1 and ε_2.

Level| function	ψ_G	ψ_T	$\psi_{T'}$	ψ_1	ψ_2	ψ_E
ε_2	–	α	β	β	α	$\alpha\beta$
ε_1	$\alpha\beta$	α	β	α	β	–

Here E is a doubly excited electronic state, and T and T' are two of three possible triplet states of the same energy. If we require that any state should be an eigenfunction of the $\hat{S}^2$ operator (it also needs to be an eigenfunction of $\hat{S}_z$, but this condition is fortunately fulfilled by all the functions listed above), it appears that only ψ_1 and ψ_2 are illegal. However, their combinations,

$$\psi_S = \frac{1}{\sqrt{2}}(\psi_1 - \psi_2), \tag{8.74}$$

$$\psi_{T''} = \frac{1}{\sqrt{2}}(\psi_1 + \psi_2), \tag{8.75}$$

are legal. The first describes the singlet state, and the second the triplet state (the third function missing from the complete triplet set).[89] This may be easily checked by inserting the spin orbitals into the determinants, expanding the determinants, and separating the spin part. For ψ_S, the spin part is typical for the singlet, $\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$; for T, T', and T'' the spin parts are, respectively, $\alpha(1)\alpha(2)$, $\beta(1)\beta(2)$, and $\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$. This is expected for triplet functions with components of total spin equal to 1, -1, and 0, respectively (Appendix R).

Now let us calculate the mean values of the Hamiltonian using the states mentioned above. Here we will use the Slater–Condon rules (see the diagram on p. 718), which when expressed in orbitals[90] produce in the MO representation

$$E_G = 2h_{11} + \mathcal{J}_{11}, \tag{8.76}$$

$$E_T = h_{11} + h_{22} + \mathcal{J}_{12} - \mathcal{K}_{12}, \tag{8.77}$$

(for all three components of the triplet)

$$E_S = h_{11} + h_{22} + \mathcal{J}_{12} + \mathcal{K}_{12}, \tag{8.78}$$

$$E_E = 2h_{22} + \mathcal{J}_{22}, \tag{8.79}$$

---

[89] Let us make a convention: in the Slater determinant $\frac{1}{\sqrt{2}} \det|\phi_1(1)\phi_2(2)|$, the spin orbitals are organized according to increasing orbital energy, because then the signs in Eqs. (8.74) and (8.75) are valid.

[90] For E_G the derivation of the final formula is given on p. 490 (E'_{RHF}). The other derivations are simpler.

where $h_{ii} = (\varphi_i|\hat{h}|\varphi_i)$, $\hat{h}$ is a one-electron operator (the same as that appearing in the Slater–Condon rules and *explicitly* shown on p. 470), and $\mathcal{J}_{ij}$ and $\mathcal{K}_{ij}$ are the two-electron integrals (Coulombic and exchange): $\mathcal{J}_{ij} = (ij|ij)$ and $\mathcal{K}_{ij} = (ij|ji)$.

The orbital energies of a molecule (calculated for the state with the doubly occupied φ_1 orbital) are

$$\varepsilon_i = (\varphi_i|\hat{\mathcal{F}}|\varphi_i) = (\varphi_i|\hat{h} + 2\hat{\mathcal{J}} - \hat{\mathcal{K}}|\varphi_i). \tag{8.80}$$

Thus, we get

$$\varepsilon_1 = h_{11} + \mathcal{J}_{11}, \tag{8.81}$$

$$\varepsilon_2 = h_{22} + 2\mathcal{J}_{12} - \mathcal{K}_{12}. \tag{8.82}$$

Now, the energies of the electronic states can be expressed in terms of orbital energies, i.e.,

$$E_G = 2\varepsilon_1 - \mathcal{J}_{11}, \tag{8.83}$$

$$E_T = \varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11} - \mathcal{J}_{12} \tag{8.84}$$

(for the ground singlet state and for the three triplet components of the common energy E_T). The distinguished role of φ_1 (in E_T) may be surprising (since the electrons reside on φ_1 and φ_2), but φ_1 is indeed distinguished, because the ε_i values are derived from the Hartree–Fock problem with the *only* occupied orbital φ_1. So we get

$$E_S = \varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11} - \mathcal{J}_{12} + 2\mathcal{K}_{12}, \tag{8.85}$$

$$E_E = 2\varepsilon_2 + \mathcal{J}_{22} - 4\mathcal{J}_{12} + 2\mathcal{K}_{12}. \tag{8.86}$$

Now it is time for conclusions.

8.8.2 Singlet or triplet excitation?

The Jabłoński diagram plays an important role in molecular spectroscopy (Fig. 8.26). It shows three energy levels: the ground state (G), the first excited singlet state (S), and the metastable in-between state. Later on researchers identified this metastable state as the lowest triplet (T).

Aleksander Jabłoński (1898–1980), Polish theoretical physicist, professor at the Stefan Batory University in Vilnius, then at the Nicolaus Copernicus University in Toruń, studied photoluminescence problems. Courtesy of Nicolaus Copernicus University, Poland.

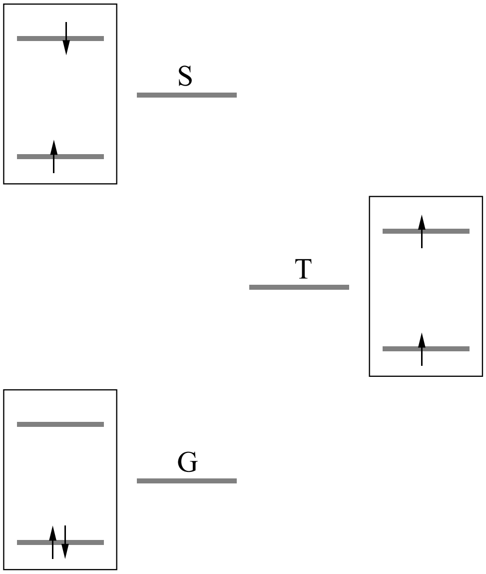


Fig. 8.26. The Jabłoński diagram. The ground state is G. The energy of the singlet excited state (S) is *higher* than the energy of the corresponding triplet state (T; that resulting from use of the same orbitals).

Let us compute the energy difference between the singlet and triplet states. We obtain

$$E_T - E_S = -2\mathcal{K}_{12} < 0. \tag{8.87}$$

This equation says that

> a molecule always has lower energy in the excited triplet state than in the excited singlet state (both states resulting from the use of the same orbitals),

because $\mathcal{K}_{12} = (\, \varphi_1(1)\varphi_2(2)|\frac{1}{r_{12}}|\varphi_2(1)\varphi_1(2)\,)$ is always positive, being the interaction of two identical charge distributions (interpretation of an integral, real functions assumed). This rule holds firmly for the energy of the two lowest singlet and triplet states.

8.8.3 Hund's rules

The difference between the energies of the ground and triplet states is

$$E_T - E_G = (\varepsilon_2 - \varepsilon_1) - \mathcal{J}_{12}. \tag{8.88}$$

This result has a simple interpretation. The excitation of a single electron (to the triplet state) costs some energy ($\varepsilon_2 - \varepsilon_1$), but (since $\mathcal{J}_{12} = (12|12) = \int \mathrm{d}V_1 \mathrm{d}V_2 |\varphi_1(1)|^2 \frac{1}{r_{12}} |\varphi_2(2)|^2 > 0$) there is also an energy gain ($-\mathcal{J}_{12}$) connected with the removal of the (mutually repulsing) electrons from the "common apartment" (orbital φ_1) to the two separate "apartments" (φ_1 and φ_2). Apartment φ_2 is admittedly on a higher floor ($\varepsilon_2 > \varepsilon_1$), but if $\varepsilon_2 - \varepsilon_1$ is small, then it may still pay to move.

In the limiting case, if $\varepsilon_2 - \varepsilon_1 = 0$, the system prefers to put electrons in separate orbitals and with the same spins (the empirical Hund rule, Fig. 8.27).

8.8.4 Hund's rules for the atomic terms

The question whether electron pairing is energetically favorable is most delicate for atoms. In the atomic case one has to do with quite a lot of possible electronic configurations ("occupancies") and the question is which of them better describes reality, i.e., which of them is of the lowest energy.

Friedrich Hermann Hund (1896–1997), professor of theoretical physics at the Universities in Rostock, Leipzig (1929–1946), Jena, Frankfurt am Main, and finally Göttingen, where in his youth he had worked with Born and Franck. He applied quantum theory to atoms, ions, and molecules and discovered his famous empirical rule in 1925. He is considered by many as one of the founders of quantum chemistry (biography in German: *Intern. J. Quantum Chem.*, S11(1977)6).

Hund discovered that this can be determined by using some empirical rules (now known as Hund's rules). They are the following:

- Hund's rule I: the lowest energy corresponds to that configuration which corresponds to the maximum of the spin angular momentum $|\mathbf{S}|$, where $|\mathbf{S}|^2 = S(S+1)\hbar^2$, where S may be either integer or half-integer. A large value of S requires a "same spin situation," which is only possible for different orbitals for valence electrons, i.e., a single occupation of the orbitals (low electronic repulsion).

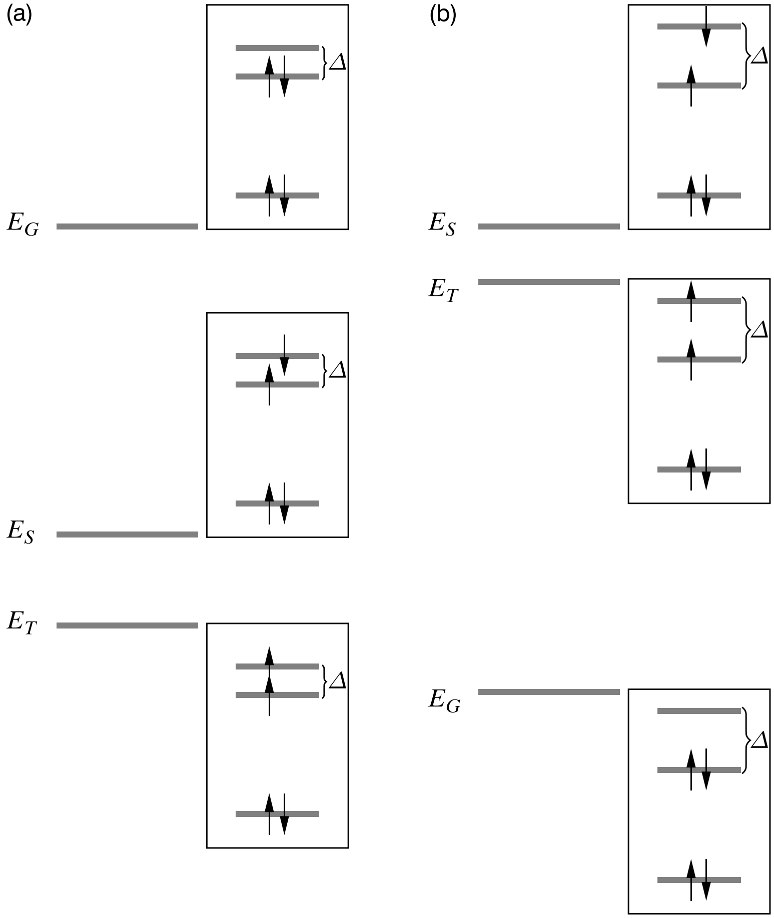


Fig. 8.27. Hund's rule. Energy of each configuration (left side of the picture) corresponds to an electron occupation of the orbital energy levels (shown in boxes). Two electrons of the highest orbital energy face a dilemma (two upper diagrams):

– is it better to occupy a common apartment on the lower floor (...but electrons do not like each other), or

– is it better for one of them (fortunately, they are not distinguishable...) to make a sacrifice and move to the upper-floor apartment (then they can avoid each other).

If the upper floor is not too high in the energy scale (small Δ, (a)), then the electrons prefer the second case: each of them occupies a separate apartment and they feel best having their spins parallel (triplet state). But when the upper-floor energy is very high (large Δ, (b)), then both electrons are forced to live in the same apartment, and in that case they are forced to have antiparallel spins.

Hund's rule pertains to the first case in its extreme form ($\Delta = 0$). When there are several orbitals of the same energy and there are many possibilities of their occupation, then the state with the lowest energy is such that the electrons go each to a separate orbital, and the alignment of their spins is "parallel" (see p. 37).

- Hund's rule II: if several such configurations come into play, the lowest energy corresponds to that one which has the largest orbital angular momentum $|\mathbf{L}|$, with $|\mathbf{L}|^2 = L(L+1)\hbar^2$, where $L = 0, 1, 2, \ldots$.
- Hund's rule III is a relativistic correction to the first two rules, introducing a splitting of the terms given by the previous rules. The energy operator (Hamiltonian) commutes with the square of the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$ and therefore the energy levels depend rather on the *total* momentum $|\mathbf{J}|^2 = J(J+1)\hbar^2$. This means that they depend on the mutual orientation of $\mathbf{L}$ and $\mathbf{S}$ (this is a relativistic effect due to the spin–orbit coupling in the Hamiltonian). The vectors $\mathbf{L}$ and $\mathbf{S}$ add in quantum mechanics in a specific way (see Chapter 7, p. 404): one has $J = |L+S|, |L+S-1|, \ldots |L-S|$. *Hund's rule III says that if the shell is less than half-filled, the lowest energy corresponds to* $J = |L-S|$, *while if it is more than half-filled it corresponds to* $J = |L+S|$.

Hund's rules not only allow one to identify the lowest energy level (or term $^{2S+1}[L]_J$, where we use the symbol $[L] = S$ (for $L = 0$), P (for $L = 1$), D (for $L = 2$), ...), but also enable one to give their sequence on the energy scale.

Example 5 (Closed shells – a neon atom). We write first the neon electronic configuration, $1s^2 2s^2 2p^6$, with a convention that we choose as AOs those which represent the eigenfunctions of the $\hat{L}_z$ operator (in our case $2p_0, 2p_1, 2p_{-1}$, not $2p_x, 2p_y, 2p_z$). Hund's rules are about the orbital angular momentum and spin angular momentum. Let us calculate the z component of the total *orbital* angular momentum as a sum of the corresponding components for individual electrons: $L_z = 2\cdot 0\hbar + 2\cdot 0\hbar + 2\cdot(-1)\hbar + 2\cdot 0\hbar + +2\cdot(+1)\hbar = 0$. The only possibility for the total orbital angular momentum to have *only* this value of the z component is $L = 0$. The term has to have therefore the symbol $[L] = S$ (I am very sorry, this is a heritage of a misleading quantum tradition: we cannot mix this symbol with the spin quantum number S!). Now about the z component of the *spin* angular momentum as a sum of contributions of the individual electrons: $S_z = (\frac{1}{2} - \frac{1}{2})\hbar + (\frac{1}{2} - \frac{1}{2})\hbar + 3\cdot(\frac{1}{2} - \frac{1}{2})\hbar = 0$. This can happen only if $S = 0$ (this time S means spin!). The multiplicity of the term is $2S + 1 = 1$, and therefore the term symbol $^{2S+1}[L]$ becomes in our case ^1S (we say "singlet S"). Since $\mathbf{L} = \mathbf{0}$ and $\mathbf{S} = \mathbf{0}$ they may add only to $\mathbf{J} = \mathbf{0}$, which means $J = 0$. Therefore, the full term symbol should be $^1S_0$, but to avoid the banality we always write it as ^1S. This term is the only term possible for the electronic configuration $1s^2 2s^2 2p^6$ as well as for any other closed electronic shell.

Any closed shell gives zero contribution to the orbital angular momentum $\mathbf{L}$ and to the spin angular momentum $\mathbf{S}$. This makes it possible, when determining the terms, just to ignore any closed shell contribution.

Example 6 (The lowest-energy term for the carbon atom). The carbon atom electronic configuration reads as $1s^2 2s^2 2p^2$. Since the closed shells $1s^2 2s^2$ do not contribute anything, we just ignore them. What matters is the configuration $2p^2$. Three orbitals $2p$ allow for six spin orbitals, occupied by two electrons only. The number of possible Slater determinants is therefore $\binom{6}{2} = 15$. The lowest-energy Slater determinant will be determined by using Hund's rules to find the lowest-energy term. To this end we write down a box diagram of orbital occupations with the convention that an up or down arrow stands for an electron with the α or β spin function, respectively. The boxes in the diagram (each one corresponding to an atomic orbital with quantum numbers n, l, m) begin by the one with maximum value of m and continue in descending order. Then we begin the occupation of the boxes by the electrons (maximum two electrons – of opposite spins – per box). Our aim is to find the lowest-energy term by using Hund's rules. We begin by Hund's rule I and this means maximization of S (same spins). To get this we place electrons in separate boxes, but to satisfy also Hund's rule II we begin to do this from the left-hand side of the diagram:

$$\begin{array}{ccc} 2p_1 & 2p_0 & 2p_{-1} \\ \uparrow & \uparrow & \end{array}.$$

We get $M_S = \frac{1}{2} + \frac{1}{2} = 1$ (such a projection implies $S = 1$) and $M_L = 1 + 0 = 1$ (this projection implies $L = 1$).

Hence, the term indicated by Hund's first two rules that explains these projections is ${}^3P$. This is what the nonrelativistic approach gives as the ground state.

However, the relativistic effects will split this triple-degenerate energy level into three separate levels according to three possible mutual orientations of the vectors **L** and **S**, leading to $J = |L + S|, \ldots J = |L - S|$, i.e., to $J = 2, 1, 0$. Therefore, we get three terms: ${}^3P_2, {}^3P_1, {}^3P_0$, and Hund's rule III predicts the following energy sequence:

$${}^3P_0 <{}^3 P_1 <{}^3 P_2.$$

This sequence is valid for all atoms having the $2p^2$ configuration: C, Si, Ge, Sn, Pb, with increasing splitting in this series (from about 43 cm^{-1} for C to about 10 650 cm^{-1} for Pb[91]).

8.8.5 Ionization potential and electron affinity (Koopmans' theorem)

The ionization potential of the molecule M is defined as the minimum energy needed for an electron to detach from the molecule. The electron affinity energy of the molecule M is defined

---

[91] This reminds us about a general rule: the heavier the atom, the more important relativistic effects.

as the minimum energy for an electron detachment from M^-. Let us naively assume again that *during these operations the MOs and the orbital energies do not undergo any changes*. In fact, of course, everything changes, and the computations should be repeated for each system separately (the same applies in the previous section for excitations).

In our two-electron system, which is a model of any closed shell molecule, the electron removal leaves the molecule with one electron only, and its energy has to be

$$E_+ = h_{11}. \tag{8.89}$$

However,

$$h_{11} = \varepsilon_1 - \mathcal{J}_{11}. \tag{8.90}$$

This formula looks like trouble! After the ionization there is only a single electron in the molecule, while here some electron–electron repulsion (integral $\mathcal{J}$) appears! But everything is fine, because we still use the two-electron problem as a reference, and ε_1 relates to the two-electron problem, in which $\varepsilon_1 = h_{11} + \mathcal{J}_{11}$.

Hence,

IONIZATION ENERGY

The ionization energy is equal to the negative of the orbital energy of an electron, i.e.,

$$E_+ - E_G = -\varepsilon_1. \tag{8.91}$$

To calculate the electron affinity energy we need to consider a determinant as large as 3 × 3, but this proves easy if the useful Slater–Condon rules (Appendix N) are applied. Rule I gives (we write everything using the ROHF spin orbitals, note that the three spin orbitals are derived from two orbitals, and then sum over the spin variables)

$$E_- = 2h_{11} + h_{22} + \mathcal{J}_{11} + 2\mathcal{J}_{12} - \mathcal{K}_{12}, \tag{8.92}$$

and introducing the orbital energies we get

$$E_- = 2\varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11}, \tag{8.93}$$

which gives

$$E_- - E_G = \varepsilon_2. \tag{8.94}$$

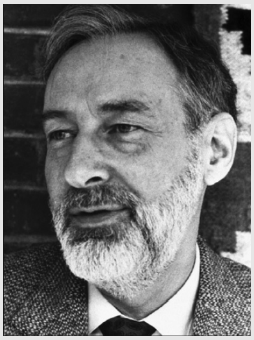
Hence,

ELECTRON AFFINITY

The electron affinity is the difference of the energies of the system without an additional electron and that of the anion, $E_G - E_- = -\varepsilon_2$. It is approximately equal to the negative energy of the virtual orbital on which the electron lands (if $\varepsilon_2 < 0$ attaching an electron means energy lowering).

A comment on Koopmans' theorem

Tjalling Charles Koopmans (1910–1985), American econometrist of Dutch origin, professor at Yale University (USA), introduced mathematical procedures of linear programming to economics and received the Nobel Prize in 1975 "*for work on the theory of optimum allocation of resources*."



The MO theory represents a rough approximation to reality. So is Koopmans' theorem, which proves to be poorly satisfied for most molecules. But these approximations are often used for practical purposes. This is illustrated by a certain quantitative relationship, derived by Grochala et al.[92]

The authors noted that a very simple relationship holds surprisingly well for the equilibrium *bond lengths* R of four objects (the ground state M_0 of the closed shell molecule, its excited triplet state M_T, its radical cation $M^{+\cdot}$, and its radical anion $M^{-\cdot}$):

$$R(M_T) = R(M^{-\cdot}) + R(M^{+\cdot}) - R(M_0). \tag{8.95}$$

The cyclobutadiene in the triplet state has square symmetry (the two adjacent CC bonds of equal length), while in the ground state the molecule is rectangular. Therefore, if one takes Eq. (8.95) for the first CC bond and then for the second one, we should get the same length in the triplet state (all values calculated by the DFT method, see Chapter V2-3). Let us insert the bond lengths for the first CC bond. We find

$$1.378\ \text{Å} + 1.363\ \text{Å} - 1.318\ \text{Å} = 1.423\ \text{Å},$$

while for the second CC bond

$$1.501\ \text{Å} + 1.489\ \text{Å} - 1.565\ \text{Å} = 1.425\ \text{Å},$$

---

[92] W. Grochala, A.C. Albrecht, R. Hoffmann, *J. Phys. Chem. A*, 104(2000)2195.

while the calculated CC bond length for the triplet state is equal to 1.426 Å, i.e., both values are very close to what has been calculated independently.

The above relationship is similar to that pertaining to the corresponding energies,

$$E(M_T) = E(M^{-\cdot}) + E(M^{+\cdot}) - E(M_0),$$

which may be deduced, basing on certain approximations, from Koopmans' theorem,[93] or from the Schrödinger equation while neglecting the two-electron operators (i.e., Coulomb and exchange). The difference between these two expressions is substantial: the latter holds for the four species at *the same* nuclear geometry, while the former describes the geometry *changes* for the "relaxed" species.[94] The first equation proved to be satisfied for a variety of molecules: ethylene, cyclobutadiene, divinylbenzene, diphenylacetylene, *trans*-N_2H_2, CO, CN^-, N_2, and NO^+. It is not yet clear if it would hold beyond the one-electron approximation or for experimental bond lengths (these are usually missing, especially for polyatomic molecules).

8.9 Towards a chemical picture – localization of molecular orbitals

The canonical MOs derived from the RHF method are usually delocalized over the whole molecule, i.e., their amplitudes are in general nonzero for all atoms in the molecule. This applies, however, mainly to high-energy MOs, which exhibit a similar AO amplitude for most atoms. Yet the canonical MOs of the inner shells are usually very well localized. The canonical MOs are occupied, as usual, by putting two electrons on each low-lying orbital (the Pauli exclusion principle).

> The picture obtained contrasts with chemical intuition, which indicates – chemists believe that the electron pairs are localized within the chemical bonds, free electron pairs, and inner atomic shells. The picture which agrees with intuition may be obtained after the localization of the MOs.

The localization is based on making new orbitals to be linear combinations of the canonical MOs, a fully legal procedure (see p. 475). Then, the determinantal wave function, as shown on

---

[93] Let us check it using the formulae derived by us: $E(M_T) = \varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11} - \mathcal{J}_{12}$, and $E(M^{-\cdot}) + E(M^{+\cdot}) - E(M_0) = \left[2\varepsilon_1 + \varepsilon_2 - \mathcal{J}_{11}\right] + \left[\varepsilon_1 - \mathcal{J}_{11}\right] - \left[2\varepsilon_1 - \mathcal{J}_{11}\right] = \varepsilon_2 + \varepsilon_1 - \mathcal{J}_{11}$. Equality is obtained after neglecting $\mathcal{J}_{12}$ as compared to $\mathcal{J}_{11}$.

[94] If we assume that a geometry change in these states induces an energy increase proportional to the square of the change, and that the curvature of all these parabolas is identical, then the above relationship would be easily proved. The problem is that these states have significantly different force constants, and the curvature of parabolas strongly varies among them.

p. 475, expressed in the new spin orbitals, takes the form $\psi' = (det\mathbf{A})\psi$, and the total energy will remain unchanged. If linear transformation applied ($\mathbf{A}$) is an orthogonal transformation, i.e., $\mathbf{A}^T\mathbf{A} = \mathbf{1}$, or a unitary one, $\mathbf{A}^\dagger\mathbf{A} = \mathbf{1}$, the new MOs preserve orthonormality (like the canonical ones), as shown on p. 477. We emphasize that we can make *any nonsingular*[95] *linear transformation* $\mathbf{A}$, not only orthogonal or unitary ones. This means something important, namely,

> the solution in the Hartree–Fock method depends on the *space* spanned by the occupied orbitals (i.e., on the set of all linear combinations which can be formed from the occupied MOs), and not on some particular set of the MOs. The new orbitals do not satisfy the Fock equations (8.32); these are satisfied by canonical orbitals only.

The localized orbitals (being some other orthonormal basis set in the space spanned by the canonical orbitals) satisfy the Fock equation (8.19) with the off-diagonal Lagrange multipliers.

8.9.1 Can a chemical bond be defined in a polyatomic molecule?

Unfortunately, the view to which chemists get used, i.e., the chemical bonds between pairs of atoms, lone electron pairs, and inner shells, can be derived in an infinite number of ways (because of the arbitrariness of transformation $\mathbf{A}$), and in each case the effects of localization vary. Hence,

> we cannot uniquely define the chemical bond in a polyatomic molecule.

It is not a drama, however, because what really matters is the probability density, i.e., the square of the complex modulus of the *total* many-electron wave function and this stays constant for an orthogonal (or unitary) transformation. The concept of the (localized or delocalized) MOs represents simply an attempt to divide this total density into various spatially separated although overlapping parts, each belonging to a single MO. It is similar to dividing an apple into N equal parts. The freedom of such a division is unlimited. For example, we could envisage that each part would have the dimension of the apple ("delocalized orbitals"), or an apple would be simply cut axially, horizontally, concentrically, etc., into N equal parts, forming an analog of the localized orbitals. Yet each time the full apple could be reconstructed from these parts.

As we will soon convince ourselves, the problem of defining a chemical bond in a polyatomic molecule is not so hopeless as it looks now, because various methods lead to essentially the same results.

---

[95] For any singular matrix $\det\mathbf{A} = 0$, and this should not be allowed (p. 475).

Now let us consider some practical methods of localization. There are two categories of these: internal and external.[96] In the external localization methods we plan where the future MOs will be localized, and the localization procedure only slightly alters our plans. This is in contrast with the internal methods, where certain general conditions are imposed that induce automatically localization of the orbitals.

8.9.2 The external localization methods

Projection method

This is an amazing method,[97] in which *we* first construct some *arbitrary*[98] (but linearly independent[99]) orbitals χ_i of the bonds, lone pairs, and inner shells, the total number of these being equal to the number of the occupied MOs. Now let us project them on the space of the occupied RHF MOs $\{\varphi_j\}$ using the projection operator $\hat{P}$. We obtain

$$\hat{P}\chi_i \equiv \left(\sum_j^{MO} |\varphi_j\rangle\langle\varphi_j|\right)\chi_i. \tag{8.96}$$

The projection operator is used to create the new orbitals

$$\varphi_i' = \sum_j^{MO} \langle\varphi_j|\chi_i\rangle\varphi_j. \tag{8.97}$$

The new orbitals φ_i', as linearly independent combinations of the occupied canonical orbitals φ_j, span the space of the canonical occupied Hartree–Fock orbitals $\{\varphi_j\}$. They are in general nonorthogonal, but we may orthogonalize them by applying the Löwdin orthogonalization procedure (symmetric orthogonalization, see Appendix 697).

Do the final localized orbitals depend on the starting χ_i in the projection method? The answer[100] can be found in Table 8.3. The influence is small.

---

[96] Like medicines.

[97] A. Meunier, B. Levy, G. Berthier, *Theoret. Chim. Acta*, 29(1973)49.

[98] This is the beauty of the projection method.

[99] A linear dependence cannot be allowed. If this happens then we need to change the set of functions χ_i.

[100] B. Lévy, P. Millié, J. Ridard, J. Vinh, *J. Electr. Spectr.*, 4(1974)13.

Table 8.3. Influence of the initial approximation (the LCAO coefficients for the CH_3F molecule, first four columns) on the final localized MOs in the projection method of localization.

Function χ for the C–F bond				The localized orbital of the C–F bond				
$2s(C)$	$2p(C)$	$2s(F)$	$2p(F)$	$2s(C)$	$2p(C)$	$2s(F)$	$2p(F)$	$1s(H)$
0.300	0.536	0.000	−0.615	0.410	0.496	−0.123	−0.654	−0.079
0.285	0.510	0.000	−0.643	0.410	0.496	−0.131	−0.655	−0.079
0.272	0.487	0.000	−0.669	0.410	0.496	−0.138	−0.656	−0.079
0.260	0.464	0.000	−0.692	0.410	0.496	−0.144	−0.656	−0.079
0.237	0.425	0.000	−0.730	0.410	0.496	−0.156	−0.658	−0.079

8.9.3 The internal localization methods

Ruedenberg method: the maximum interaction energy of the electrons occupying an MO

The basic concept of this method was given by Lennard-Jones and Pople,[101] and applied by Edmiston and Ruedenberg.[102] It may be easily shown that for a given geometry of the molecule the functional $\sum_{i,j=1}^{MO} \mathcal{J}_{ij}$ is invariant with respect to *any* unitary transformation of the orbitals, i.e.,

$$\sum_{i,j=1}^{MO} \mathcal{J}_{ij} = const. \tag{8.98}$$

The proof is very simple and similar to the one on p. 477, where we derived the invariance of the Coulombic and exchange operators in the Hartree–Fock method.

This further implies

maximization of $\sum_{i=1}^{MO} \mathcal{J}_{ii}$,

which is the very essence of the localization criterion, and it is equivalent to the *minimization* of the off-diagonal elements

$$\sum_{i<j}^{MO} \mathcal{J}_{ij}. \tag{8.99}$$

This means that to localize the MOs we try to put them as far apart as possible in space, because then their repulsion will be least.

---

[101] J.E. Lennard-Jones, J.A. Pople, *Proc. Roy. Soc. (London) A*, 202(1950)166.

[102] C. Edmiston, K. Ruedenberg, *Rev. Modern Phys.*, 34(1962)457.

Similarly, we can prove another invariance,

$$\sum_{i,j=1}^{MO} \mathcal{K}_{ij} = const'. \tag{8.100}$$

It may be also expressed in another way, given that $\sum_{i,j}^{MO} \mathcal{K}_{ij} = const' = \sum_{i}^{MO} \mathcal{K}_{ii} + 2\sum_{i<j}^{MO} \mathcal{K}_{ij} = \sum_{i}^{MO} \mathcal{J}_{ii} + 2\sum_{i<,j}^{MO} \mathcal{K}_{ij}$. Since we maximize the $\sum_{i}^{MO} \mathcal{J}_{ii}$, simultaneously

we minimize the sum of the exchange contributions

$$\sum_{i<j}^{MO} \mathcal{K}_{ij}. \tag{8.101}$$

Boys method: the minimum distance between electrons occupying an MO

In this method[103] we minimize the functional[104]

$$\sum_{i}^{MO} (ii|r_{12}^2|ii), \tag{8.102}$$

where the symbol $(ii|r_{12}^2|ii)$ denotes an integral similar to $\mathcal{J}_{ii} = (ii|ii)$, but instead of the $1/r_{12}$ operator, we have r_{12}^2. Functional (8.102) is invariant with respect to any unitary transformation of the MOs.[105] Since the integral $(ii|r_{12}^2|ii)$ represents the definition of the mean square of the distance between two electrons described by $\varphi_i(1)\varphi_i(2)$, the Boys criterion means that we try to obtain the localized orbitals as small as possible (small orbital *dimensions)*, i.e., localized in some small volume in space.[106] The detailed technique of localization will be given in a moment.

---

[103] S.F. Boys, in *Quantum Theory of Atoms, Molecules and the Solid State*, P.O. Löwdin, ed., Acad. Press, New York, 1966, p. 253.

[104] Minimization of the interelectronic distance (Boys method) is in fact similar in concept to the maximization of the Coulombic interaction of two electrons in the same orbital (Ruedenberg method).

[105] To prove this we need to represent the orbitals as components of a vector, the double sum as two scalar products of such vectors, transform the orbitals, and then show that the matrix transformation in the integrand results in a unit matrix.

[106] The integrals (8.102) are trivial. Indeed, using Pythagoras' theorem, we finally get simple one-electron integrals of the following type:

$$\begin{aligned}\big(i(1)i(2)|(x_2-x_1)^2|i(1)i(2)\big) &= \big(i(2)|x_2^2|i(2)\big) + \big(i(1)|x_1^2|i(1)\big) - 2\big(i(1)|x_1|i(1)\big)\big(i(2)|x_2|i(2)\big)\\ &= 2\big(i|x^2|i\big) - 2\big(i|x|i\big)^2.\end{aligned}$$

8.9.4 Examples of localization

Despite the freedom of the localization criterion choice, the results are usually similar. The orbitals of the C–C and C–H bonds in ethane, obtained by various approaches, are shown in Table 8.4.

Table 8.4. The LCAO coefficients of the localized orbitals of ethane in the antiperiplanar conformation (P. Millié, B. Lévy, G. Berthier, in: "*Localization and Delocalization in Quantum Chemistry*," eds. O. Chalvet, R. Daudel, S. Diner, J.P. Malrieu, Reidel Publish. Co., Dordrecht (1975)). Only the nonequivalent AOs have been shown in the table (four significant digits) for the C–C and one of the equivalent C–H bonds (with the proton H(1), Fig. 8.28). The z axis is along the C–C' bond. The localized MOs corresponding to the carbon inner shells $1s$ are not listed.

	The projection method	**Minimum distance method**	**Maximum repulsion energy**
	C–C' bond		
$1s$ (C)	−0.0494	−0.1010	−0.0476
$2s$ (C)	0.3446	0.3520	0.3505
$2p_z$ (C)	0.4797	0.4752	0.4750
$1s$ (H)	−0.0759	−0.0727	−0.0735
	C–H bond		
$1s$ (C)	−0.0513	−0.1024	−0.0485
$2s$ (C)	0.3397	0.3373	0.3371
$2p_z$ (C)	−0.1676	−0.1714	−0.1709
$2p_x$ (C)	0.4715	0.4715	0.4715
$1s$ (C')	0.0073	0.0081	0.0044
$2s$ (C')	−0.0521	−0.0544	−0.054
$2p_z$ (C')	−0.0472	−0.0503	−0.0507
$2p_x$ (C')	−0.0082	−0.0082	−0.0082
$1s$ (H1)	0.5383	0.5395	0.5387
$1s$ (H2)	−0.0942	−0.0930	−0.0938
$1s$ (H3)	−0.0942	−0.0930	−0.0938
$1s$ (H4)	0.0580	0.0584	0.0586
$1s$ (H5)	−0.0340	−0.0336	−0.0344
$1s$ (H6)	−0.0340	−0.0336	−0.0344

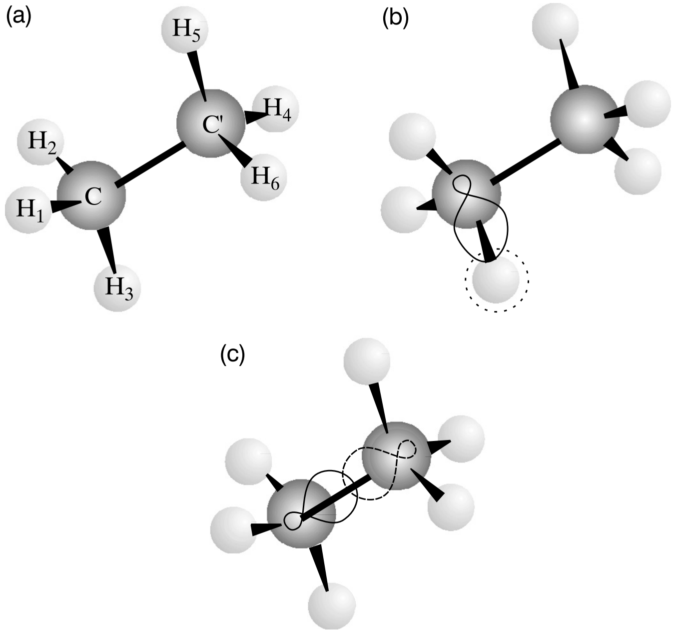


Fig. 8.28. (a) The ethane molecule in the antiperiplanar configuration. (b) The localized orbital of the C–H bond. (c) The localized orbital of the C–C bond. The carbon atom hybrid forming the C–H bond is quite similar to the hybrid forming the C–C bond.

Let us try to understand the table. First note the similarity of the results of various localization methods. The methods are different, the starting points are different, and yet we get almost the same in the end. It is both striking and important that

> the results of various localizations are similar to one another, and in practical terms (not theoretically) we can speak of the unique definition of a chemical bond in a polyatomic molecule.

Nobody would reject the statement that a human body is composed of the head, the hands, the legs, etc. Yet a purist (i.e., theoretician) might get into troubles defining, e.g., a hand (where does it end?). Therefore, purists would claim that it is impossible to define a hand, and as a consequence there is no such a thing as a hand – it simply does not exist. This situation is quite similar to the definition of the chemical bonding between two atoms in a polyatomic molecule.

Note that the localized orbitals are concentrated mainly in *one* particular bond between two atoms. For example, in the C–C bond orbital, the coefficients at the $1s$ orbitals of the hydrogen atom are small (-0.08), i.e., what really counts belongs to the carbon atoms. Similarly, the

$2s$ and $2p$ orbitals of *one* carbon atom and the $1s$ orbital of *one* (the closest) of the hydrogen atoms dominate the C–H bond orbital. Of course, localization is never complete. The oscillating "tails" of the localized orbital may be found even at distant atoms. They ensure the mutual orthogonality of the localized orbitals.

8.9.5 Localization in practice – computational technique

Let us take as an example the maximization of the electron interaction within all the orbitals (Ruedenberg method). We have

$$I = \sum_i^{MO} \mathcal{J}_{ii} = \sum_i^{MO} (ii|ii). \tag{8.103}$$

Suppose we want to make an orthogonal transformation (i.e., a rotation in the Hilbert space, Appendix B) of – so far only two – orbitals,[107] $|i\rangle$ and $|j\rangle$, in order to maximize I. The rotation (an orthogonal transformation which preserves the orthonormality of the orbitals) can be written as

$$\begin{aligned} \left|i'(\vartheta)\right\rangle &= |i\rangle \cos\vartheta + |j\rangle \sin\vartheta, \\ \left|j'(\vartheta)\right\rangle &= -|i\rangle \sin\vartheta + |j\rangle \cos\vartheta, \end{aligned}$$

where ϑ is an angle measuring the rotation (we are going to find the optimum angle ϑ). The contribution from the changed orbitals to I is

$$I(\vartheta) = (i'i'|i'i') + (j'j'|j'j'). \tag{8.104}$$

Then[108]

$$I(\vartheta) = I(0)\left(1 - \frac{1}{2}\sin^2 2\vartheta\right) + \big(2(ii|jj) + (ij|ij)\big)\sin^2 2\vartheta + \big((ii|ij) - (jj|ij)\big)\sin 4\vartheta, \tag{8.105}$$

where $I(0) = \left(ii|ii\right) + \left(jj|jj\right)$ is the contribution of the orbitals before their rotation.

Requesting that $\frac{\mathrm{d}I(\vartheta)}{\mathrm{d}\vartheta} = 0$, we easily get the condition for optimum $\vartheta = \vartheta_{opt}$:

$$\begin{aligned} &-2I(0)\sin 2\vartheta_{opt}\cos 2\vartheta_{opt} + \big(2(ii|jj) + (ij|ij)\big)4\sin 2\vartheta_{opt}\cos 2\vartheta_{opt} \\ &\quad + \big((ii|ij) - (jj|ij)\big)4\cos 4\vartheta_{opt} = 0, \end{aligned} \tag{8.106}$$

---

[107] The procedure is an iterative one. First we rotate one pair of orbitals, then we choose another pair and make another rotation, etc., until the next rotations do not introduce anything new.

[108] Derivation of this formula is simple and takes one page.

and hence

$$\mathrm{tg}(4\vartheta_{opt}) = 2\frac{(ij|jj) - (ii|ij)}{2(ii|jj) + (ij|ij) - \frac{1}{2}I(0)}. \tag{8.107}$$

The operation described here needs to be performed for all pairs of orbitals, and then repeated (iterations) until the numerator vanishes for each pair, i.e.,

$$(ij|jj) - (ii|ij) = 0. \tag{8.108}$$

The value of the numerator for each pair of orbitals is thus the criterion for whether a rotation is necessary for this pair or not. The matrix of the full orthogonal transformation represents the product of the matrices of these successive rotations.

The same technique of successive 2 × 2 rotations applies to other localization criteria.

8.9.6 The chemical bonds of σ, π, δ symmetry

Localization of the MOs leads to the orbitals corresponding to chemical bonds (as well as lone pairs and inner shells). In the case of a bond orbital, a given localized MO is in practice dominated by the AOs of *two atoms only*, i.e., those which create the bond.[109] According to the discussion on p. 511, the larger the overlap integral of the AOs, the stronger the bonding. The energy of a molecule is most effectively decreased if the AOs are oriented in such a way as to maximize their overlap integral (Fig. 8.29). We will now analyze the kind and the mutual orientation of these AOs.

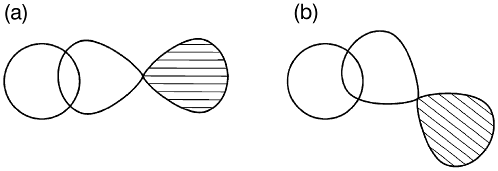


Fig. 8.29. Maximization of the AO overlap requests position (a), while position (b) is less preferred.

As shown in Fig. 8.30, the orbitals σ, π, δ (either canonical or not) have the following features:

- the σ-type orbital has cylindrical symmetry, and *has no nodal plane containing the bond axis*,

---

[109] That is, they have the largest absolute values of LCAO coefficients.

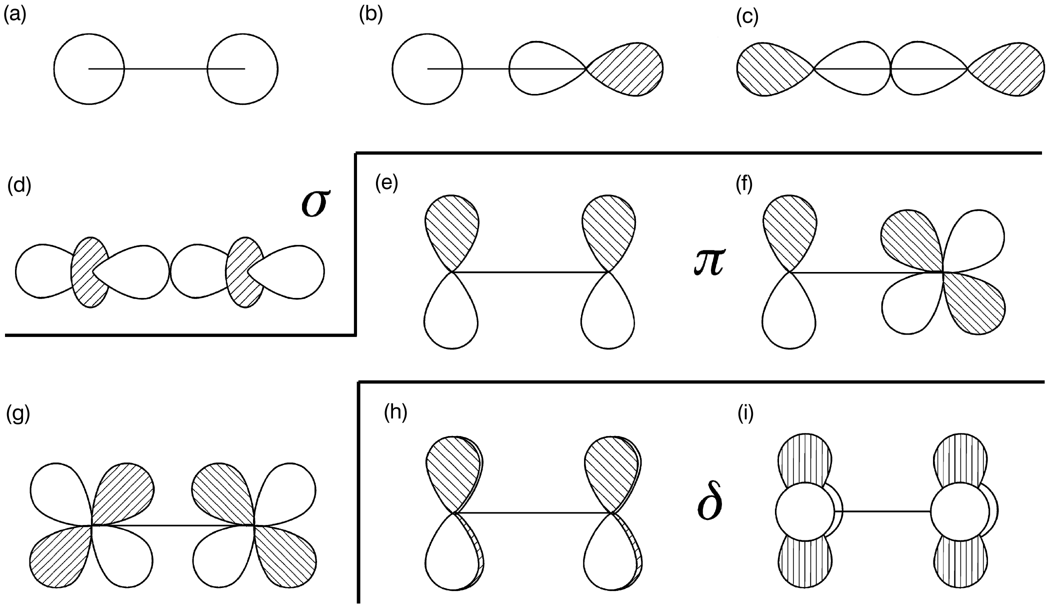


Fig. 8.30. Symmetry of the MOs results from the mutual arrangement of those AOs of both atoms which have the largest LCAO coefficients. (a)–(d) The σ-type bonds. (e)–(g) The π-type bonds. (h, i) The δ-type bonds. The σ bond orbitals have no nodal plane (containing the nuclei), the π orbitals have one such plane, the δ ones have two such planes.
If the z axis is set as the bond axis and the x axis is set in the plane of the figure, then cases (b)–(i) correspond (compare with Chapter 4) to the overlap of the following AOs: (b) s with p_z, (c) p_z with p_z, (d) $3d_{3z^2-r^2}$ with $3d_{3z^2-r^2}$, (e) p_x with p_x, (f) p_x with $3d_{xz}$, (g) $3d_{xz}$-$3d_{xz}$, (h) $3d_{xy}$ with $3d_{xy}$, and (i) $3d_{x^2-y^2}$ with $3d_{x^2-y^2}$. The figures show such AOs which correspond to the *bonding* MOs. To get the corresponding *antibonding* MOs, we need to change the sign of *one* of the two AOs.

- the π-type orbital has *one such a nodal plane*,
- the δ-type orbital has *two such nodal planes.*

If an MO is antibonding, then a little star (tradition) is added to its symbol, e.g., σ^*, π^*, etc. Usually we also give the orbital quantum number (in order of increasing energy), e.g., $1\sigma, 2\sigma$, ..., etc. For homonuclear diatomics additional notation is used (Fig. 8.31) showing the main AOs participating in the MO, e.g., (normalization omitted) $\sigma 1s = 1s_a + 1s_b$, $\sigma^* 1s = 1s_a - 1s_b, \sigma 2s = 2s_a + 2s_b, \sigma^* 2s = 2s_a - 2s_b$, etc.

The very fact that the π and δ MOs have value zero at the positions of the nuclei (the region most important for lowering the potential energy of electrons) suggests that they are bound to be of higher energy than the σ ones, and they are indeed.

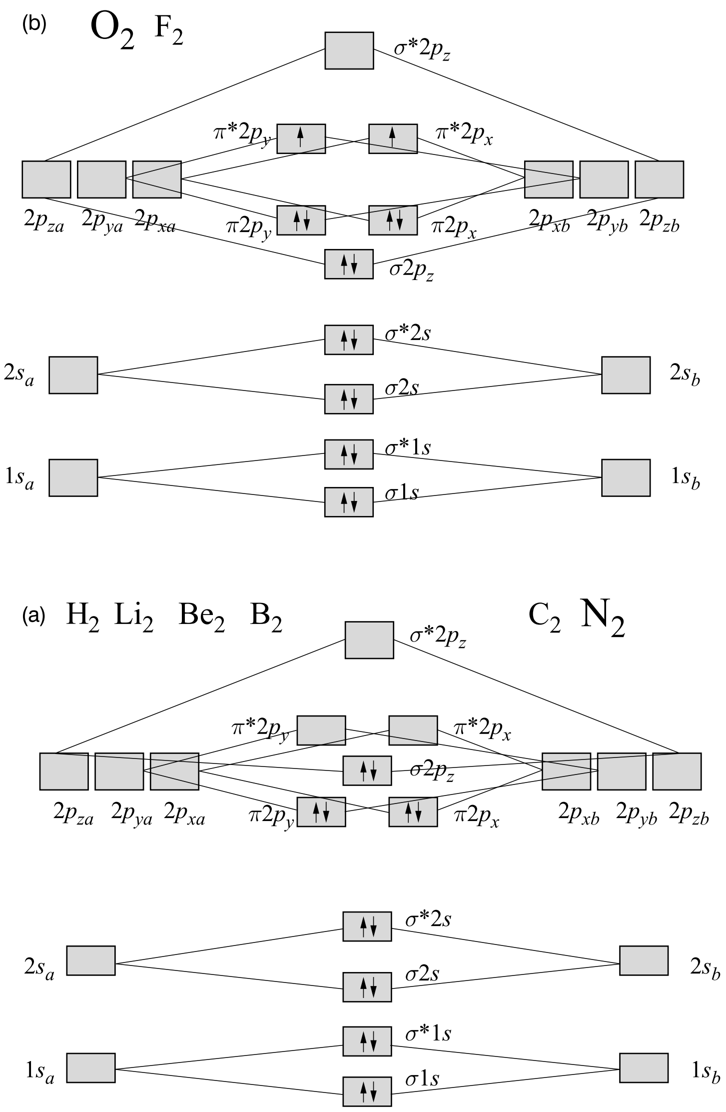


Fig. 8.31. Scheme of the bonding and antibonding MOs in homonuclear diatomics from H_2 through F_2. This scheme is better understood after we recall the rules of effective mixing of AOs (p. 503). All the orbital energies become lower in this series (due to increasing of the nuclear charge), but lowering of the bonding π orbitals leads to changing the *order* of the orbital energies, when going from N_2 to F_2. This is why we get two sequences of orbital energies (schematically) for the molecules (a) from H_2 through N_2 (the occupancy is shown for N_2) and (b) for O_2 and F_2 (the occupancy for O_2).

8.9.7 Electron pair dimensions and the foundations of chemistry

What are the dimensions of the electron pairs described by the localized MOs? Well, but how to define such dimensions? All orbitals extend to infinity, so you cannot measure them using a ruler, but some may be more diffuse than others. It also depends on the molecule itself, the role of a given MO in the molecular electronic structure (the bonding orbital, a lone electron pair, or the inner shell), the influence of neighboring atoms, etc. These are fascinating problems, and the issue is at the heart of structural studies of chemistry.

Several concepts may be given to calculate the dimensions of the molecular orbitals mentioned above. For example, we may take the integrals $(ii|r_{12}^2|ii) \equiv \langle r^2 \rangle$ calculated within the Boys localization procedure and use them to estimate the square of the dimension of the (normalized) MO φ_i. Indeed, $\langle r^2 \rangle$ is the mean value of the interelectronic distance for a two-electron state $\varphi_i(1)\varphi_i(2)$, and $\rho_i(\text{Boys}) = \sqrt{\langle r^2 \rangle}$ may be viewed as a measure of the φ_i orbital dimension. Or, we may do a similar thing by the Ruedenberg method, by noting that the Coulombic integral $\mathcal{J}_{ii}$, calculated in a.u., is nothing more than the mean value of the *inverse of the distance* between two electrons described by the φ_i orbital. In this case, the dimension of the φ_i orbital may be proposed as $\rho_i(\text{Ruedenberg}) = \frac{1}{\mathcal{J}_{ii}}$. Below, the calculations are reported, in which the concept of $\rho_i(\text{Boys})$ is used.

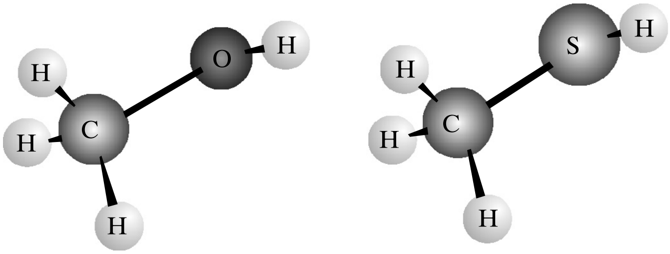


Fig. 8.32. Methanol (CH_3OH) and mercaptan (CH_3SH).

Let us compare the results for CH_3OH and CH_3SH (Fig. 8.32) in order to see what makes these two molecules so different (Table 8.5). The most fundamental about the molecular structure is whether formally the same chemical bonds (say, the C–H ones) are indeed similar for both molecules. A purist approach says that each molecule is a New World, and thus these are two different bonds by definition. Yet chemical intuition says that some local interactions (in the vicinity of a given bond) should mainly influence the bonding. If such local interactions are similar, the bonds should turn out to be similar as well. Of course, the purist approach is formally right, but the local interactions turn out to be rather small. If chemists desperately clung to purist theory, they would know some 0.01% or so of what they now know about molecules. *It is of fundamental importance for chemistry that we do not study particular cases,*

Table 8.5. The dimensions of the electron pairs, i.e., $\sqrt{\langle r^2\rangle}$ (a.u.) for CH_3OH and CH_3SH according to Csizmadia.[a] "Core" means the $1s$ orbital of the atom indicated.

CH_3OH		CH_3SH	
core O	0.270	core S ($1s$)	0.148
core C	0.353	core C	0.357
		S (L shell)	0.489
			0.489
			0.483
			0.484
CO	1.499	CS	2.031
CH_1[b]	1.576	CH_1[b]	1.606
$CH_{2,3}$[b]	1.576	$CH_{2,3}$[b]	1.589
OH	1.353	SH	1.847
lone $pair_{1,2}$[c]	1.293	lone $pair_{1,2}$[c]	1.886

[a] I.G. Csizmadia, in "*Localization and Delocalization in Quantum Chemistry*," eds. O. Chalvet and R. Daudel, D. Reidel Publ. Co., Dordrecht, 1975.
[b] Different electron pair dimensions originate from their different positions versus the OH or SH group.
[c] There are two lone pairs in the molecule.

case by case, but derive general rules. Strictly speaking, these rules are false from the very beginning, for they are valid to some extent only, but they enable chemists to understand, to operate, and to be efficient; otherwise there would be no chemistry at all.

The periodicity of chemical elements discovered by Mendeleev is another great idea of chemistry. It has its source in the shell structure of atoms. This causes, e.g., that the compounds of sulphur with hydrogen should be similar to the compounds of oxygen with hydrogen,[110] because sulphur and oxygen have analogous electronic configurations of the valence electrons (i.e., those of the highest energies), and they differ only in the inner shells (O: $[He]2s^2 2p^4$ as compared to S: $[Ne]3s^2 3p^4$).

Take a look of Table 8.5. Note that:

1. The dimension of the electron pair localized on the $1s$ orbital of the sulphur atom is twice as small as the dimension of a similar pair of the $1s$ orbital of the oxygen atom. Nothing special. The electrons occupying the $1s$ orbital of S experience a strong electric field of

---

[110] Only those who have carried out experiments in person with mercaptan, or who have had neighbors (even distant ones) involved in such experiments, understand how important the difference between the O–H and S–H bonds really is. In view of the theoretical results reported, I am sure they also appreciate the blessing of theoretical work. According to the Guinness book of records, CH_3SH (mercaptan) is the most smelly substance in the Universe.

the nucleus charged +16, while the charge of the O nucleus is only +8. Let us note that the core of the carbon atom is even larger, because it is controlled by an even less charged nucleus[111] (+6).

2. The dimension of the electron pair of the 1s orbital of the carbon atom (core C) for CH_3OH is very similar to that of the corresponding orbital for CH_3SH (0.353 versus 0.357).[112] This means that the influence of the S atom (as compared to the oxygen atom) on the 1s orbital of the *neighboring atom* is small. The *local character of the interactions* is thus the most decisive.
3. The influence of the S and O atoms on the C–H bonds of the methyl group is only slightly larger. For example, in CH_3OH one of the C–H bond-localized orbitals has the dimension of 1.576 a.u., while in CH_3SH it is 1.606 a.u.
4. The three C–H bonds in methanol are very similar to each other (the numbers in Table 8.5 are identical), yet only two of them are strictly equivalent due to symmetry. It is even more interesting that the C–H bonds in CH_3SH are also similar to them, although the differences between the various C–H bonds of mercaptan, and between the corresponding C–H bonds in methanol and mercaptan, are clearer. So, *even despite the different atomic environment, the chemical bond preserves its principal and individual features*.
5. Let me apologize for a banality: CH_3SH differs from CH_3OH in that the O atom is replaced by the S atom. No wonder then that large differences in the close vicinity of the O and S atoms are easily noticeable. The dimensions of the electron pairs at the S atom (lone pairs and the S–H and C–S bonds) are always larger than the corresponding pair at the O atom. The differences are at the 30% level. The sulphur atom is simply larger than the oxygen atom, indicating that the electrons are more loosely bound when we go down within a given group of the periodic table.

These conclusions are instructive and strongly encouraging, because *we see a locality in chemistry, and therefore chemistry is easier than it might be* (e.g., C–H bonds have similar properties in two different molecules). On the other hand, we may play a subtle game with local differences on purpose, by making suitable chemical substitutions. In this way we have a possibility of *tuning the chemical and physical properties of reactants, which is of prime importance in designing and synthesis of new materials*.

---

[111] Let us check what the Slater orbitals give for the dimension of the 1s core orbital for the C, O, and S atoms. The nuclear charge has to be diminished by 0.3, so in these three cases they should be $Z = 5.70, 7.70, 15.70$, respectively. The mean value of the nucleus–electron distance can be easily computed as $\sqrt{\langle 1s|r^2|1s\rangle} = \sqrt{\frac{Z^3}{\pi}\int_0^\infty r^4\exp(-2Zr)dr\int_0^\pi \sin\theta d\theta\int_0^{2\pi} d\phi} = \sqrt{\frac{Z^3}{\pi}\cdot 4!\,(2Z)^{-5}\cdot 2\cdot 2\pi} = \frac{\sqrt{3}}{Z}$. The results are 0.30, 0.22, 0.11, whereas the reported (more accurate) data are 0.35, 0.27, 0.15 a.u., respectively.

[112] Even these small changes may be detected experimentally by the removal of electrons from the molecules by monochromatic X-ray radiation and subsequent measurement of the kinetic energy of the removed electrons. Those which were more strongly bound run slower.

8.9.8 Hybridization or mixing one-center AOs

The localized orbitals may serve to present the idea of a *hybrid atomic orbital*. A given localized orbital φ of a bond represents an LCAO of *mainly two atoms* – the partners which form the chemical bond, say, a and b. If so, then (for each localized bond orbital) all the AOs of atom a may be added together with their specific LCAO coefficients,[113] and the same can be done for atom b. These two sums represent two *normalized hybrid atomic orbitals* h_a and h_b multiplied by the resulting coefficients c_a and c_b and together form the approximate[114] bond orbital

$$\varphi \approx c_a h_a + c_b h_b$$

with the corresponding LCAO expansions

$$h_a = \sum_{j \in a} c_{ji} \chi_j,$$

$$h_b = \sum_{j \in b} c_{ji} \chi_j.$$

Such a definition of the hybrid orbitals is not unique, since the localized orbitals used are also not unique. However, as shown above, this ambiguity is of secondary importance. The advantage of such an approach to hybridization is that:

- It can be determined for any configuration of the nuclei, e.g., for the tetrahedral as well as for any other configuration of CH_4, etc.
- The definition is applicable at any LCAO basis set used.
- It gives a clear message that *all* the atoms in a molecule are hybridized (why not?), e.g., the carbon atom in the methane molecule as well as all the hydrogen atoms. The only difference between these two hybridizations is that the h_a for the carbon atom does not resemble any of the χ_j in $\sum_{j \in a} c_{ji} \chi_j$ (because of comparable values[115] of $|c_{ji}|$ meaning an effective mixing of the AOs), while the h_b for the hydrogen atom is *dominated by a single AO* $1s_b$, which should be understood as a lack of hybridization.[116]

How will the hybridization in the optimized geometry of methane look? Well, among five doubly occupied localized MOs, four[117] protrude from the carbon nucleus towards one of the

---

[113] That serve to express the localized orbital through the atomic basis set.

[114] The “tails” of the localized orbital, i.e., its amplitudes on other atoms, have been neglected.

[115] Mainly of $2s_a$ and $2p_a$, which have the highest values of the LCAO coefficients.

[116] The reason why the carbon atom (and some other atoms such as N, O, etc.) is effectively hybridized, while the hydrogen atom is not, is that the $2s$ and $2p$ orbital energy levels in those atoms are close in the energy scale, while the energy difference between the $1s$ hydrogen orbital energy and higher-energy hydrogen orbitals is larger.

[117] The fifth will be composed mainly of the $1s$ carbon orbital.

hydrogens (four hydrogens form a regular tetrahedron) and will have only some marginal amplitudes on the three other hydrogens. If we neglect these "tails" on the other atoms and the contributions of the AOs other than $2s$ and $2p$ (i.e., their c_{ji}'s) of the carbon atom (also eliminating from the MO the $1s$ orbital of the partner hydrogen atom), we obtain the following normalized hybrid carbon orbitals:

$$h_i = \frac{1}{\sqrt{1+\lambda_i^2}} \left[(2s) + \lambda_i (2p_i) \right],$$

with $i = 1, 2, 3, 4$ denoting the four directions of p_i and therefore of h_i. If we force the four hybrids to be equivalent, then this means $\lambda_i = \lambda$. Forcing the hybrids to be mutually orthogonal,[118]

$$\langle h_i | h_j \rangle = \frac{1}{1+\lambda^2} \left[1 + \lambda^2 \langle 2p_i | 2p_j \rangle \right] = \frac{1}{1+\lambda^2} \left[1 + \lambda^2 \cos\theta_{ij} \right] = 0,$$

we obtain as the $2s$ and $2p$ mixing ratio

$$\lambda = \sqrt{\frac{-1}{\cos\theta_{ij}}}. \tag{8.109}$$

sp^3 *hybridization (tetrahedral)*

For the tetrahedral configuration $\theta_{ij} = 109°28'$; hence, from Eq. (8.109), $\cos 109°28' = -\frac{1}{3}$ and, therefore, $\lambda_i = \sqrt{3}$. Therefore the orthogonal hybrids on the carbon atom (Figs. 8.33 and 8.34a) read as

$$h_i\left(sp^3\right) = \frac{1}{2}\left[(2s) + \sqrt{3}(2p_i) \right],$$

where $2s$ and $2p_i$ are the normalized carbon atomic orbitals[119] with i denoting the direction of the hybrid, one of the four directions from the carbon atom towards the tetrahedrally located hydrogen atoms.[120]

---

[118] "Orthogonal" also means "absolutely independent."

[119] Say, the STOs (p. 494).

[120] Such orientation of the (normalized) $2p_i$'s may be achieved by the following choices (just look at the vortices of a cube with the carbon atom at its center and the four directions forming the tetrahedron):
$2p_1 = \frac{1}{\sqrt{3}}(2p_x + 2p_y + 2p_z)$,
$2p_2 = \frac{1}{\sqrt{3}}(2p_x - 2p_y - 2p_z)$,
$2p_3 = \frac{1}{\sqrt{3}}(-2p_x + 2p_y - 2p_z)$,
$2p_4 = \frac{1}{\sqrt{3}}(-2p_x - 2p_y + 2p_z)$.
The normalization of the above functions is obvious, since $2p_x, 2p_y, 2p_z$ are orthogonal.

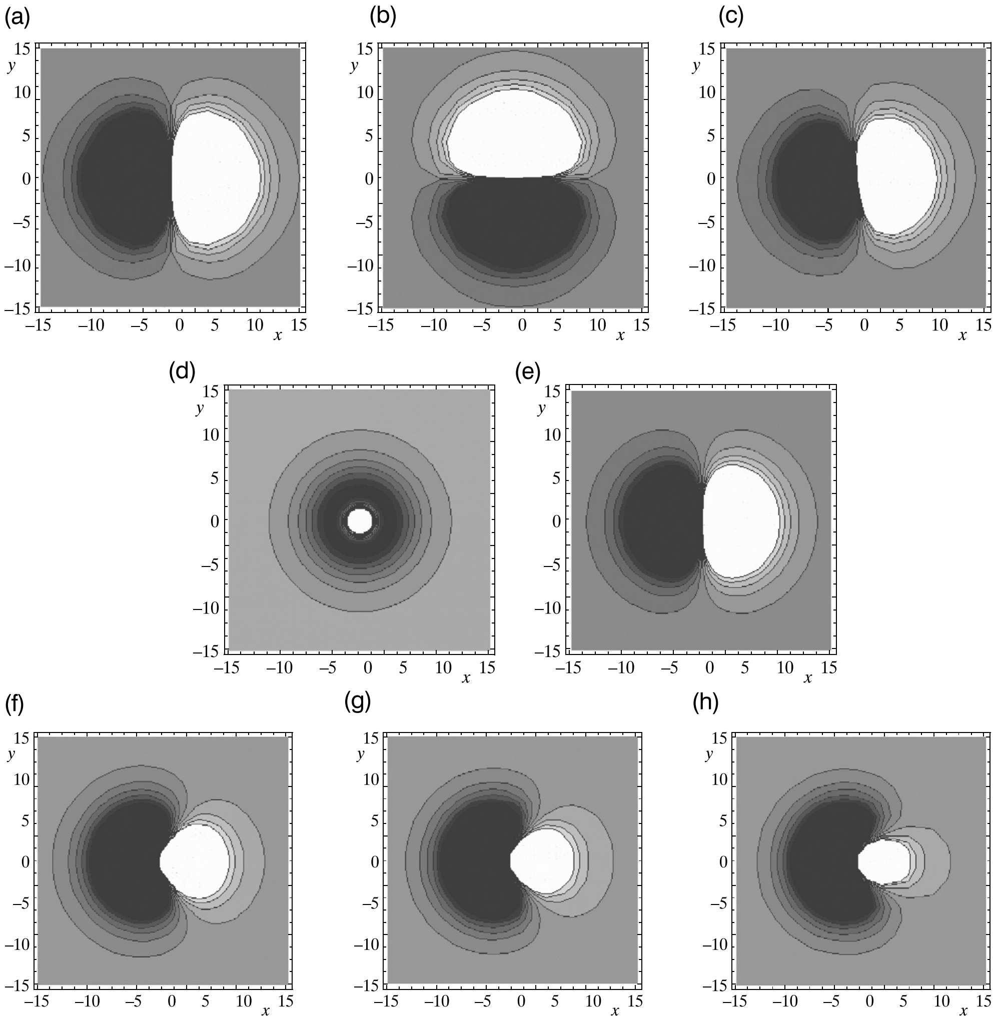


Fig. 8.33. The Slater-type orbitals shown as contours of the section at $z = 0$. The background corresponds to the zero value of the orbital, the darker regions to the negative, and the brighter to the positive value of the orbital. (a) $2p_x$, (b) $2p_y$, and (c) their linear combination $\cos 5°2p_x + \sin 5°2p_y$, which is also a $2p$ orbital, but rotated by 5° with respect to the $2p_x$ orbital. In (d) and (e) we show the normalized $2s$ and $2p$ orbitals, which will now be mixed in various proportions. (f) The 1:1 ratio, i.e., the sp hybridization. (g) The $1{:}\sqrt{2}$ ratio, i.e., the sp^2 hybridization. (h) The $1{:}\sqrt{3}$ ratio, i.e., the sp^3 hybridization.

sp^2 hybridization (trigonal)

If we tried to find the lowest-energy configuration of ethylene (C_2H_4), it would correspond to a planar structure (Fig. 8.34b) of D_{2h} symmetry. After analyzing the localized MOs, it would turn

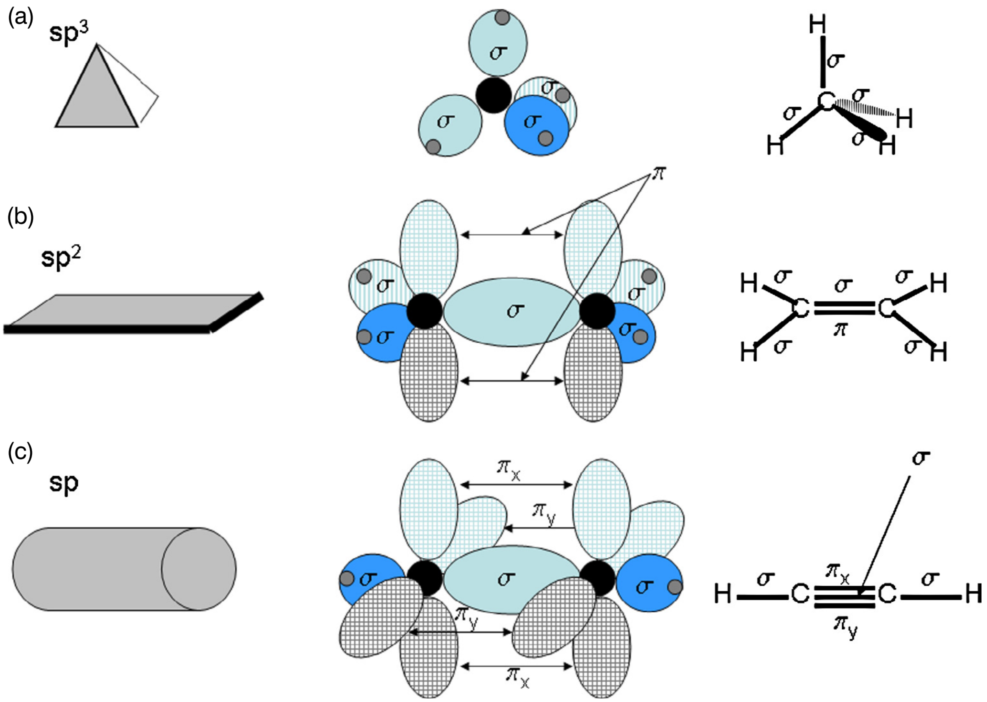


Fig. 8.34. (a) The sp^3 hybridization in the methane molecule in its tetrahedral equilibrium geometry (which corresponds to the minimum of the ground-state electronic energy $E_0^0(\mathbf{R})$, see p. 320). There are four doubly occupied CH-localized MOs and one that is essentially the doubly occupied $1s$ carbon AO. Each of the CH MOs (of *nearly* cylindrical symmetry) is composed mainly of the carbon hybrid and the hydrogen $1s$ AO. The figure shows a scheme of the four carbon hybrids called the sp^3 hybrids. (b) An example of the nearly perfect sp^2 hybridization of the carbon atoms in ethylene (C_2H_4), which is perfectly planar in its ground electronic state (D_{2h} symmetry). Such a hybridization is only approximate, because the C–C–H angle has to differ from the H–C–H angle; both slightly deviate from 120°. The localized MOs are the following (occupied by altogether 16 electrons):
— two essentially $1s$ carbon orbitals,
— four CH orbitals and one CC orbital having *nearly* cylindrical symmetry (i.e., σ type),
— one bond orbital being antisymmetric with respect to the reflection in the molecular plane (i.e., of π symmetry).
(c) An example of sp hybridization: the acetylene molecule. The Hartree–Fock geometry optimization gives the lowest-energy linear configuration: HCCH. The localization gives seven localized molecular orbitals:
— two of them are essentially the $1s$ carbon orbitals,
— two represent the cylindrical CH orbitals (σ),
— one cylindrical CC σ orbital,
— two CC orbitals that are of π symmetry (perpendicular to each other).

out that three hybrids protrude from each carbon nucleus, their directions lying in the molecular plane (say, xy). These hybrids form angles very close to 120°.

For the trigonal hybridization (i.e., pure sp^2 hybridization, with $\theta_{ij} = 120°$ angles) we obtain from Eq. (8.109) $\lambda = \sqrt{2}$, and, therefore, the three orthogonal normalized sp^2 hybrids are

$$h_i\left(sp^2\right) = \frac{1}{\sqrt{3}}\left[(2s) + \sqrt{2}(2p_i)\right],$$

where the directions $i = 1, 2, 3$ form the Mercedes logo on a plane.

sp hybridization (digonal)

Such hybridization is said to occur in acetylene (HCCH), which after optimization of the Hartree–Fock energy corresponds to the linear symmetric configuration. According to this explanation, each carbon atom exposes two hybrids (Fig. 8.34c): one towards its carbon and one towards its hydrogen partner. These hybrids use the two carbon $2s$ and the two carbon $2p_z$ and, together with the two $1s$ orbitals of the hydrogens, form the two H–C σ bonds and one C–C σ bond. This means that each carbon atom has two electrons left, which occupy its $2p_x$ and $2p_y$ orbitals (perpendicular to the molecular axis). The $2p_x$ orbitals of the two carbon atoms form the doubly occupied π_x bond-localized MO and the same happens to the $2p_y$ orbitals. In this way the carbon atoms form the C≡C triple bond composed of one σ and two π (i.e., π_x and π_y) bonds.

The angle between the two equivalent orthonormal hybrids should be $\theta_{ij} = 180°$; then the mixing ratio will be determined by $\lambda = 1$. Two such hybrids[121] the two opposite directions, $h_1(sp) = \frac{1}{\sqrt{2}}\left[(2s) + (2p_z)\right]$ and $h_2\,(sp) = \frac{1}{\sqrt{2}}\left[(2s) - (2p_z)\right]$.

Is hybridization of any value?

The general chemistry textbook descriptions of hybridization for methane, ethylene, and acetylene usually start from the electronic configuration of the carbon atom: $1s^2 2s^2 2p^2$. Then it is said that, according to valence bond theory (VB, see Chapter V2-2), this configuration predicts CH_2 as the carbon hydride (bivalent carbon atom) with the C–H bonds forming the right angle.[122] This differs very much from the way the methane molecule looks in reality (regular tetrahedron and tetravalent carbon). If the carbon atom were excited (this might happen at the expense of future energy gains – known as "promotion"), then the configuration might look like

---

[121] This cannot be exact (cf. the ethylene case), because the two hybrids must not be equivalent. One corresponds to the C–C, the other to the C–H bond.

[122] Because $2p^2$ means, say, $2p_x^1 2p_y^1$, and these singly occupied AOs form the two C–H bonds with two $1s$ hydrogen orbitals.

$1s^2 2s^1 2p_x^1 2p_y^1 2p_z^1$. The textbooks usually go directly to the mixing of the valence atomic orbitals $2s$, $2p_x 2p_y 2p_z$ to form four equivalent sp^3 hybrids, which lead directly to the tetrahedral hydride: methane. Note, however, that being in less of a rush, we would draw the conclusion that the $1s^2 2s^1 2p_x^1 2p_y^1 2p_z^1$ configuration leads to four *nonequivalent* C–H bonds in the CH_4 hydride.[123] Only equivalent mixing (hybridization) gives the correct picture. When aiming at ethylene or acetylene the reasoning changes, because some orbitals are left without mixing. We assume sp^2 (one orbital left) or sp (two orbitals left) hybridizations, respectively, which leads to the correct compounds and (almost) correct structures. *It looks as if when we know what we are aiming for, we decide what mixes and what does not.* This seems to be not fair.

Example 7 (Water molecule). Let us carry out the Hartree–Fock calculations for the water molecule.[124] We focus on a subsequent calculation of the localized MOs and get five doubly occupied MOs, shown in Fig. 8.35.

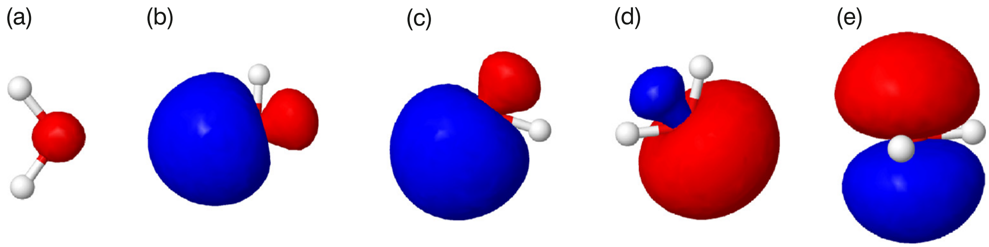


Fig. 8.35. A shock therapy. The localized MOs of a water molecule. (a) An orbital, which is strongly localized on the oxygen nucleus and nearly spherical (practically $1s$ of the oxygen atom). (b, c) The O–H bond orbitals, one for each O–H bond. (d) The first of the two lone pair orbitals is a hybride with its axis within the molecular plane. (e) The second of the two lone pair orbitals (orthogonal to the plane of the molecule) is a pure $2p$ orbital of oxygen. It is clear that a sum of the last two orbitals leads to a hybride, which has to go up off the plane. On the other hand, a subtraction of these orbitals brings a twin hybride but down from the plane. In this way we recover (in a fully correct way) the equivalent lone pairs of water molecule known from textbooks.

The lowest mean value of the orbital energy[125] corresponds to an orbital which is small and strongly localized on the oxygen nucleus; we identify it as practically $1s$ orbital of oxygen (a). Next, we have two bond orbitals (b, c), each of them looking at first sight like a hybride of the oxygen atom oriented towards a given hydrogen, but it is a bit misleading because it absorbed also the $1s$ orbital of the corresponding hydrogen atom. Then we find something that seems

---

[123] Three C–H bonds would form right angles (because of $2p_x^1, 2p_y^1, 2p_z^1$), one C–H bond however (formed by $2s^1$ together with the corresponding $1s$ hydrogen orbital) would have a quite different character. This contradicts what we get from experiments.

[124] For example, by using the web server webmo.com (Gaussian program).

[125] The localized orbitals are not the eigenfunctions of the Fock operator.

embarrassing (d, e); the last two localized MOs – the two lone pairs: one of them (d), with its axis within the plane of the molecule, looks as an sp^2 orbital of oxygen, while the second one (e) is orthogonal to the molecular plane and represents nothing but a pure $2p$ orbital of oxygen! Is it simply nonsense? Maybe the computer made a mistake! Well, it may be that the reader studied too literally the popular literature, where the two lone pairs in the water molecule always protrude right from the oxygen as in a tetrahedral configuration. Everything is all right! If the reader added/subtracted the above two orbitals (which is a perfectly legal operation!), he/she would get the nearly tetrahedrally oriented sp^3 lone pair orbitals known from the textbooks. They are no better than the ones obtained in our calculations. Well, maybe a bit nicer, because they will look similar. Both sets of the orbitals lead to the same total electronic density and the same total energy.

Example 8 (Methane molecule). Let us check how important the role of hybridization is in the formation of chemical bonds in methane. Let us imagine four scientists performing Hartree–Fock computations for methane in its tetrahedral configuration[126] of nuclei. They use four LCAO basis sets. Professor A believes that in this situation it is important to remember sp^3 hybridization and uses the following basis set:

A: $1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, h_1\left(sp^3\right), h_2\left(sp^3\right), h_3\left(sp^3\right), h_4\left(sp^3\right)$.

Student B did not read anything about hybridization and just uses the common orbitals:

B: $1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, 2s_C, 2p_{x,C}, 2p_{y,C}, 2p_{z,C}$.

Students C and D are not the brightest; they have mixed up the hybridization for methane with that for ethylene and acetylene and used the following basis sets:

C: $1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, 2p_{x,C}, h_1\left(sp^2\right), h_2\left(sp^2\right), h_3\left(sp^2\right)$.

D: $1s_{H1}, 1s_{H2}, 1s_{H3}, 1s_{H4}, 1s_C, 2p_{x,C}, 2p_{y,C}, h_1(sp), h_2(sp)$.

Who of these scientists will obtain the lowest total energy, i.e., the best approximation to the wave function?

Well, we could perform these calculations, but it is a waste of time. Indeed, each of the scientists used different basis sets, but they all used *the same space* spanned by the AOs.[127] This is because all these hybrids are linear combinations of the orbitals of student B. All the scientists are bound to obtain the same total energy, the same molecular orbitals,[128] and the same orbital energies.

---

[126] Or any other one.

[127] Any of the AOs and, therefore, also any of the MOs of a person can be expressed as an LCAO of another person.

[128] Although the LCAO coefficients will be, of course, different, because the expansion functions are different. The orbital plots will be the same.

Hybridization is useful before the calculations are performed

Is hybridization a useless concept then? No, it is not. It serves as a first indicator (when calculations are not yet performed) of what happens to a local atomic electronic structure if the atomic configuration is tetrahedral, trigonal, etc. For example, the trigonal hybrids describe the main features of the electronic configuration in the benzene molecule (Fig. 8.36).

Let us take the slightly more complicated example of a molecule that is of great importance in biology (Fig. 8.37).

It is important to remember that we always start from some chemical intuition[129] and use the structural formula given in Fig. 8.37a. Most often we do not even consider other possibilities (isomers), like those shown in Fig. 8.37b. Now, we try to imagine what kind of local electronic structure we have around the particular atoms. Let us start from the methyl ($–CH_3$) functional groups. Of course, such a group resembles methane, except that one carbon hybrid extends to another atom (not hydrogen). Thus, we expect hybridization over there close to the sp^3 one (with all consequences, i.e., angles, etc.). Next, we have the carbon atom that is believed[130] to make the double bond with the oxygen atom. The double bond means an ethylene-like situation, i.e., both atoms should have hybridizations similar to sp^2. Let us begin from the oxygen atom (Fig. 8.37c). The sp^2 means three hybrids (planar configuration) protruding from the O atom. One of them will certainly bind to a similar one protruding from the carbon atom (O–C σ bond); it therefore needs only a single electron from the oxygen. The oxygen atom has six valence electrons, therefore there remain five to think of. Four of them will occupy the other two hybrids protruding into space (nothing to bind; they are lone pairs). Hence there is one electron left. This is very good, because it will participate in the O–C π bond. Let us go to the partner carbon atom. It is supposed to make a double bond with the oxygen. Hence it is reasonable to ascribe to it an ethylene-like hybridization as well. Out of four valence electrons for carbon, two are already used up by the σ and π C–O bonds. Two other sp^2 hybrids remain that, of course, accommodate the two electrons and therefore are able to make two σ bonds: one with $–CH_3$ and one with the nitrogen atom. Then we go to the nitrogen atom. It has three substituents in most cases in the (almost) planar configuration (we know this from experiment).

---

[129] Based on the vast experience of chemists.

[130] Here we rely on the concept of what is known as the *valency* of atoms, i.e., the number of bonds a given atom is able to form with its neighbors. The valency is equal to the number of valence electrons or valence holes, e.g., the valency of the carbon atom is four (because its electron configuration is $K2s^22p^2$, four valence electrons), that of the oxygen atom is two (because its electron configuration is $K2s^22p^4$, two valence holes). An element may have several valencies, because of possible opening of several electronic shells.
Note that we are making several assumptions based on chemical intuition or knowledge. The reason is that we want to go quickly without performing any computations. *This ambiguity disappears if we make real computations, e.g., using the Hartree–Fock method. Then the chemical bonds, hybrids, etc., are obtained as a result of the computations.*

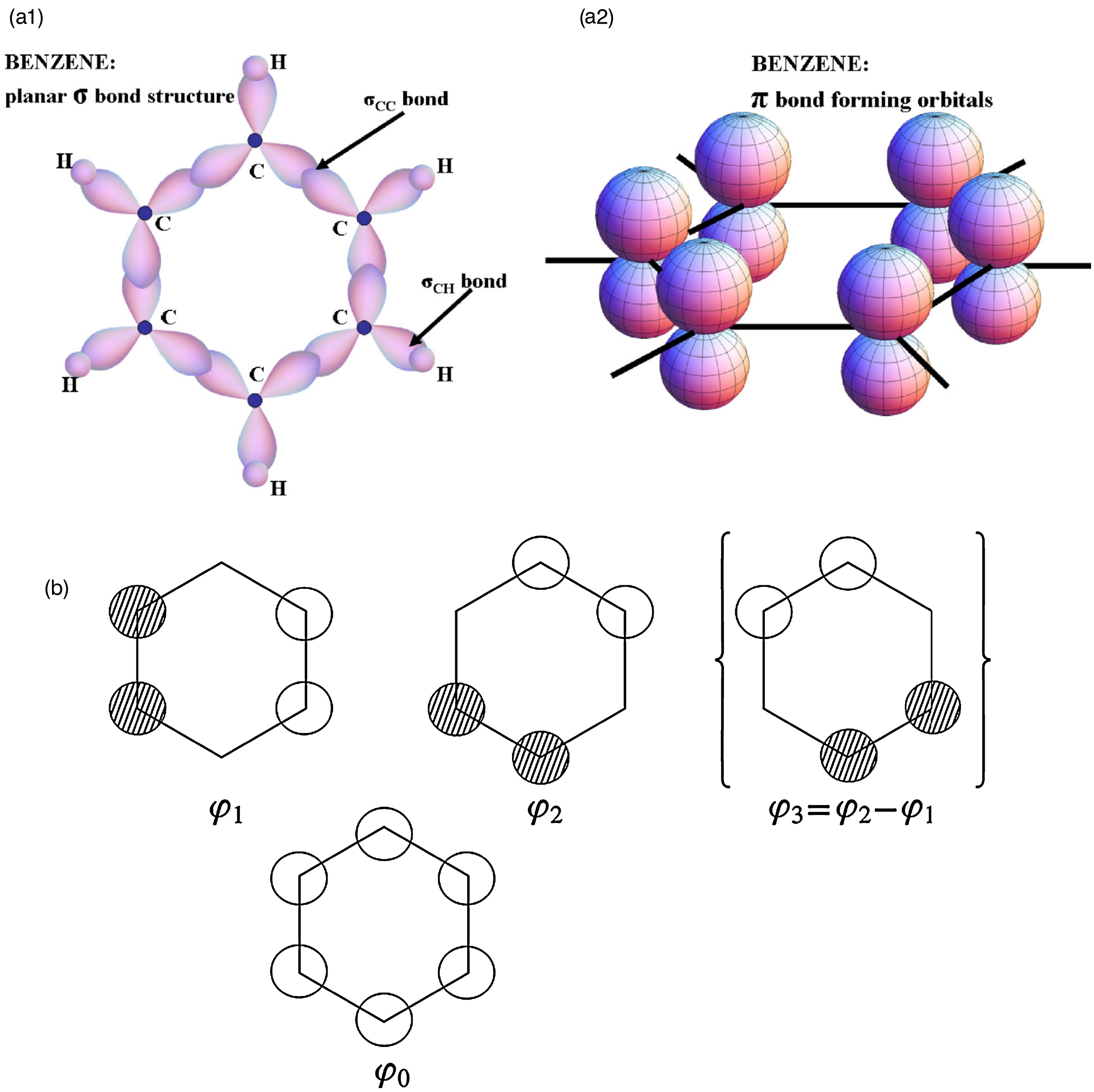


Fig. 8.36. The benzene molecule. (a) The hybridization concept allows us to link the actual geometry of a molecule with its electronic structure. The sp^2 hybrids of the six carbon atoms form the six σ C–C bonds and the structure is planar. Each carbon atom thus uses two out of its three sp^2 hybrids, the third one lying in the same plane protrudes towards a hydrogen atom and forms the σ C–H bond. In this way, each carbon atom uses its three valence electrons. The fourth one resides on the $2p$ orbital that is perpendicular to the molecular plane. (b) The six $2p$ orbitals form six π molecular orbitals, out of which three are doubly occupied and three are empty. The doubly occupied ones are shown. The φ_0 of the lowest energy is an all-in-phase linear combination of the $2p$ atomic orbitals (only their upper lobes are shown). The φ_1 and φ_2 correspond to the same higher energy, and have a single node (apart from the node plane of the AOs). The φ_3 orbital that apparently completes all combinations of single-node MOs is redundant (that is why it is shown in parentheses), because the orbital represents a linear combination of the φ_1 and φ_2.

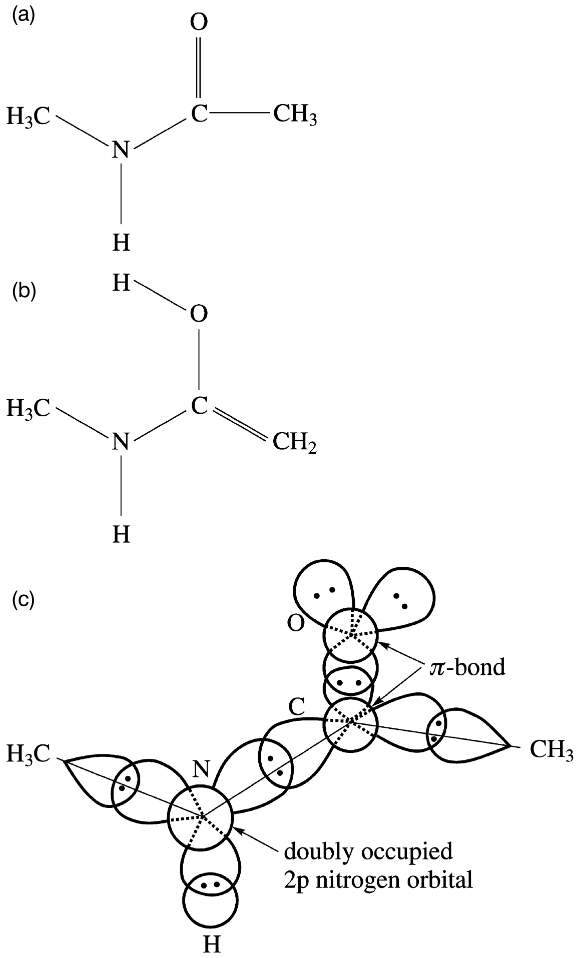


Fig. 8.37. How does the hybridization concept help? The figure shows the all important example of the peptide bond. (a) We assume a certain pattern of the chemical bonds (this choice is knowledge-based), ignoring other possibilities, such as the isomers shown in (b). Apart from the methyl groups (they have the familiar tetrahedral configuration) the molecule is planar. Usually in chemistry, knowing the geometry, we make a conjecture pertaining to the hybridization of particular atoms. This leads to the electron count for each atom: the electrons left are supposed to participate in bonds with other atoms. (c) In the example shown, the sp^2 hybridization is assumed for the central carbon and for the nitrogen and oxygen atoms. A π bonding interaction of the nitrogen, carbon, and oxygen should therefore stabilize the planarity of the system, which is indeed an experimental fact.

To make the analysis simple, we assume an sp^2 ideal hybridization. The nitrogen atom has five valence electrons. Three of them will go to form the σ N–C, N–H, N–CH_3 bonds. Note that although the configuration at N is assumed to be planar, this plane may not coincide with the analogous plane on the carbon atom. Finally, we predict the last two valence electrons of the nitrogen will occupy the $2p$ orbital perpendicular to the plane determined by the substituents of the nitrogen. Note that the $2p$ orbital could overlap (making a bonding effect) with the analogous $2p$ orbital of the carbon atom *provided that the two planes will coincide*. This is why we could expect the planarity of the C–N bond. This bond (known as the amide bond) plays a prominent role in proteins, because it is responsible for making the chain of amino acid residues. It is an experimental fact that *deviations of the amide bond from planarity are very small*.

The value of the analyses as that given above is limited to qualitative predictions. Of course, computations would give us a much more precise picture of the molecule. In such computations the orbitals would be more precise, or would be not present at all, because, to tell the truth, there is no such thing as orbitals. We badly need to interpret the numbers, to communicate them to others in a understandable way, to say whether we expect such numbers or they are totally unexpected. Reasoning like that given above has a great value as part of our understanding of chemistry, of speaking about chemistry, of predicting and discussing the structures. This is why we need hybridization. Moreover, if our calculations were performed within the valence bond method (in its simplest formulation; the details of the method will be explained in Chapter V2-2), then the lowest energy would be obtained by Professor A (who assumed the sp^3 hybridization), because the energy gain over there is very much connected to the overlap of the AOs forming the basis, and the overlap with the $1s$ hydrogen orbitals is the best for the basis set of Professor A. The other people would get high total energies, because of poor overlap of their AOs with the $1s$ hydrogen orbitals.

8.10 A minimal model of a molecule

It is easy to agree that our world is a complex business. It would be great, however, to understand how this business is operating. At least sometimes answers look more and more complex as we go from crude to more and more accurate theories. Therefore, we would like to consider a simpler world (a model of our real world), which:

- would work to very good accuracy, i.e., resembled the real world quite well, and
- would be based on such simple rules that we could understand it in detail.

We could explain these rules to anybody who were interested. Not only could we predict a lot for a molecular system, but we ourselves could be confident that we understand *most* of chemistry, because it is based on *several simple rules*. Moreover, why worry about details? Most often we just want to grasp the essence of the problem. On top of that, if this essence were free, only sometimes would we be interested in a more detailed (and expensive) picture.

Is this utopia or can such a model of chemistry be built?

Well, it seems that theoretical chemistry nowadays offers such a model describing chemical structures.

The model is based on the following basic simplifications of the real world:

- The *nonrelativistic approach*, i.e., the speed of light is assumed to be infinite, which leads to the Schrödinger equation (Chapter 2).
- The Born–Oppenheimer approximation (Chapter 6) that separates the motion of the nuclei from the motion of the electrons. This approximation allows us to introduce the concept of a *three-dimensional structure* of the molecule: the heavy nuclear framework of the molecule kept together by "electronic glue" moves (translation), and at the same time rotates in space.
- The mean-field approximation of the present chapter offers us the *orbital model* of the electronic structure of molecules within the RHF approach. In this picture the electrons are described by the doubly occupied MOs. Localization of the orbitals gives the doubly occupied *inner shell*, *lone pair*, and *bond* MOs. The first and second are sitting on atoms, the latter on chemical bonds. Not all atoms are bound with all, but instead the *molecule has a pattern of chemical bonds*.
- These bonds are traditionally and formally represented by graphs suggesting a *single*, e.g., C–H; *double*, e.g., C=C; or *triple*, e.g., C≡C, bond, although some intermediate situations usually take place. The total number of these formal bonds of a given atom is equal to its valency. The idea of valency helps a lot in selecting the chemical bond pattern, which afterwards may be checked against experiment (e.g., bond distances).[131] In most cases a single bond is of the σ type, a double one is composed of one σ and one π, and a triple bond means one σ and two π bonds (cf. p. 555).
- The minimal model of a molecule may explain most of the chemical reactions, if besides the closed shell configuration (double occupancy of the MOs, including HOMO) we consider excited configurations corresponding to electron transfer(s) from the HOMO to the LUMO (see "*Ideas of Quantum Chemistry. Vol. 2: Interactions*").
- The bonds behave very much like *springs of a certain strength and length*,[132] and therefore, apart from the translational and rotational motion, the atoms vibrate about their equilibrium positions.[133] As to the structural problems (not chemical reactions), these vibrations may be treated as harmonic.

---

[131] For some molecules this procedure is not unique, i.e., several chemical bond patterns may be conceived (sometimes called "resonance structures" or "mesomeric forms," cf. the valence bond method in Chapter V2-2). In such cases the real electronic structure corresponds to an averaging of all of them (in space or in time).

[132] Both depend first of all of the elements making the bond; also, a single bond is the weakest and longest, the triple is the strongest and shortest.

[133] The molecular model visualized in virtually all popular computer programs shows spherical atoms and chemical bonds as shining rods connecting them. First of all, atoms are not spherical, as is revealed by Bader analysis (p. V2-197) and atomic multipole representations (Appendix V2-D). Second, a chemical bond resembles more

- The three-dimensional shape of simple chemical structures can be correctly predicted using the Hartree–Fock model. The main features of this three-dimensional structure can be also predicted (without any calculation) by using the concept of the *minimum repulsion energy of the electrons pairs*. Within the molecular orbital model, such repulsion is given by Eq. (8.99).

Valence Shell Electron Pair Repulsion (VSEPR) algorithm.[134]

Is it possible to predict – in a minute, without any quantum chemical calculations – the shape of a molecule corresponding to the lowest energy, or more exactly the configuration of ligands around a central atom? It turns out that often[135] such a goal is feasible and moreover the rules (the VSEPR algorithm) behind such prediction are very simple! Here they are:

- The VSEPR algorithm starts from choosing the central atom (C) in the molecule and defines the rest as a set of its ligands (L).[136]
- The key step of the VSEPR algorithm is to calculate an integer number (N) of electrons assigned to the valence shell of the central atom according to the formula $N = n_C + n_L + n_{ion}$, where n_C stands for the number of the valence electrons of the central atom itself, n_L denotes the number of electrons offered by the ligands L, while n_{ion} is an obvious correction for the number of electrons if one considers an ion, not a neutral molecule.
    - Calculation of n_C. This number is calculated from the position of the central atom in the Mendeleev periodic table (n_C is the group number): for alkali metals $n_C = 1$, for alkali earth metals $n_C = 2$, for analogs of boron $n_C = 3$, for analogs of carbon $n_C = 4$, for analogs of nitrogen $n_C = 5$, for analogs of oxygen $n_C = 6$, for halogens $n_C = 7$, and for noble gases $n_C = 8$ (except helium for which $n_C = 2$).
    - Calculation of n_L. The integer n_L is a sum of integer contributions from all individual ligands. Each ligand bound to the central atom by a single bond contributes 1, the ligands O, S, CH_2, and NH (i.e., those bound to the central atom by a double bond) contribute 0,

---

a "rope" (higher values) of electronic density than a cylindrical rod. The "rope" is not quite straight and is slimmest at a critical point (see p. V2-201). Moreover, the rope, when cut perpendicularly, has a circular cross-section for pure σ bonds, and an oval cross-section for the double bond σ and π.

[134] The VSEPR algorithm comes from Ronald Gillespie and Ronald Nyholm (R.J. Gillespie, R.S. Nyholm, *Quart. Rev. Chem. Soc.*, 11(1957)339). Witold Mizerski from the University of Warsaw, Chemistry Department for many years edited a high-level popular chemistry journal, "Kurier chemiczny." The present description of the VSEPR is based on Mizerski's article "Prediction of molecular shape" in this journal.

[135] The VSEPR works correctly for small molecules almost without exceptions, for larger molecules one meets difficulties of two types: there may be problems with heavy atoms and the VSEPR algorithm is unable to solve conformational problems (i.e., to show the lowest-energy conformer).

[136] As the central atom can be chosen any atom of the molecule, one may also repeat the VSEPR procedure for different choices of the central atom and check whether the outputs are consistent.

the ligands N and CH (i.e., those bound to the central atom by a triple bond) contribute -1.

- Calculation of n_{ion}. If the total system is a cation of charge $+n|e|$ the number $n_{ion} = -n$, for an anion of charge $-n|e|$ the number $n_{ion} = +n$.

- One calculates *the number of the valence electronic pairs assigned to the central atom* as[137] $P = \frac{N}{2}$.
- Distribution of the electron pairs found around the central atom. The pairs are distributed on a surface of a sphere in such a way as to have their repulsion the smallest possible (the largest distances between them).
- Distribution of the ligands around the central atom (i.e., the shape of the molecule). Each ligand is assigned to one of the electron pairs. They are called the ligand pairs; the other ones will play the role of the lone electron pairs. If the number of ligands is equal to P, the structure is determined. If the number of ligands is larger than P, there is an error in our calculations! If the number of ligands is smaller than P, we have a problem of several structures possible. *There comes the last VSEPR rule: two lone pairs repel stronger than the lone pair–ligand pair and even stronger than the ligand pair–ligand pair.*[138] This sequence leads to a unique VSEPR-predicted structure; they are listed in Fig. 8.38 for various P and n_L.

Example 9 (Water). Let us first try the VSEPR model. For the oxygen atom as the central one we have $N = n_C + n_L + n_{ion} = 6 + 2 + 0 = 8$. Hence the number of electronic pairs $P = N/2 = 4$. The largest distances between the pairs will be ensured by the tetrahedral configuration (Fig. 8.38, the angles equal 109°28′). Any assignment of the ligands (H) gives, of course, the same bent H–O–H structure. Recalling the strong repulsion between lone pairs, we can predict the H–O–H angle to be smaller than 109°28′. This is a quite good guess, because the experimental H–O–H angle is equal to 104.5°.

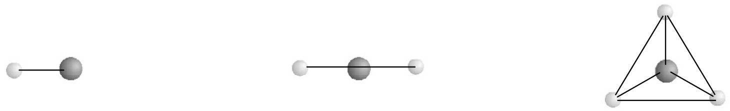
Let us check now what the minimal model gives. The model (STO 6-31G** basis set, optimized geometry) predicts correctly that we have to do with two *equivalent* O–H bonds, because both bonds have the same length[139] $R_{\mathrm{OH}} = 0.943$ Å, which is quite close to what can be deduced

---

[137] If N is an odd number (rare), one rounds off the number to $(N + 1)$.

[138] Note that at a given geometry a minimization of the electron pair repulsion given by Eq. (8.99) means localization of the MOs. If, however, a change of geometry is considered (which is the crux of the VSEPR), a smaller electron repulsion (i.e., a smaller *const* in (8.98)) stabilizes the structure. For small variation of angles L–C–L one may expect a marginal change of the self-repulsion, i.e., $\sum_i^{MO} \mathcal{J}_{ii} = \sum_i^{MO} (ii|ii)$, because each term depends uniquely on a localized orbital on C or on an individual C–L, not involving any two ligands. The repulsion is expected to vary for *different* localized orbitals, because their distance will vary. The larger the distance, the smaller the value of (8.98) and (8.99). This may be seen as a theoretical hint for validity of the VSEPR.

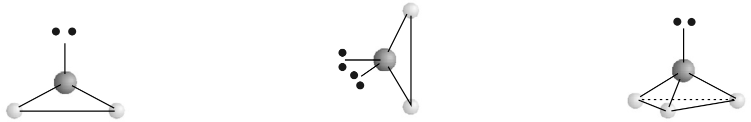
[139] The VSEPR model does not provide information about interatomic distances.



P=1, lig.=1
linear molecule
HCl, LiH, H_2

P=2, lig.=2
linear molecule
CO_2, $BeCl_2$, BeH_2

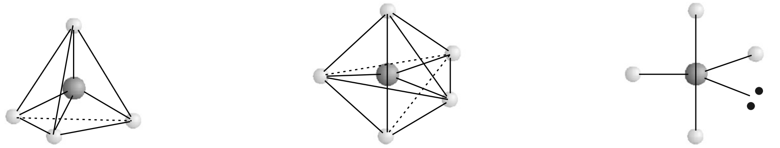
P=3, lig.=3
equilateral triangle
SO_3, CO_3^{2-}, NO_3^-, BCl_3, HCHO



P=3, lig.=2
bent mol. (angle<120°)
SO_2, NO_2^-, NO_2, $SnCl_2$, O_3

P=4, lig.=2
bent mol. (angle<109°28')
H_2O, ClO_2^-, OF_2

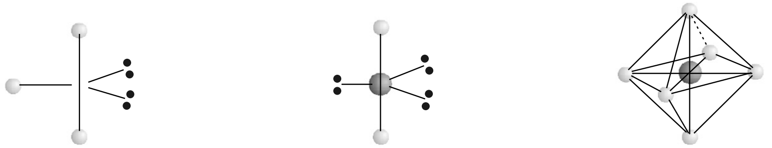
P=4, lig.=3
pyramide (angle<109°28')
NH_3, SO_3^{2-}, $SOCl_2$, PCl_3, H_3O^+



P=4, lig.=4
tetrahedral (angle=109°28')
CH_4, NH_4^+, CCl_4, SO_4^{2-}, PO_4^{3-}

P=5, lig.=5
trigon.bipyramide (120°, 90°)
PCl_5, SOF_4, IO_5^{3-}

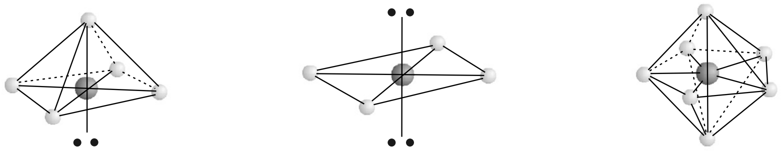
P=5, lig.=4
deformed tetrahedron
SF_4, $TeCl_4$, $ClOF_3$



P=5, lig.=3
T-shaped
ClF_3, BrF_3, $XeF_3{}^+$

P=5, lig.=2
linear molecule
XeF_2, ICl_2^-, I_3^-

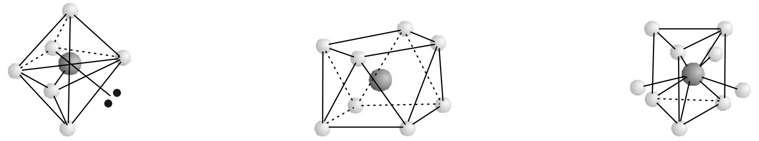
P=6, lig.=6
octahedral (angle 90°)
SF_6, AlF_6^{3-}, XeO_6^{4-}, IO_6^{5-}



P=6, lig.=5
square pyramide
IF_5, TeF_5^-

P=6, lig.=4
square (angles=90°)
XeF_4, ICl_4^-

P=7, lig.=7
pentagon.bipyram. (90°, 108°)
IF_7



P=7, lig.=6
deformed octahedron
XeF_6

P=8, lig.=8
square antiprism
ReF_8^-

P=9, lig.=9
face centered trigonal prism
ReH_9^{12-}

Fig. 8.38. The most stable structures according to VSEPR. The structures are assigned by assuming that the lone pair–lone pair repulsion is stronger than the lone pair–bonding pair repulsion, the latter still stronger than the bonding pair–bonding pair repulsion.

from the microwave spectroscopy[140]: $R_{\mathrm{OH}} = 0.957$ Å. The minimal model used predicts also that the molecule is bent (C_{2v} symmetry), the H–O–H angle being 106.0° (the Hartree–Fock limit, i.e., calculated with the complete AO set is 105.3°). *These results are quite typical for the minimal model: it is able to predict the bond lengths*[141] *to the accuracy of about* 0.01 Å, *and the bond–bond angles with an accuracy of about* 1°. There is no H–H bond in H_2O at this geometry.[142]

The minimal model predicts three harmonic modes for the water molecule (Fig. 7.8 on p. 423): antisymmetric stretching vibration with the wave number 4264 cm^{-1}, symmetric stretching vibration with 4147 cm^{-1}, and bending vibration with 1770 cm^{-1}. It is not so easy to say how to compare these numbers to the experimental absorption lines observed in spectra. What an experimentalist measures is related to the vibrational excitations from the ground state $v = 0$ to the excited state with $v = 1$, but of an anharmonic oscillator. One may deduce what the corresponding frequencies would be if the potential energy well were parabolic; they will be equal: $3942,\ 3832,\ 1648\ \mathrm{cm}^{-1}$. *This also is typical: the minimal model with its harmonic approximation gives frequencies that are higher by 7%–8% than the experimental values.*[143]

Example 10 (Ozone). This simple molecule is composed of three oxygen atoms (O_3), which at first sight should, therefore, play the same role. The concept of chemical valence (valence equals two) seems to confirm this idea allowing a structure of an equilateral triangle of single bonds. What does a more serious approach offer? One has to solve the Schrödinger equation (even at the level of the Hartree–Fock method). Before we go to this let us turn first to what the VSEPR algorithm predicts. We choose one of the oxygen atoms as a "central" one, the two other oxygen atoms will be the ligands. We count: $N = n_C + n_L + n_{ion} = 6 + 2 \cdot 0 + 0 = 6$. Hence the number of electronic pairs is $P = 3$. The repulsion principle (Fig. 8.38) gives the configuration in the form of the Mercedes sign (with angle 120°). We assign two pairs to the ligands, one lone pair is left. Therefore, the VSEPR predicts the O–O–O angle to be a bit

---

[140] The microwave spectroscopy determines the moments of inertia.

[141] We mean here molecules with well-closed electronic shells.

[142] This is common knowledge in chemistry, but if you think about how you know this, then it turns out you have to quote your teachers... As to a more serious argument, this conclusion can be drawn either from an electronic population analysis described in Appendix V2-D or by performing Bader analysis (Chapter V2-3). In the first case we would get the positive (and equivalent) populations between atoms O and H, which will result from net bonding interactions, while the population between H and H will be negative informing about an antibonding interaction. The Bader analysis would reveal a kind of electronic density "rope" between O and H and the absence of such a "rope" for H and H. No doubt, the H–H bond will appear if for any reason the H–O–H angle is forced to be small.

[143] This systematic error of theoretical description is quite often "corrected" by a proper scaling of the results. If anharmonicity becomes very large (like in the case of vibronic coupling) the error may be much larger.

smaller than 120° (due to a stronger lone pair–ligand pair repulsion), which excludes the three oxygen atoms to be equivalent!

Let us see what the Hartree–Fock method has to say about it.[144] Let us start from an equilateral triangle configuration; we optimize the geometry and get an equilateral angle ($R_{OO} = 1.373$ Å) with the energy -224.245 a.u. Well, it may be that it is a saddle point. We calculate the normal modes; all the three frequencies turn out to be real, it is a true minimum. No way, the VSEPR must have made a mistake... The VSEPR algorithm is certainly primitive and no wonder it is sometimes misleading. Well, before the last farewell with the VSEPR, let us perform a Hartree–Fock calculation (with geometry optimization), this time starting from the configuration predicted by the VSEPR. To our amazement we find another stable configuration: an isosceles triangle (C_{2v} symmetry) with the energy -224.261 a.u.,[145] which is *lower* than that computed before! Experiment confirms the VSEPR and our last result: symmetry C_{2v}. The side of the isosceles triangle is $R_{OO} = 1.204$ Å (the experimental value is[146] 1.272 Å), whereas the largest O–O distance is computed as $R_{OO} = 2.076$ Å. The computed angle is 119°, experiment gives 116.78°, while the VSEPR predicted "*less than* 120°." A triumph of the simple VSEPR algorithm, even more amazing since obtained in a quite subtle situation of two nonequivalent and competing minima on the hypersurface of the ground-state electronic energy).

Example 11 (I_3^- anion). Our experience with H_2O and O_3 suggests (hypothesis) that if one has three atoms, one gets C_{2v} symmetry of the molecule. We may check our hypothesis in case of $\mathbf{I_3^-}$, a product of dissolving iodine in a solution of potassium iodide.

Well, what does the VSEPR algorithm say about the structure? We count (for an iodine atom taken as a central one): $N = n_C + n_L + n_{ion} = 7 + 2 \cdot 1 + 1 = 10$. We get $P = 5$, which means a configuration of a trigonal bipyramid (Fig. 8.38). In the center of the bipyramid one has the central iodine atom, but where will the ligands be? There are three possibilities: two iodine atoms occupy two axial positions (I–I–I angle equal to 180°), one of them occupies an axial position while the other occupies an equatorial position (I–I–I angle equal to 90°), or both are equatorial (I–I–I angle equal to 120°). Fig. 8.38, which summarizes the result of the simple VSEPR assumption about the sequence of the strength of the electron pair repulsion, indicates unambiguously the first possibility. Thus, the VSEPR predicts a linear configuration, exactly what one finds in experiments.

---

[144] There are 24 electrons in the system, which is quite a lot for quantum mechanical methods. Meanwhile the VSEPR algorithm is *insensitive* to the number of electrons!

[145] The computed harmonic vibrational mode frequencies are 1537, 1453, and 849 cm^{-1}, which proves we have to do with a true minimum.

[146] T. Tanaka, Y. Morino, *Spectroscopy*, 33(1970)538.

What says the minimal quantum mechanical model about this? Let us optimize the geometry[147] starting from several distinct configurations of the nuclei.[148] We get the energy $-20\,552.059$ a.u., always corresponding to a symmetric linear configuration with the I–I distance equal to 2.94 Å. The experimental I–I distance is about 2.90 Å (and changes a bit depending on the accompanying cation in a salt). Therefore, the Hartree–Fock calculations (as well as the VSEPR!) gave the structure confirmed by experiments.

Example 12 (The infernal[149] ClF_3). The VSEPR algorithm gives (we choose chlorine as the central atom): $N = n_C + n_L + n_{ion} = 7 + 3 \cdot 1 + 0 = 10$. So, we have $P = 5$, and this means again a trigonal bipyramid (Fig. 8.38). The three chlorine atoms may be in the following configurations: (a) two axial and one equatorial; (b) one axial and two equatorial; or (c) three equatorial. The VSEPR algorithm is able to distinguish these three situations (Fig. 8.38) and indicates case (a) as being of the lowest energy. This however means a very strange molecule: planar and T-shaped. The roof of the T is formed by F–Cl–F lying on the straight line, while there is an additional fluorine protruding from the chlorine. Does this make any sense at all? Let us optimize the geometry within the Hartree–Fock method.[150] We get indeed a T-shaped molecule, only the roof deviates from the straight line, but not too much: $\measuredangle$ F–Cl–F $= 172.5°$ (the experimental value is 175.0°). The two fluorine atoms (of the roof of the T) are equivalent (the net charge computed by the Mulliken population analysis is -0.515, the distance $r_{ClF} = 1.67$ Å, its experimental value is 1.70 Å). The third fluorine atom differs (!): its net charge is -0.308, and $r_{ClF} = 1.58$ Å; the experimental value is 1.60 Å. When the optimization starts from the configurations (b) or (c) it ends up again in the strange configuration (a), which suggests that these structures (b and c) are simply nonstable. This underlines once more the success of the VSEPR algorithm, for a very unusual molecular shape.

8.11 The isolobal analogy

What is the electronic structure of the $Fe_2(CO)_8$ molecule? Not easy to say just from scratch. Maybe there are some multiple bonds between iron atoms of σ, π, or δ character? Well, the isolobal analogy will allow us to say the following: in fact $Fe_2(CO)_8$ is an ethylene-like molecule, with all consequences concerning the iron–iron bond (double), the planarity of the

---

[147] 160 electrons is quite a lot (not for the VSEPR algorithm though!). This is the reason why the calculations are carried out within the minimal basis set (STO-3G).

[148] This result is obtained from a starting point corresponding to a linear symmetric structure. A start from a bent structure (the angle about 120°) gave the same linear configuration. A start from a strongly bent structure (the angle equal to 90°) led to a dissociation into $I^- + I_2$ (of higher energy).

[149] When in contact with many organic substances it creates fire spontaneously and has many other vicious surprises.

[150] 6-31G(d) basis set.

$Fe_2(CO)_4$ fragment, etc. Also, we will be able to predict that the $Fe(CO)_4$ moiety resembles ("is isolobal with") simply CH_2, that it will bind to CH_2, forming $H_2C{=}Fe(CO)_4$, and that it can just replace CH_2 in many other instances in organic chemistry.

Now, a very short story about the isolobal analogy idea.[151]

The concept of localized MOs not only brought the quantum chemical approach much closer to the everyday chemists' reasoning, but also enabled them to see chemistry in a kind of holistic perspective, beyond any useless details.[152] It became evident that to form a chemical structure (e.g., a single chemical bond) it is sufficient to fulfill necessary conditions concerning AOs, like for instance to have two particular atoms, each offering a hybridized orbital protruding in space (of similar energy, "a lobe") and occupied by a single electron. It is not important how the hybridized AO looks in detail. It may have either an sp, sp^2, sp^3 orbital or, e.g., dsp^3, d^2sp^3, etc..[153] Such outermost crucial orbitals are known as *frontier orbitals* (they may be identified with the HOMO or LUMO molecular orbitals).

The isolobal analogy introduced by Roald Hoffmann says that

> the electronic structure – typically of inorganic d-electron complexes – can be rationalized by reducing it to the structures made by much simpler organic radicals like CH_3, CH_2, and CH, with *similar* frontier orbitals, and therefore similar chemical properties. More formally, two fragments (or molecules) are called isolobal if they have the same number of frontier orbitals of the same symmetry and electron occupation. So, isolobal molecular fragments will bind similarly and for similar reasons, while isolobal molecules will have similar electronic structures.

Let us consider some examples.

In Fig. 8.39 we see several isolobal chemical objects. Let us consider the methane molecule (upper left corner). By detaching consecutive hydrogen atoms one moves down the first column and the result are the radicals CH_3, CH_2, CH *with one, two, and three monooccupied lobes (together with the protons, all in the tetrahedral configuration)*, respectively. Now, consider the CrL_6 complex of chromium and six ligands L (first row). Each ligand binds to the chromium atom, because the chromium offers low-energy empty orbitals,[154] while ligands (like CO, NH_3,

---

[151] R. Hoffmann, *Nobel Lecture*, December 8, 1981.

[152] Stefan Banach: "*A good mathematician sees analogies among theorems, an outstanding one sees the analogies among theories, and a genius sees the analogies among analogies*."

[153] To see what such a hybride is, please, take a look of Fig. 4.25 on p. 240. If you add to what you see there (it is a linear combination of $3d_{z^2-x^2}$ and $3d_{z^2-y^2}$) a $3s$ and a $3p_z$ orbital, you increase one of the lobes and diminish the other one along z. The result is a lobe protruding towards $+z$.

[154] In this case six d^2sp^3-type hybrid lobes protruding towards the six corners of the tetragonal bipyramid.

etc.) provide an electron pair each. The CrL_6 complex has, therefore, the electronic configuration $4s^2 3d^4 L^{12}$, i.e., 18 valence electrons, which means the very stable Lewis-type closed shell of the argon atom. The next column begins by $Mn(CO)_3Cp$, where Cp means the cyclopentadienyl ligand offering five electrons. The electron count gives $4s^2 3d^5 (CO)^6 Cp^5$, again the stable system of 18 valence electrons.

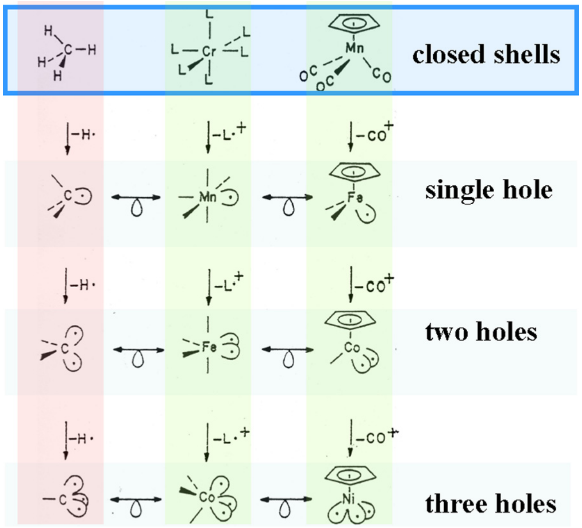


Fig. 8.39. The isolobal analogy between molecular fragments (symbolized by a double arrow with a loop). All species shown are electrically neutral and are derived conceptually from the first row that contains the "generic" compounds: CH_4, CrL_6, $Mn(CO)_3(C_5H_4)$, where L are ligands that offer an electron pair each. The second row is a result of an alchemical transmutation: we remove an electron pair together with a ligand L, but compensate it by increasing the atomic number of the central atom by 1 (its additional electron enters the empty lobe). In such a way all species in the second row have an orbital lobe carrying a single electron and each pair of them is isolobal. Similarly, in the third row all species share the same two-lobe structure (all are mutually isolobal), while in the fourth row we have the same three-lobe isolobal structures.

Now we will go down the columns (except the first column already discussed); we begin by CrL_6. Going towards the second row we *formally* remove one of the ligands L with its *one* electron from the lone pair (thus, we leave another *single electron on the protruding lobe*). In fact, we are doing this in a more reasonable way by removing L with *two* electrons, but at the same time replacing the central atom by the next one in the periodic table, i.e., by manganese. Therefore, the extra electron from the manganese atom (when compared to chromium) enters the lobe, which is thus singly occupied (see Fig. 8.39). We do the same in the third column (with the tetrahedral complex containing Cp). The electron configuration is $4s^2 3d^4 (\text{lobe})^1 L^{10}$, which means 17 electrons or one electron hole in the 18-electron shell. Going down once more (third row) we create two singly occupied lobes by the same type of "alchemical transmutation": removing L, but increasing the central atom atomic number by 1 (two holes in the 18-electron shell). Finally, we repeat this for the fourth row, receiving the three singly occupied lobes (three electron holes).

The philosophy behind the isolobal analogy is to see complex molecular structures as interrelated and systemized through a simple idea. It allows to predict the essence of the electronic structure just from basic knowledge about much simpler molecules, and go across the often too detailed quantum mechanical description, in other words to "understand" why the results *should be* like that. Understanding has direct consequences in chemical synthesis; the isolobal analogy helped a lot in producing new organometallic materials.

Summary

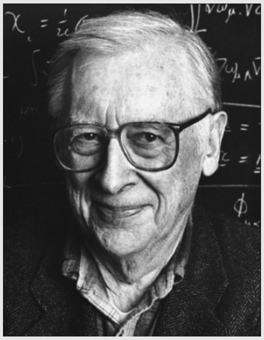
- The Hartree–Fock procedure represents a *variational* method. The variational function takes the form of *a single Slater determinant* ψ built of orthonormal molecular spin orbitals: $\psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \ldots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \ldots & \phi_2(N) \\ \ldots & \ldots & \ldots & \ldots \\ \phi_N(1) & \phi_N(2) & \ldots & \phi_N(N) \end{vmatrix}$.
- A molecular spin orbital $\phi_i(1)$ is a one-electron function of the coordinates x_1, y_1, z_1, σ_1. In the RHF method, it is the product $\varphi_i(x_1, y_1, z_1)\alpha(\sigma_1)$ or $\varphi_i(x_1, y_1, z_1)\beta(\sigma_1)$ of a real *molecular orbital* $\varphi_i(x_1, y_1, z_1)$ and of the *spin function* $\alpha(\sigma_1)$ or $\beta(\sigma_1)$, respectively. In the GHF method, a spin orbital is a complex function which depends both on $\alpha(\sigma_1)$ and on $\beta(\sigma_1)$. The UHF method uses, instead, real orbitals, which are all different and are multiplied either by α or β ("different orbitals for different spins").
- Minimization of the mean value of the Hamiltonian, $\varepsilon = \frac{\langle \psi | \hat{H}\psi \rangle}{\langle \psi | \psi \rangle}$, with respect to the orthonormal spin orbitals ϕ_i (GHF) leads to equations for *optimum spin orbitals* (Fock equations): $\hat{F}(1)\phi_i(1) = \varepsilon_i \phi_i(1)$, where the Fock operator $\hat{F}$ is $\hat{F}(1) = \hat{h}(1) + \hat{J}(1) - \hat{K}(1)$, the Coulombic operator is defined by $\hat{J}(1)u(1) = \sum_j \hat{J}_j(1)u(1)$ and $\hat{J}_j(1)u(1) = \int d\tau_2 \frac{1}{r_{12}} \phi_j^*(2)\phi_j(2)u(1)$, and the exchange operator is defined by $\hat{K}(1)u(1) = \sum_j \hat{K}_j(1)u(1)$ and $\hat{K}_j(1)u(1) = \int d\tau_2 \frac{1}{r_{12}} \phi_j^*(2)u(2)\phi_j(1)$.
- In the RHF method for closed shell systems, we assume *double occupancy of orbitals*, i.e., we form two spin orbitals out of each MO (by multiplying either by α or β).
- The Fock equations are solved by an iterative approach (with *an arbitrary* starting point) and as a result we obtain approximations to:
  - the total energy,
  - the wave function (the optimum Slater determinant),
  - the canonical MOs (spin orbitals),
  - the orbital energies.

- Use of the LCAO expansion leads to the Hartree–Fock–Roothaan equations $\mathbf{Fc} = \mathbf{Sc}\boldsymbol{\varepsilon}$. Our job is then to find the LCAO coefficients **c**. This is achieved by transforming the matrix equation to the form of the eigenvalue problem, and to diagonalize the corresponding Hermitian matrix. The canonical MOs obtained are LCAOs. The lowest-energy orbitals are occupied by electrons; those of higher energy are called virtual and are left empty.
- Using the H_2^+ and H_2 examples, we found that a chemical bond results from a quantum effect of an electron density flow towards the bond region. This results from a superposition of AOs due to the variational principle.
- In the simplest MO picture:
  - the *excited triplet state has lower energy than the corresponding excited singlet state*;
  - in case of orbital degeneracy, the system prefers *parallel electron spins* (Hund's rule);
  - the ionization energy is equal to the negative of the *orbital energy of the removed* electron; the electron affinity is equal to the negative of the *orbital energy corresponding to the virtual orbital accommodating the added electron* (Koopmans theorem).
- The canonical MOs for closed shell systems (the RHF method) may be transformed to orbitals localized in the chemical bonds, lone pairs, and inner shells.
- There are many methods of localization. The most important ones are the projection method, the method of minimum distance between two electrons from the same orbital (Boys approach), and the method of maximum interaction of electrons from the same orbital (Ruedenberg approach).
- Different localization methods lead to sets of localized molecular orbitals which are slightly different but their general shape is similar.
- The MOs (localized as well as canonical) can be classified as to the number of nodal surfaces going through the nuclei. A σ bond orbital has no nodal surface at all, a π bond orbital has a single nodal surface, and a δ bond orbital has two such surfaces.
- The localization allows for comparison of the molecular fragments of different molecules. It appears that the features of the MO localized on the A–B bond relatively weakly depend on the molecule in which this bond is found. This is a strong argument and a true source of experimental tactics in chemistry, which is to tune the properties of particular atoms by changing their neighborhood in a controlled way.
- Localization may serve to determine hybrids.
- In everyday practice, chemists have to use a minimal model of molecules that enables them to compare the geometry and vibrational frequencies with experiment to the accuracy of about 0.01 Å for bond lengths and about 1° for bond angles. This model assumes that the speed of light is infinite (nonrelativistic effects only), the Born–Oppenheimer approximation is valid (i.e., the molecule has a three-dimensional structure), the nuclei are bound by chemical bonds and vibrate (often harmonic vibrations are assumed), and the molecule moves (translation) and rotates as a whole in space.
- In many cases we can successfully predict the three-dimensional structure of a molecule by using a very simple tool: the Valence Shell Electron Pair Repulsion concept.
- One may view chemistry as interaction of molecular fragments, which are interrelated, systemized and simplified through a unifying idea of the isolobal analogy.

Main concepts, new terms

AMO method (p. 517)
antibonding orbital (p. 511)
atomic basis set (p. 504)
atomic orbital (p. 504)
atomic orbital dimension (p. 496)
basis sets (p. 504)
bonding orbital (p. 511)
Boys method (p. 551)
centering AO (p. 494)
chemical bond (p. 511)
closed shell (p. 464)
conditional extremum (p. 472)
Coulomb integral (p. 491)
Coulombic operator (p. 474)
digonal hybridization (p. 562)
Dirac notation (p. 469)
effectiveness of AO mixing (p. 503)
electron affinity (p. 537)
electronic configuration (p. 522)
electronic pair dimension (p. 558)
electronic shells (p. 522)
energy functional (p. 470)
exchange integral (p. 491)
exchange operator (p. 474)
excitation energy (p. 537)
external localization (p. 549)
Fock equation (p. 478)
Fukutome classes (p. 514)
Gaussian-type orbital (p. 501)
General Hartree–Fock (GHF) method (p. 478)
Hartree–Fock method (p. 478)
Hartree–Fock–Roothaan method (p. 506)
Hartree method (p. 488)
HOMO (p. 481)
Hund’s rule (p. 541)
hybridization (p. 562)
instability (p. 514)
internal localization (p. 550)
invariance with respect to a unitary transformation (p. 477)
ionization potential (p. 537)
Jabłoński diagram (p. 539)
Koopmans theorem (p. 544)
Lagrange multipliers (p. 472)
LCAO (p. 499)
localized MOs (p. 547)
LUMO (p. 481)
mean field (p. 486)
Mendeleev periodic table (p. 522)
minimal model of a molecule (p. 571)
molecular spin orbital (p. 464)
molecular orbital (p. 464)
occupied orbital (p. 481)
open shell (p. 482)
orbital centering (p. 501)
orbital localization (p. 547)
orbital size (p. 497)
penetration energy (p. 533)
Restricted Hartree–Fock (RHF) method (p. 480)
Ruedenberg method (p. 551)
Self-consistent field (p. 489)
Slater determinant (p. 465)
Slater orbital (p. 501)
Slater-type orbital (p. 501)
spin orbital (p. 464)
stationary state (p. 472)
σ, π, δ molecular orbitals (p. 555)
tetrahedral hybridization (p. 562)
trigonal hybridization (p. 562)
triplet instability (p. 515)
unitary transformation (p. 475)
Unrestricted Hartree–Fock (UHF) method (p. 478)
Valence Shell Electron Pair Repulsion (VSEPR) (p. 573)
variation of a spin orbital (p. 472)
virtual orbital (p. 481)

From the research front



John Pople (1925–2004), British mathematician and one of the founders of the modern quantum chemistry. His childhood was in difficult war time in England (every day 25-mile train journeys, sometimes under bombing). He came from a lower middle class family (drapers and farmers), but his parents were ambitious for the future of their children. At the age of twelve John developed an intense interest in mathematics. He entered Cambridge University after receiving a special scholarship. John Pople made important contributions to theoretical chemistry, including proposing semiempirical methods – the famous PPP method for π electron systems, the once very popular CNDO approach for all-valence calculations, and the monumental joint work on GAUSSIAN – a system of programs that constitutes one of the most important computational tools for quantum chemists. John Pople received the Nobel prize in 1998 "*for his development of computational methods in quantum chemistry*," sharing it with Walter Kohn. ©The Nobel Foundation.

The Hartree–Fock method belongs to a narrow two–three-member class of standard methods of quantum chemistry. It is the source of basic information about the electronic ground state of a molecule. It also allows for its geometry optimization. At present, the available computational codes limit the calculations to systems built of several hundreds of atoms. Moreover, the programs allow calculations to be made by clicking the mouse. The Hartree–Fock method is always at their core. The GAUSSIAN is one of the best-known programs. It is the result of many years of coding by several tens of quantum chemists working under John Pople. Pople was awarded the Nobel Prize in 1998 mainly for this achievement. To get a flavor of the kind of data needed, I provide below a typical data set necessary for GAUSSIAN to perform the Hartree–Fock computations for the water molecule:

```
#HF/STO-3G opt freq pop
water, the STO-3G basis set
0 1
O
H1 1 r12
H2 1 r12 2 a213
r12=0.96
a213=104.5
```

The explanatory comments, line by line:

- #HF/STO-3G opt freq pop is a command which informs GAUSSIAN that the computations are of the Hartree–Fock type, that the basis set used is of the STO-3G type (each STO is expanded into three GTOs), and that we want to optimize geometry (opt), compute the harmonic vibrational frequencies

(freq), and perform the charge population analysis for the atoms (known as Mulliken population analysis, see Appendix V2-D, p. V2-595).
- Just a comment line.
- 0 1 means that the total charge of the system is equal to 0, and the singlet state is to be computed (1).
- O means that the first atom in the list is oxygen.
- H1 1 r12 means that the second atom in the list is hydrogen named H1, distant from the first atom by r12.
- H2 1 r12 2 a213 means that the third atom in the list is hydrogen named H2, distant from atom number 1 by r12, and forming the 2-1-3 angle equal to a213.
- r12=0.96 is a starting O–H bond length in Å.
- a213=104.5 is a starting angle in degrees.

Similar inputs are needed for other molecules. The initial geometry is to some extent arbitrary, and therefore in fact it cannot be considered as real input data. The only true information is the number and charge (kind) of the nuclei, the total molecular charge (i.e., we know how many electrons are in the system), and the multiplicity of the electronic state to be computed. The basis set issue (STO-3G) is purely technical, and gives information about the quality of the results.

Ad futurum

Along with the development of computational techniques and with progress in the domain of electronic correlation, the importance of the Hartree–Fock method as a source of information about total energy, or total electron density, will most probably decrease. Simply, much larger molecules (beyond the Hartree–Fock level) will be within the reach of future computers. Yet Hartree–Fock calculations will still be carried out, and their results will be carefully analyzed. There are at least two reasons for this:

- Hartree–Fock calculations are most often the necessary step before more precise computations are performed.
- Hartree–Fock computations result in the MO model: the MOs and the orbital energies scheme ("minimal model"), and thus they provide the *conceptual framework for the molecule*. It is the sort of model which may be discussed, thought of, and used to search for explanations of physical and chemical phenomena. So far such a possibility does not exist for advanced methods, where often we obtain very good results, but it is extremely difficult to get an idea *why* they agree so well with experiments.[155]

Additional literature

A. Szabo, N.S. Ostlund, "Modern Quantum Chemistry," McGraw-Hill, New York, 1989, p. 108–231.

Excellent book.

---

[155] Solution of the Schrödinger equation, unfortunately, in most cases does not instruct us on the nature of physical phenomena.

T. Helgaker, P. Jørgensen, J. Olsen, "Molecular Electronic-Structure Theory," Wiley, Chichester, 2000, p. 433–513.

Very good book. It is a contemporary compendium of computational quantum chemistry.

"Localization and Delocalization in Quantum Chemistry," eds. O. Chalvet, R. Daudel, S. Diner, J.-P. Malrieu, D. Reidel Publish. Co., Dordrecht, 1975.

A set of very interesting articles by the leading quantum chemists of that time.

Questions

1. The Hartree–Fock method for a system with N electrons (closed shell case) leads to a wave function:
   a. which satisfies the Schrödinger equation.
   b. in the form of a Slater determinant, which in the Hilbert space is the closest to the solution of the Schrödinger equation.
   c. in the form of such a Slater determinant which gives the lowest mean value of the electronic Hamiltonian.
   d. in the form of an antisymmetrized product of the spin orbitals, each satisfying the Fock equation.
2. The canonical orbitals:
   a. represent a minimal basis set of the atomic orbitals of the atoms present in the molecule.
   b. when inserted into the Slater determinant give the lowest (among all orbitals possible) mean value of the Hamiltonian.
   c. satisfy the Fock equation.
   d. are either the core orbitals, or the bond orbitals, or the lone pair orbitals.
3. The localized molecular orbitals:
   a. are mutually orthogonal.
   b. when inserted into the Slater determinant give the lowest mean value of the Hamiltonian of all the Slater determinants possible.
   c. when transformed by a particular unitary transformation give the canonical orbitals.
   d. are localized on individual atoms of the molecule.
4. Localization of molecular orbitals:
   a. is done for determining which molecular orbitals are bonding, antibonding, and nonbonding.
   b. may lead to different sets of the localized orbitals.
   c. lowers the mean value of the Hamiltonian.
   d. enables one to obtain the orbitals of individual chemical bonds, of the lone pairs, and of the atomic core orbitals.
5. The orbital energy:
   a. for each molecular orbital multiplied by its occupancy and summed up over all molecular orbitals is equal to the total Hartree–Fock electronic energy.
   b. is equal to the mean value of the Fock operator with the corresponding molecular orbital.
   c. is a sum of the energies of the two electrons occupying the molecular orbital.
   d. is equal to the mean value of the energy per one electron of the molecule.

6. The Fock operator contains the following operators:
   a. of the kinetic energy of the nuclei.
   b. of the electrostatic attraction of the electron pairs by the nuclei.
   c. the Coulomb operator of a repulsion of a point-like electron with the charge density distribution of all electrons.
   d. the electrostatic attraction electron–nucleus.
7. In the LCAO MO method each molecular orbital represents:
   a. a function of position of an electron in three-dimensional space.
   b. a function that depends on the coordinates of the two electrons occupying this molecular orbital.
   c. a function of class Q.
   d. a linear combination of atomic orbitals (that belong to the atomic basis set chosen).
8. In the Hartree–Fock method (closed shell case, $U =$ a sum of the orbital energies of the doubly occupied molecular orbitals, V_{nn} stands for the repulsion of the nuclei) the total energy is:
   a. lower than $2U + V_{nn}$.
   b. equal to $2U$.
   c. equal to $2U + V_{nn}$.
   d. lower than $2U - \frac{1}{2}V_{ee}+V_{nn}$.
9. Comparing the RHF and UHF methods:
   a. in the UHF method one always gets a spin contamination of the wave function.
   b. the Slater determinant in the RHF method represents an eigenfunction of the operator of the square of the spin angular momentum.
   c. $E_{UHF} < E_{RHF}$.
   d. $E_{UHF} \leq E_{RHF}$, and both functions are eigenfunctions of the z component of the total spin of the electrons.
10. The RHF method for the hydrogen molecule (R is the internuclear distance):
   a. gives wrong results for large R, due to a HOMO-LUMO quasidegeneracy.
   b. gives a correct description of the dissociation producing two ground-state hydrogen atoms.
   c. gives wrong products of dissociation because of a too large difference between the bonding and antibonding energy levels.
   d. does not take into account any correlation of motion of the two electrons.

Answers

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

APPENDIX A

Reminding Matrices and Determinants

Matrices

Definition

A $n \times m$ matrix **A** represents a rectangular table of numbers[1] A_{ij} standing like soldiers in n perfect rows and m columns (index i tells us in which row, and index j in which column the number A_{ij} is located)

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nm} \end{pmatrix}.$$

Such a notation allows us to operate the whole matrices (like troops), instead of specifying what happens to each number ("soldier") separately. If matrices were not invented, then the equations would be very long and clumsy, instead of short and clear.

Addition

Two matrices **A** and **B** may be *added* if their dimensions n and m match. The result is matrix $\mathbf{C} = \mathbf{A} + \mathbf{B}$ (of the same dimensions as **A** and **B**), where each element of **C** is a sum of the corresponding elements of **A** and **B**, i.e.,

$$C_{ij} = A_{ij} + B_{ij},$$

e.g.,

$$\begin{pmatrix} 1 & -1 \\ -3 & 4 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ -5 & 7 \end{pmatrix}.$$

---

[1] If instead of the numbers a matrix contained functions, then everything below would remain valid (at particular values of the variables, instead of the functions we would have their values).

Multiplying by a number

A matrix may be multiplied by a number by multiplying every element of the matrix by this number: $c\mathbf{A} = \mathbf{B}$ with $B_{ij} = cA_{ij}$. For example, $2\begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ 6 & -4 \end{pmatrix}$.

Matrix product

A product of two matrices **A** and **B** is matrix **C**, denoted by $\mathbf{C} = \mathbf{AB}$, and its elements are calculated using elements of **A** and **B**, i.e.,

$$C_{ij} = \sum_{k=1}^{N} A_{ik} B_{kj},$$

where the number of columns (N) of matrix **A** has to be equal to the number of rows of matrix **B**. The resulting matrix **C** has the number of rows equal to the number of rows of **A** and the number of columns equal to the number of columns of **B**. Let us see how it works in an example. The product $\mathbf{AB} = \mathbf{C}$ is

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} & \mathbf{A}_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \mathbf{B}_{13} & B_{14} & B_{15} & B_{16} & B_{17} \\ B_{21} & B_{22} & \mathbf{B}_{23} & B_{24} & B_{25} & B_{26} & B_{27} \\ B_{31} & B_{32} & \mathbf{B}_{33} & B_{34} & B_{35} & B_{36} & B_{37} \\ B_{41} & B_{42} & \mathbf{B}_{43} & B_{44} & B_{45} & B_{46} & B_{47} \end{pmatrix} =$$

$$\begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} \\ C_{21} & C_{22} & \mathbf{C}_{23} & C_{24} & C_{25} & C_{26} & C_{27} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} & C_{37} \end{pmatrix}.$$

For example, C_{23} is the dot product of two vectors, or in matrix notation,

$$C_{23} = \begin{pmatrix} A_{21} & A_{22} & A_{23} & A_{24} \end{pmatrix} \cdot \begin{pmatrix} B_{13} \\ B_{23} \\ B_{33} \\ B_{43} \end{pmatrix} =$$

$$A_{21}B_{13} + A_{22}B_{23} + A_{23}B_{33} + A_{24}B_{43}.$$

Several remarks:

- The result of matrix multiplication depends in general on whether one has to multiply **AB** or **BA**, i.e.,[2] $\mathbf{AB} \neq \mathbf{BA}$.

---

[2] Although it may happen that $\mathbf{AB} = \mathbf{BA}$.

- Matrix multiplication satisfies the following relation (easy to check): $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$, i.e., the parentheses do not count and we can write simply $\mathbf{ABC}$.
- Often we will have multiplication of a square matrix $\mathbf{A}$ by a matrix $\mathbf{B}$ composed of one column. Then, using the rule of matrix multiplication we obtain the matrix $\mathbf{C}$ in the form of a single column (with the number of elements identical to the dimension of $\mathbf{A}$):

$$\begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \dots & \dots & \dots & \dots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \dots \\ B_m \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ \dots \\ C_m \end{pmatrix}.$$

Transposed matrix

For a given matrix $\mathbf{A}$ we may define the transposed matrix $\mathbf{A}^T$ defined as $\left(\mathbf{A}^T\right)_{ij} = \mathbf{A}_{ji}$.

For example, if $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -2 & 3 \end{pmatrix}$, then $\mathbf{A}^T = \begin{pmatrix} 1 & -2 \\ 2 & 3 \end{pmatrix}$.

If matrix $\mathbf{A} = \mathbf{BC}$, then $\mathbf{A}^T = \mathbf{C}^T\mathbf{B}^T$, i.e., the order of multiplication is reversed. Indeed, $(\mathbf{C}^T\mathbf{B}^T)_{ij} = \sum_k \left(\mathbf{C}^T\right)_{ik} \left(\mathbf{B}^T\right)_{kj} = \sum_k \mathbf{C}_{ki}\mathbf{B}_{jk} = \sum_k \mathbf{B}_{jk}\mathbf{C}_{ki} = (\mathbf{BC})_{ji} = \left(\mathbf{A}^T\right)_{ij}$.

Inverse matrix

For some square matrices $\mathbf{A}$ (which will be called nonsingular) we can define the so-called inverse matrix, denoted as $\mathbf{A}^{-1}$, which has the following property: $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{1}$, where $\mathbf{1}$ stands for the unit matrix:

$$\mathbf{1} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

For example, for the matrix $\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ we can find $\mathbf{A}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{pmatrix}$.

For square matrices $\mathbf{A1} = \mathbf{1A} = \mathbf{A}$.

If we cannot find $\mathbf{A}^{-1}$ (because it does not exist), then $\mathbf{A}$ is called a singular matrix. For example, the matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is singular. The inverse matrix for $\mathbf{A} = \mathbf{BC}$ is $\mathbf{A}^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}$. Indeed, $\mathbf{A}\mathbf{A}^{-1} = \mathbf{B}\mathbf{C}\mathbf{C}^{-1}\mathbf{B}^{-1} = \mathbf{B}\mathbf{1}\mathbf{B}^{-1} = \mathbf{B}\mathbf{B}^{-1} = \mathbf{1}$.

Adjoint, Hermitian, symmetric matrices

If the matrix $\mathbf{A}$ is transposed and in addition all its elements are changed to their complex conjugate, then we obtain the *adjoint matrix*, denoted as $\mathbf{A}^\dagger = \left(\mathbf{A}^T\right)^* = (\mathbf{A}^*)^T$. If for a square matrix we have $\mathbf{A}^\dagger = \mathbf{A}$, then $\mathbf{A}$ is called *Hermitian*. If $\mathbf{A}$ is real, then, of course, $\mathbf{A}^\dagger = \mathbf{A}^T$. If in addition for a real square matrix $\mathbf{A}^T = \mathbf{A}$, then $\mathbf{A}$ is called *symmetric*. Examples:

$$\mathbf{A} = \begin{pmatrix} 1+i & 3-2i \\ 2+i & 3-i \end{pmatrix} \quad \mathbf{A}^T = \begin{pmatrix} 1+i & 2+i \\ 3-2i & 3-i \end{pmatrix} \quad \mathbf{A}^\dagger = \begin{pmatrix} 1-i & 2-i \\ 3+2i & 3+i \end{pmatrix}.$$

Matrix $\mathbf{A} = \begin{pmatrix} 1 & -i \\ i & -2 \end{pmatrix}$ represents an example of a Hermitian matrix, because $\mathbf{A}^\dagger = \mathbf{A}$. Matrix $\mathbf{A} = \begin{pmatrix} 1 & -5 \\ -5 & -2 \end{pmatrix}$ is a symmetric matrix.

Unitary and orthogonal matrices

If for a square matrix $\mathbf{A}$ we have $\mathbf{A}^\dagger = \mathbf{A}^{-1}$, then $\mathbf{A}$ is called a *unitary* matrix. If $\mathbf{B}$ is Hermitian, then the matrix $\exp(i\mathbf{B})$ is unitary, where we define $\exp(i\mathbf{B})$ by using the Taylor expansion: $\exp(i\mathbf{B}) = \mathbf{1} + i\mathbf{B} + \frac{1}{2!}(i\mathbf{B})^2 + ...$ Indeed, $\left[\exp(i\mathbf{B})\right]^\dagger = \mathbf{1} - i\mathbf{B}^T + \frac{1}{2!}\left(-i\mathbf{B}^T\right)^2 + ... = \mathbf{1} - i\mathbf{B} + \frac{1}{2!}(-i\mathbf{B})^2 + ... = \exp(-i\mathbf{B})$, while $\exp(i\mathbf{B})\exp(-i\mathbf{B}) = \mathbf{1}$.

If $\mathbf{A}$ is a real unitary matrix $\mathbf{A}^\dagger = \mathbf{A}^T$, then it is called *orthogonal* with the property $\mathbf{A}^T = \mathbf{A}^{-1}$. For example, if $\mathbf{A} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, then $\mathbf{A}^T = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \mathbf{A}^{-1}$. Indeed, $\mathbf{A}\mathbf{A}^T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Determinants

Definition

For any square matrix $\mathbf{A} = \left\{A_{ij}\right\}$ we may calculate a number called its determinant and denoted by $\det\mathbf{A}$ or $|\mathbf{A}|$. The determinant is computed by using the Laplace expansion

$$\det\mathbf{A} = \sum_i^N (-1)^{i+j} A_{ij}\bar{A}_{ij} = \sum_j^N (-1)^{i+j} A_{ij}\bar{A}_{ij},$$

where (N is the dimension of the matrix) the result does not depend on which column j has been chosen in the first expression or which row i in the second expression. The symbol $\bar{A}_{ij}$

stands for the determinant of the matrix, which is obtained from $\mathbf{A}$ by removing the i-th row and the j-th column. Thus, we have defined a determinant (of dimension N) by saying that it is a certain linear combination of determinants (of dimension $N-1$). It is sufficient then to tell what we mean by the determinant that contains only one number c (i.e., having only one row and one column); this is simply $\det c \equiv c$.

For example, the matrix $\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 2 & 4 \\ 3 & -2 & -3 \end{pmatrix}$; and the determinant

$$\det\mathbf{A} = \begin{vmatrix} 1 & 0 & -1 \\ 2 & 2 & 4 \\ 3 & -2 & -3 \end{vmatrix} =$$

$$(-1)^{1+1} \times 1 \times \begin{vmatrix} 2 & 4 \\ -2 & -3 \end{vmatrix} + (-1)^{1+2} \times 0 \times \begin{vmatrix} 2 & 4 \\ 3 & -3 \end{vmatrix} + (-1)^{1+3} \times (-1) \times \begin{vmatrix} 2 & 2 \\ 3 & -2 \end{vmatrix} =$$

$$\begin{vmatrix} 2 & 4 \\ -2 & -3 \end{vmatrix} - \begin{vmatrix} 2 & 2 \\ 3 & -2 \end{vmatrix} = (2 \times (-3) - 4 \times (-2)) - (2 \times (-2) - 2 \times 3) = 2 + 10 = 12.$$

In particular, $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$

By repeating again and again (i.e., expanding $\bar{A}_{ij}$, etc.) the Laplace expansion we arrive finally at a linear combination of products of the elements,

$$\det\mathbf{A} = \sum_P (-1)^p \, \hat{P} \left[A_{1\mathbf{1}} A_{2\mathbf{2}} \ldots A_{N\mathbf{N}} \right],$$

where the summation goes over all permutations and the permutation operator $\hat{P}$ pertains to the second indices (shown in bold) and p is the parity of the permutation $\hat{P}$.

Slater determinant

In this book we will most often have to do with determinants of matrices whose elements are functions, not numbers. In particular, the most important will be the so-called Slater determinants. A Slater determinant for the N-electron system is built of the functions called spin orbitals $\phi_i\,(j)$, $i = 1, 2, \ldots, N$, where the *symbol* j means the space and spin coordinates $\left(x_j, y_j, z_j, \sigma_j\right)$ of electron j:

$$\psi\,(1, 2, \ldots, N) = \begin{vmatrix} \phi_1\,(1) & \phi_1\,(2) & \ldots & \phi_1\,(N) \\ \phi_2\,(1) & \phi_2\,(2) & \ldots & \phi_2\,(N) \\ \ldots & \ldots & \ldots & \ldots \\ \phi_N\,(1) & \phi_N\,(2) & \ldots & \phi_N\,(N) \end{vmatrix}.$$

Then the Laplace expansion gives

$$\psi(1,2,...,N)=\sum_{P}(-1)^{p}\,\hat{P}\left[\phi_1(1)\,\phi_2(2)\,...\phi_N(N)\right],$$

where the summation is over $N!$ permutations of the N electrons, $\hat{P}$ stands for the permutation operator that acts on the *arguments* of the product of the spin orbitals $[\phi_1(1)\,\phi_2(2)\,...\phi_N(N)]$, and p is the parity of the permutation $\hat{P}$ (i.e., the number of the transpositions that change $[\phi_1(1)\,\phi_2(2)\,...\phi_N(N)]$ into $\hat{P}\,[\phi_1(1)\,\phi_2(2)\,...\phi_N(N)]$).

All properties of determinants pertain also to the Slater determinants.

Some useful properties

- We have $\det\mathbf{A}^T=\det\mathbf{A}$.
- From the Laplace expansion it follows that if one of the spin orbitals is composed of two functions $\phi_i=\xi+\zeta$, then the Slater determinant is a sum of the two Slater determinants, one with ξ instead of ϕ_i, the second with ζ instead of ϕ_i.
- If we add to a row (column) any linear combination of other rows (columns), the value of the determinant does not change.
- If a row (column) is a linear combination of other rows (columns), then $\det\mathbf{A}=0$. In particular, if two rows (columns) are identical, then $\det\mathbf{A}=0$. Conclusion: in a Slater determinant the spin orbitals have to be linearly independent, otherwise the Slater determinant is equal to zero.
- If in a matrix $\mathbf{A}$ we exchange two rows (columns), then $\det\mathbf{A}$ changes the sign. Conclusion: the exchange of the coordinates of any two electrons leads to the change of the sign of the Slater determinant (Pauli exclusion principle).
- We have $\det(\mathbf{AB})=\det\mathbf{A}\quad\det\mathbf{B}$.
- From the Laplace expansion it follows that multiplying the determinant by a number is equivalent to multiplication of an arbitrary row (column) by this number. Therefore, $\det(c\mathbf{A})=c^N\det\mathbf{A}$, where N is the matrix dimension.[3]
- If matrix $\mathbf{U}$ is unitary, then $\det\mathbf{U}=\exp(i\phi)$, where ϕ is a real number. This means that if $\mathbf{U}$ is an orthogonal matrix, then $\det\mathbf{U}=\pm1$.

---

[3] Note that to multiply a matrix by a number we have to multiply every element of the matrix by this number. However, to multiply a determinant by a number means multiplication by this number of one row (column).

APPENDIX B

A Few Words on Spaces, Vectors, and Functions

Vector space

A vector space means a set V of elements $x, y, \dots$ (i.e., $x, y, \dots \in V$) that form an Abelian group (see Appendix C) and can be "added" together[1] and "multiplied" by numbers α, β, thus producing $z = \alpha x + \beta y, z \in V$. The multiplication ($\alpha, \beta$ are, in general, complex numbers) satisfies the usual rules:

$$
\begin{aligned}
&1 \cdot x = x,\\
&\alpha(\beta x) = (\alpha\beta)x,\\
&\alpha(x + y) = \alpha x + \alpha y,\\
&(\alpha + \beta)x = \alpha x + \beta x.
\end{aligned}
$$

Example 1 (Integers). The elements $x, y, \dots$ are integers, the "addition" means simply the usual addition of integers, the numbers α, β, ... are also integers, "multiplication" means just usual multiplication. Does the set of integers form a vector space? Let us see. The integers form a group (with addition as the operation in the group). Checking all the above axioms one easily proves that they are satisfied by integers. Thus, the integers (with the operations defined above) form a vector space.

Example 2 (Integers with real multipliers). If, in the previous example, we admitted α, β being real, then the multiplication of integers x, y by the real numbers would give real numbers (not necessarily integers). Therefore, in this case $x, y, \dots$ *do not* represent any vector space.

Example 3 (Vectors). Suppose $x, y, \dots$ are vectors, each represented by a N-element sequence of real numbers (they are called the vector "components") $x = (a_1, a_2, \dots a_N)$, $y = (b_1, b_2, \dots b_N)$, etc. Their addition $x + y$ is an operation that produces the vector $z = (a_1 + b_1, a_2 + b_2, \dots, a_N + b_N)$. The vectors form an Abelian group, because $x + y = y + x$, the unit ("neutral") element is $(0, 0, \dots 0)$, and the inverse element to $(a_1, a_2, \dots a_N)$ is equal to $(-a_1, -a_2, \dots - a_N)$. Thus, the vectors form a group. "Multiplication" of a vector by a real number α means $\alpha(a_1, a_2, \dots a_N) = (\alpha a_1, \alpha a_2, \dots \alpha a_N)$. Please check that the above four axioms are satisfied. Conclusion: the vectors form a vector space.

---

[1] See Appendix C; to form a group any pair of the elements can be "added" (operation in the group), the addition is associative, there exists a unit element, and to each element the inverse one exists.

Note that if only the positive vector components were allowed, then they would not form an Abelian group (no neutral element), and on top of that their addition (which might mean a subtraction of components, because α, β could be negative) could produce vectors with nonpositive components. Thus, the vectors with all positive components do not form a vector space.

Example 4 (Functions). This example is important in the scope of the book. This time the vectors have real components.[2] Their "addition" means the addition of two functions $f(x) = f_1(x) + f_2(x)$. "Multiplication" means multiplication by a real number. The unit ("neutral") function means $f = 0$, the "inverse" function to f is $-f(x)$. Therefore, the functions form an Abelian group. A few seconds are needed to show that the above four axioms are satisfied. Such functions form a vector space.

Linear independence. A set of vectors is called a set of linearly independent vectors if no vector of the set can be expressed as a linear combination of the other vectors of the set. The number of the linear independent vectors in a vector space is called the dimension of the space.

Basis. A set of n linearly independent vectors in the n-dimensional space.

Euclidean space

A vector space (with real multiplying numbers α, β) represents a Euclidean space if for any two vectors ϕ, ψ of the space we assign a real number called an *inner product* $\langle\phi|\psi\rangle$ with the following properties:

- $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle$,
- $\langle\alpha\phi|\psi\rangle = \alpha\,\langle\phi|\psi\rangle$,
- $\langle\phi_1 + \phi_2|\psi\rangle = \langle\phi_1|\psi\rangle + \langle\phi_2|\psi\rangle$,
- $\langle\phi|\phi\rangle = 0$, only if $\phi = 0$.

Inner product and distance

The concept of the inner product is used to introduce

- the *length of the vector* ϕ defined as $\|\phi\| \equiv \sqrt{\langle\phi|\phi\rangle}$, and
- the *distance between two vectors* ϕ and ψ as a nonnegative number $\|\phi - \psi\| = \sqrt{\langle\phi - \psi|\phi - \psi\rangle}$. The distance satisfies some conditions, which we treat as obvious from everyday experience:

---

[2] Note a similarity of the present example with the previous one: a function $f(x)$ may be treated as a vector with the infinite number of the components. The components are listed in the sequence of increasing $x \in R$, the component $f(x)$ corresponding to x.

- the distance from Cracow to Cracow has to be equal to zero (just insert $\phi = \psi$),
- the distance from Cracow to Warsaw has to be the same as from Warsaw to Cracow (just exchange $\phi \leftrightarrow \psi$),
- the Cracow–Warsaw distance is shorter than or equal to the sum of two distances: Cracow–X and X–Warsaw for any town X (a little more difficult to show).

Schwarz inequality

For any two vectors belonging to the Euclidean space the Schwarz inequality holds,[3] i.e.,

$$|\langle\phi|\psi\rangle| \leq \|\phi\| \, \|\psi\| \, , \tag{B.1}$$

or equivalently,

$$|\langle\phi|\psi\rangle|^2 \leq \|\phi\|^2 \, \|\psi\|^2 \, .$$

Orthogonal basis. All basis vectors ϕ_j, $j = 1, 2, ...N$, are orthogonal to each other, i.e., $\langle\phi_i|\phi_j\rangle = 0$ for $i \neq j$.

Orthonormal basis. An orthogonal basis set with all the basis vectors having length $\|\phi_i\| = 1$. Thus, for the orthonormal basis set we have $\langle\phi_i|\phi_j\rangle = \delta_{ij}$.

Example 5 (Dot product). Let us take the vector space from Example 3 and let us introduce the dot product (representing the inner product) defined as

$$\langle\phi|\psi\rangle = \sum_{i=1}^{N} a_i b_i \, . \tag{B.2}$$

Let us check whether this definition satisfies the properties required for an inner product:

- $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle$, because the order of a and b in the sum is irrelevant,
- $\langle\alpha\phi|\psi\rangle = \alpha \langle\phi|\psi\rangle$, because the sum says that multiplication of each a_i by α is equivalent to multiplying by α the inner product,
- $\langle\phi_1 + \phi_2|\psi\rangle = \langle\phi_1|\psi\rangle + \langle\phi_2|\psi\rangle$, because if the vector ϕ is decomposed into two vectors $\phi = \phi_1 + \phi_2$ in such a way that $a_i = a_{i1} + a_{i2}$ (with a_{i1}, a_{i2} being the components of ϕ_1, ϕ_2, respectively), then the summation of $\langle\phi_1|\psi\rangle + \langle\phi_2|\psi\rangle$ gives $\langle\phi|\psi\rangle$,
- $\langle\phi|\phi\rangle = \sum_{i=1}^{N} (a_i)^2$, and this is equal to zero if and only if all the components $a_i = 0$. Therefore, the proposed formula operates as the inner product definition requires.

---

[3] The Schwarz inequality agrees with what everyone recalls about the dot product of two vectors: $\langle x|y\rangle = \|x\| \, \|y\| \cos\theta$, where θ is the angle between the two vectors. Taking the absolute value of both sides we obtain $|\langle x|y\rangle| = \|x\| \, \|y\| \, |\cos\theta| \leq \|x\| \, \|y\|$.

Unitary space

If three changes are introduced in the definition of the Euclidean space, then we would obtain another space: the unitary space. These changes are as follows:

- the numbers α, β, ... instead of real are complex,
- the inner product instead of $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle$ has the property $\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$,
- instead of $\langle\alpha\phi|\psi\rangle = \alpha\,\langle\phi|\psi\rangle$ we have[4] $\langle\alpha\phi|\psi\rangle = \alpha^*\,\langle\phi|\psi\rangle$.

After the new inner product definition is introduced the related quantities, the length of a vector and the distance between the vectors, are defined in exactly the same way as in the Euclidean space. Also the definitions of the orthogonality and the Schwarz inequality remain unchanged.

Hilbert space

This is for us the most important unitary space – its elements are wave functions, which will be often denoted as $f, g, ..., \phi, \chi, \psi, ...$ etc. The wave functions with which we are dealing in quantum mechanics (according to John von Neumann) are the elements (i.e., vectors) of the Hilbert space. The inner product of two functions f and g means $\langle f|g\rangle \equiv \int f^* g\,\mathrm{d}\tau$, where the integration is over the whole space of variables, on which both functions depend. The length of vector f is denoted by $||f|| = \sqrt{\langle f|f\rangle}$. Consequently, the orthogonality of two functions f and g means $\langle f|g\rangle = 0$, i.e., an integral $\int f^* g\ \mathrm{d}\tau = 0$ over the whole range of the coordinates on which the function f depends. The Dirac notation (1.9) is in fact the inner product of such functions in a unitary space.

David Hilbert (1862–1943), German mathematician, professor at the University of Göttingen. At the II Congress of Mathematicians in Paris Hilbert formulated 23 goals of mathematics considered by him as very important. This had a great impact on mathematics and led to some unexpected results (e.g., Gödel theorem). Hilbert's investigations in 1900–1910 on integral equations resulted in the concept of the Hilbert space. Hilbert worked also on the foundations of mathematics, on mathematical physics, number theory, variational calculus, etc. This hard-working and extremely prolific mathematician became deeply depressed by Hitler's seizing of power. He regularly came to his office, but did not write any single sheet of paper.

---

[4] While still $\langle x|\alpha y\rangle = \alpha\,\langle x|y\rangle$.

Let us imagine an infinite sequence of functions (i.e., vectors) f_1, f_2, f_3, ... in a unitary space (Fig. B.1). The sequence will be called a Cauchy sequence if for a given $\varepsilon > 0$ a natural number N can be found such that for $i > N$ we will have $\| f_{i+1} - f_i \| < \varepsilon$. In other words, in a Cauchy sequence the distances between the consecutive vectors (functions) decrease when we go to sufficiently large indices, i.e., the functions become more and more similar to each other. *If the converging Cauchy sequences have their limits (functions) which belong to the unitary space, then such a space is called the Hilbert space*.

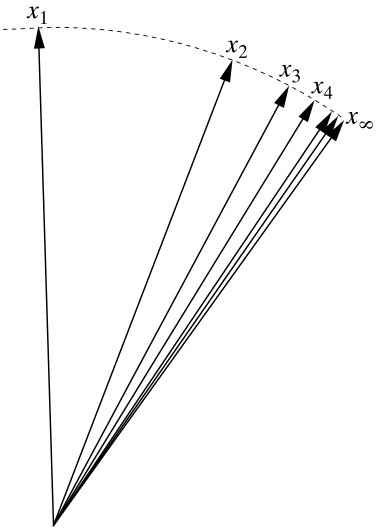


Fig. B.1. A pictorial representation of the Hilbert space. We have a vector space (each vector represents a wave function) and a series of unit vectors f_i that differ less and less (Cauchy series). If any convergent Cauchy series has its limit belonging to the vector space, then the space represents the *Hilbert space.*

A basis in the Hilbert space is such a set of the linearly independent functions (vectors) that any function belonging to the space can be expressed as a linear combination of the basis set functions. Because of the infinite number of dimensions the number of the basis set functions is infinite. This is difficult to imagine. In analogy with a three-dimensional Euclidean space we may imagine an orthonormal basis as the unit vectors protruding from the origin towards an infinite number of directions (like a “hedgehog,” Fig. B.2).

Each vector (function) can be represented as a linear combination of the “hedgehog” functions. It is seen that we may rotate the “hedgehog” (i.e., the basis set)[5] and the completeness of the

---

[5] The new orthonormal basis set is obtained by a unitary transformation of the old one.

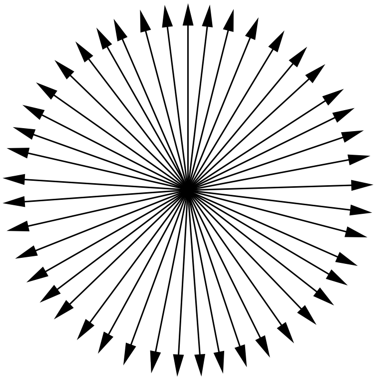


Fig. B.2. A pictorial representation of something that surely cannot be represented. Using *licencia poetica*: an orthonormal basis in the Hilbert space looks like a "hedgehog" of the unit vectors (their number equal to ∞), each pair of them orthogonal. This is in analogy to a two-dimensional or three-dimensional basis set, where the "hedgehog" has two or three orthogonal unit vectors.

basis will be preserved, i.e., any vector of the Hilbert space can be represented as a linear combination of the new basis set vectors.

Linear operator

Operator $\hat{A}$ transforms any vector ϕ from the operator's domain into vector ψ (both vectors ϕ, ψ belong to the unitary space): $\hat{A}(\phi) = \psi$, which is written as $\hat{A}\phi = \psi$. A *linear operator* satisfies $\hat{A}(c_1\phi_1 + c_2\phi_2) = c_1\hat{A}\phi_1 + c_2\hat{A}\phi_2$, where c_1 and c_2 stand for complex numbers.

We define:

- Sum of operators: $\hat{A} + \hat{B} = \hat{C}$ as $\hat{C}\phi = \hat{A}\phi + \hat{B}\phi$.
- Product of operators: $\hat{A}\hat{B} = \hat{C}$ as $\hat{C}\phi = \hat{A}(\hat{B}(\phi))$.

If for two operators we have[6] $\hat{A}\hat{B} = \hat{B}\hat{A}$, then we say they *commute*, or their *commutator* $\left[\hat{A}, \hat{B}\right] \equiv \hat{A}\hat{B} - \hat{B}\hat{A} = 0$. In general $\hat{A}\hat{B} \neq \hat{B}\hat{A}$, i.e., the operators do not commute.

- Inverse operator (if it exists): $\hat{A}^{-1}(\hat{A}\phi) = \phi$.

---

[6] When acting on any function of their domain.

Adjoint operator

If for an operator $\hat{A}$ we can find a new operator $\hat{A}^\dagger$ such that for any two vectors ϕ and ψ of the unitary space[7] we have[8]

$$\left\langle \phi | \hat{A}\psi \right\rangle = \left\langle \hat{A}^\dagger \phi | \psi \right\rangle, \tag{B.3}$$

then we say that $\hat{A}^\dagger$ is the *adjoint* operator to $\hat{A}$.

Hermitian operator

If $\hat{A}^\dagger = \hat{A}$, then the operator $\hat{A}$ will be called by us *self-adjoint* or *Hermitian* operator[9]:

$$\left\langle \phi | \hat{A}\psi \right\rangle = \left\langle \hat{A}\phi | \psi \right\rangle. \tag{B.4}$$

Unitary operator

A unitary operator $\hat{U}$ transforms a vector ϕ into $\psi = \hat{U}\phi$ both belonging to the unitary space (the domain is the unitary space) and the *inner product is preserved*:

$$\left\langle \hat{U}\phi | \hat{U}\psi \right\rangle = \langle \phi | \psi \rangle .$$

This means that any unitary transformation preserves the angle between the vectors ϕ and ψ, i.e., the angle between ϕ and ψ is the same as the angle between $\hat{U}\phi$ and $\hat{U}\psi$. The transformation preserves also the length of the vector, because $\left\langle \hat{U}\phi | \hat{U}\phi \right\rangle = \langle \phi | \phi \rangle$. This is why the operator $\hat{U}$ can be thought of as a *transformation related to a motion in the unitary space* (rotation, reflection, etc.). For a unitary operator we have $\hat{U}^\dagger \hat{U} = 1$, because $\left\langle \hat{U}\phi | \hat{U}\psi \right\rangle = \left\langle \phi | \hat{U}^\dagger \hat{U}\psi \right\rangle = \langle \phi | \psi \rangle$.

Eigenvalue equation

If for a particular vector ϕ we have

$$\hat{A}\phi = a\phi, \tag{B.5}$$

---

[7] The formal definition is less restrictive and the domains of the operators $\hat{A}^\dagger$ and $\hat{A}$ do not need to extend over the whole unitary space.

[8] Sometimes in the Dirac notation we make a useful modification: $\left\langle \phi | \hat{A}\psi \right\rangle \equiv \left\langle \phi | \hat{A} | \psi \right\rangle$.

[9] The self-adjoint and Hermitian operators differ in mathematics (the matter of domains), but we will ignore this difference in the present book.

where a is a complex number and $\phi \neq 0$, then ϕ is called an eigenvector[10] of the operator $\hat{A}$ corresponding to the eigenvalue a. The operator $\hat{A}$ may have a finite (including zero) or infinite number of eigenvalues, labeled by the subscript i:

$$\hat{A}\phi_i = a_i\phi_i.$$

Hermitian operators have the following important properties:

> If $\hat{A}$ represents a Hermitian operator, its eigenvalues a_i are real numbers, and its eigenvectors ϕ that correspond to different eigenvalues are orthogonal.

The number of linear independent eigenvectors that correspond to a given eigenvalue a is called the degree of degeneracy of the eigenvalue. Such vectors form a basis of the *invariant space of the operator* $\hat{A}$, i.e., any linear combination of the vectors represents a vector that is also an eigenvector (with the same eigenvalue a). If the eigenvectors correspond to different eigenvalues, then their linear combination *is not* an eigenvector of $\hat{A}$. Both things need a few seconds to be shown.[11]

One can show that the eigenvectors of a Hermitian operator form the complete basis set[12] in the Hilbert space, i.e., any function of class[13] Q can be expanded in a linear combination of the basis set.

---

[10] In quantum mechanics the vector ϕ will correspond to a function (a vector in the Hilbert space) and therefore is called the eigenfunction.

[11] We have the eigenvalue problem $\hat{A}\phi = a\phi$. Taking complex conjugate of both sides we obtain $(\hat{A}\phi)^* = a^*\phi^*$. Multiplying the first of the equations by ϕ^* and integrating, and then using ϕ and doing the same with the second equation, we get $\left\langle \phi|\hat{A}\phi \right\rangle = a\left\langle \phi|\phi \right\rangle$ and $\left\langle \hat{A}\phi|\phi \right\rangle = a^*\left\langle \phi|\phi \right\rangle$. But $\hat{A}$ is Hermitian, and therefore the left-hand sides of both equations are equal. Subtracting them we have $(a - a^*)\left\langle \phi|\phi \right\rangle = 0$. Since $\left\langle \phi|\phi \right\rangle \neq 0$, because $\phi \neq 0$, $a = a^*$. This is what we wanted to show.
The orthogonality of the eigenfunctions of a Hermitian operator (corresponding to different eigenvalues) may be proved as follows. We have $\hat{A}\phi_1 = a_1\phi_1$, $\hat{A}\phi_2 = a_2\phi_2$, with $a_1 \neq a_2$. Multiplying the first equation by ϕ_2^* and integrating one obtains $\left\langle \phi_2|\hat{A}\phi_1 \right\rangle = a_1\left\langle \phi_2|\phi_1 \right\rangle$. Then, let us take the complex conjugate of the second equation, $\left(\hat{A}\phi_2\right)^* = a_2\phi_2^*$, where we have used that $a_2 = a_2^*$ (this has been proved a while ago). Then, let us multiply it by ϕ_1 and integrate: $\left\langle \hat{A}\phi_2|\phi_1 \right\rangle = a_2\left\langle \phi_2|\phi_1 \right\rangle$. Subtracting the two equations we have $0 = (a_1 - a_2)\left\langle \phi_2|\phi_1 \right\rangle$, and taking into account that $a_1 - a_2 \neq 0$ this gives $\left\langle \phi_2|\phi_1 \right\rangle = 0$.

[12] This basis set may be assumed to be orthonormal, because the eigenfunctions:
- as square integrable can be normalized,
- if they correspond to different eigenvalues, are automatically orthogonal,
- if they correspond to the same eigenvalue, they can be orthogonalized (still remaining eigenfunctions) by a method described in Appendix K.

[13] That is, continuous, single-valued, and square integrable (see Fig. 2.6).

Sometimes an eigenvector ϕ of the operator $\hat{A}$ (with the eigenvalue a) is subject to an operator $f(\hat{A})$, where f is an analytic function. Then[14]

$$f(\hat{A})\phi = f(a)\phi. \tag{B.6}$$

Commutation and eigenvalues

We will use sometimes the theorem that if two linear and Hermitian operators $\hat{A}$ and $\hat{B}$ commute, then they have a common set of eigenfunctions and *vice versa*.

We will prove this theorem in the case of no degeneracy (i.e., there is only one linearly independent vector). We have an eigenvalue equation $\hat{B}\psi_n = b_n\psi_n$. Applying to both sides the operator $\hat{A}$ and using the commutation relation $\hat{A}\hat{B} = \hat{B}\hat{A}$ we have $\hat{B}\left(\hat{A}\psi_n\right) = b_n\left(\hat{A}\psi_n\right)$. This means that $\hat{A}\psi_n$ is an eigenvector of $\hat{B}$ corresponding to the eigenvalue b_n. But, we know already such a vector, this is ψ_n. No way, the two vectors have to be proportional: $\hat{A}\psi_n = a_n\psi_n$, which means that ψ_n is an eigenvector of $\hat{A}$.

Now, the inverse theorem. We have two operators and any eigenvector of $\hat{A}$ is also an eigenvector of $\hat{B}$. We want to show that the two operators commute. Let us write the two eigenvalue equations: $\hat{A}\psi_n = a_n\psi_n$ and $\hat{B}\psi_n = b_n\psi_n$. Let us take a vector ϕ. Since the eigenvectors $\{\psi_n\}$ form the complete set, we have

$$\phi = \Sigma_n c_n \psi_n.$$

Applying the commutator $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$ we have

$$\begin{gathered}[\hat{A}, \hat{B}]\phi = \hat{A}\hat{B}\phi - \hat{B}\hat{A}\phi = \hat{A}\hat{B}\Sigma_n c_n\psi_n - \hat{B}\hat{A}\Sigma_n c_n\psi_n = \hat{A}\Sigma_n c_n\hat{B}\psi_n - \hat{B}\Sigma_n c_n\hat{A}\psi_n \\ = \hat{A}\Sigma_n c_n b_n\psi_n - \hat{B}\Sigma_n c_n a_n\psi_n = \Sigma_n c_n b_n\hat{A}\psi_n - \Sigma_n c_n a_n\hat{B}\psi_n = \\ \Sigma_n c_n b_n a_n\psi_n - \Sigma_n c_n a_n b_n\psi_n = 0.\end{gathered}$$

This means that the two operators commute.

---

[14] The operator $f(\hat{A})$ is defined through the Taylor expansion of the function f: $f(\hat{A}) = c_0 + c_1\hat{A} + c_2\hat{A}^2 + \ldots$. If now the operator $f(\hat{A})$ acts on an eigenfunction of $\hat{A}$, then, because $\hat{A}^n x = a^n x$, we obtain the result.

APPENDIX C

Group Theory in Spectroscopy

The group theory in this textbook will be treated in a practical way, as one of the useful tools.[1]

Quite a lot of what we will be talking about in this appendix was invented by Evariste Galois. He was only 21 when he died in a duel (*cherchez la femme*!). Galois spent his last night writing down his group theory.
Evariste Galois (1811–1832), a French mathematician, created also many fundamental ideas in the theory of algebraic equations.

Our goal will be to predict the selection rules in the ultraviolet (UV), visual (VIS), and infrared (IR) molecular spectra.

We will try to be concise, but examples need explanations; there are few amateurs of dry formulae.

Group

Imagine a set of elements $\hat{R}_i$, $i = 1, 2, \ldots g$. We say that they form a group G of the order[2] g if the following four conditions are satisfied:

1. There exists an operation called "*multiplication*," $\hat{R}_i \cdot \hat{R}_j$, which associates with every pair of the elements of G another element of G, i.e., $\hat{R}_i \cdot \hat{R}_j = \hat{R}_k$. Hereafter the multiplication

---

[1] Rather than a field of abstract mathematics. Symmetry may be viewed either as something beautiful or primitive. It seems that from a psychological point of view symmetry stresses people's longing for simplicity, order, and understanding. On the other hand symmetry means less information and hence often a kind of wearingly dull stimuli. Possibly an interplay of these two opposite features leads us to consider as beautiful what has a *broken* symmetry. Happily enough, the trees and the leaves exhibit broken symmetry and look beautiful. Ancient architects knew secrets of creating beautiful buildings, which relied on symmetry breaking, substantial but almost invisible from a distance.

[2] g may be finite or infinite. In practical applications of the present appendix, g will be always finite.

$\hat{R}_i \cdot \hat{R}_j$ will be denoted simply as $\hat{R}_i \hat{R}_j$. Thus, the elements can multiply each other and the result always belongs to the group.

2. The multiplication is associative,[3] i.e., for any three elements of G we have $\hat{R}_i(\hat{R}_j \hat{R}_k) = (\hat{R}_i \hat{R}_j)\hat{R}_k$.
3. Among $\hat{R}_i \in G$ there exists an *identity* element, denoted by $\hat{E}$, having a nice property: $\hat{R}_i \hat{E} = \hat{R}_i$ and $\hat{E}\hat{R}_i = \hat{R}_i$ for any i.
4. To each $\hat{R}_i$ we can find such an element of G (denoted as $\hat{R}_i^{-1}$, called the *inverse* element with respect to $\hat{R}_i$) that $\hat{R}_i \hat{R}_i^{-1} = \hat{E}$, also $\hat{R}_i^{-1}\hat{R}_i = \hat{E}$.

Example 1 (A four-element group). The elements $1, -1, i, -i$, with the operation chosen as the ordinary multiplication of numbers, form a group of order 4. Indeed, any product of these numbers gives one of them. Here is the corresponding "multiplication table":

		Second in the product			
		1	**−1**	**i**	**−i**
First in the product	1	1	−1	i	−i
	−1	−1	1	−i	i
	i	i	−i	−1	1
	−i	−i	i	1	−1

Note that

> ABELIAN GROUP
>
> the table is *symmetric* with respect to the diagonal. A group with symmetric multiplication table is called Abelian.

The associativity requirement is of course satisfied. The unit element is 1. You can always find an inverse element. Indeed, for 1 it is 1, for −1 it is −1, for i it is $-i$, for $-i$ it is i. Thus, all conditions are fulfilled and $g = 4$.

Example 2 (Group of integers). Let us take as G the set of integers with the "multiplication" being the regular addition of numbers. Let us check. The sum of two integers is an integer, so requirement 1 is satisfied. The operation is associative, because the addition is. The unit element is, of course, 0. The inverse element to an integer means the opposite number. Thus, G is a group of order $g = \infty$.

---

[3] Thanks to that the expressions similar to $\hat{R}_i \hat{R}_j \hat{R}_k$ have unambiguous meaning.

Example 3 (Group of nonsingular matrices). All nonsingular $n \times n$ matrices[4] with matrix multiplication as the operation form a group. Let us see. Multiplication of a nonsingular matrix $\mathbf{A}$ (i.e., $\det\mathbf{A} \neq 0$) by a nonsingular matrix $\mathbf{B}$ gives a nonsingular matrix $\mathbf{C} = \mathbf{AB}$, because $\det\mathbf{C} = \det\mathbf{A}\det\mathbf{B} \neq 0$. The unit element is the unit matrix $\mathbf{1}$, the inverse element exists (this is why we needed the nonsingularity) and is equal to $\mathbf{A}^{-1}$. Also from the matrix multiplication rule we have $(\mathbf{AB})\,\mathbf{C} = \mathbf{A}\,(\mathbf{BC})$. This is a group of the order ∞.

Example 4 (Group of unitary matrices U(n)). In particular, all the unitary $n \times n$ matrices form a group with matrix multiplication as the group multiplication operation. Let us check. Any such multiplication is feasible and the product represents a unitary matrix (if matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ are unitary, i.e., $\mathbf{U}_1^{\dagger} = \mathbf{U}_1^{-1}$ and $\mathbf{U}_2^{\dagger} = \mathbf{U}_2^{-1}$, then $\mathbf{U} = \mathbf{U}_1\mathbf{U}_2$ is also unitary, because $\mathbf{U}^{-1} = \mathbf{U}_2^{-1}\mathbf{U}_1^{-1} = \mathbf{U}_2^{\dagger}\mathbf{U}_1^{\dagger} = (\mathbf{U}_1\mathbf{U}_2)^{\dagger} = \mathbf{U}^{\dagger}$), matrix multiplication is associative, the identity element means the $n \times n$ unit matrix, and the inverse matrix $\mathbf{U}^{-1} = \mathbf{U}^{\dagger} \equiv \left(\mathbf{U}^{T}\right)^{*}$ always exists. This group is called in physics U(n).

Example 5 (SU(n) group). The (famous in physics) group SU(n) for $n \geq 2$ is defined as the subset of U(n) of such matrices $\mathbf{U}$ that $\det\mathbf{U} = 1$ with the same multiplication operation.[5] Indeed, since $\det(\mathbf{U}_1\mathbf{U}_2) = \det\mathbf{U}_1 \det\mathbf{U}_2$, multiplication of any two elements of SU(n) gives an element of SU(n). Also of great importance in physics is the SO(n) group, that is, the SU(n) group with real (i.e., orthogonal) matrices.

Unitary versus symmetry operation

Let us take the so-called $SO(3)$ group of all rotations of the coordinate system in 3D (the *Cartesian three-dimensional Euclidean space,* see Appendix B, p. 595). The rotation operators acting in this space will be denoted by $\hat{R}$ and defined as follows: the operator $\hat{R}$ acting on a vector $\mathbf{r}$ produces the vector $\hat{R}\mathbf{r}$:

$$\hat{R}\mathbf{r} = \mathbf{R}\mathbf{r}, \tag{C.1}$$

where[6] $\mathbf{R}$ represents an orthogonal matrix of dimension 3. The orthogonality guarantees that the transformation preserves the dot product (and therefore vectors' lengths as well).

---

[4] See Appendix A.

[5] Recall (Appendix A) that for a unitary matrix $\mathbf{U}$ one has $\det\mathbf{U} = \exp(i\phi)$. For orthogonal matrices (i.e., the unitary ones with all the elements real) $\det\mathbf{U} = \pm 1$. This does not mean that the SU(n) is composed of the orthogonal matrices only. For example, all the four 2×2 matrices $\left\{\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}\right\}, \left\{\begin{matrix} -1 & 0 \\ 0 & -1 \end{matrix}\right\}, \left\{\begin{matrix} 0 & i \\ i & 0 \end{matrix}\right\}, \left\{\begin{matrix} 0 & -i \\ -i & 0 \end{matrix}\right\}$ have the determinants equal to 1 and belong to SU(2), while only the first two belong to SO(2).

[6] The point in three-dimensional space is indicated by the vector $\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$.

Let us take an arbitrary function $f(\mathbf{r})$ of position $\mathbf{r}$. Now, for each of the operators $\hat{R}$ let us construct the corresponding operator $\hat{\mathcal{R}}$ that *moves in space the function without its deformation*. Generally, we obtain another function, which means that $\hat{\mathcal{R}}$ *operates in the Hilbert space*. The construction of the operator $\hat{\mathcal{R}}$ is based on the following prescription:

$$\hat{\mathcal{R}} f(\mathbf{r}) = f(\hat{R}^{-1}\mathbf{r}). \tag{C.2}$$

This means that displacement in space of the function $f(\mathbf{r})$ is equivalent simply to leaving the function intact, but instead the *inverse displacement of the coordinate system*.[7]

The operators $\hat{\mathcal{R}}$ rotate functions without their deformation; therefore they preserve the scalar products in the Hilbert space and are unitary. They form a group isomorphic with the group of operators $\hat{R}$, because they have the same multiplication table as the operators $\hat{R}$: if $\hat{R} = \hat{R}_1\hat{R}_2$, then $\hat{\mathcal{R}} = \hat{\mathcal{R}}_1\hat{\mathcal{R}}_2$, where $\hat{\mathcal{R}}_1 f(\mathbf{r}) = f(\hat{R}_1^{-1}\mathbf{r})$ and $\hat{\mathcal{R}}_2 f(\mathbf{r}) = f(\hat{R}_2^{-1}\mathbf{r})$. Indeed[8], $\hat{\mathcal{R}} f = (\hat{\mathcal{R}}_1\hat{\mathcal{R}}_2) f(\mathbf{r}) = f(\hat{R}_2^{-1}\hat{R}_1^{-1}\mathbf{r}) = f(\hat{R}^{-1}\mathbf{r})$.

UNITARY VERSUS SYMMETRY OPERATION

A unitary operation is a symmetry operation of the function $f(\mathbf{r})$, when $\hat{\mathcal{R}} f(\mathbf{r}) = f(\mathbf{r})$.

Example 6 (Rotation of a point). The operator $\hat{R}(\alpha; z)$ of rotation of the point with the coordinates x, y, z by angle α about the z axis gives the point with the coordinates x', y', z' (Fig. C.1a), i.e.,

$$\begin{aligned} x' &= r\cos(\phi+\alpha) = r\cos\phi\cos\alpha - r\sin\phi\sin\alpha = x\cos\alpha - y\sin\alpha, \\ y' &= r\sin(\phi+\alpha) = r\sin\phi\cos\alpha + r\cos\phi\sin\alpha = x\sin\alpha + y\cos\alpha, \\ z' &= z. \end{aligned}$$

---

[7] Motion is relative. Let us concentrate on a rotation by angle α. The result is the same if:

- the coordinate system stays still, but the point rotates by angle α, or
- the point does not move, while the coordinate system rotates by angle $-\alpha$.

What if function $f(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N)$ is to be rotated? Then we will do the following: $\hat{\mathcal{R}} f(\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_N) = f(\hat{R}^{-1}\mathbf{r}_1, \hat{R}^{-1}\mathbf{r}_2, \ldots \hat{R}^{-1}\mathbf{r}_N)$.

[8] This result is correct, but the routine of notation works here in a quite misleading way when suggesting that $(\hat{\mathcal{R}}_1\hat{\mathcal{R}}_2) f(\mathbf{r})$ and $f(\hat{R}_1^{-1}\hat{R}_2^{-1}\mathbf{r})$ means the same. The correct result is derived in the following way. First, from the definition we have $\hat{\mathcal{R}}_2 f(\mathbf{r}) = f\left(R_2^{-1}\mathbf{r}\right) \equiv g_2(\mathbf{r})$. Then, we get $(\hat{\mathcal{R}}_1\hat{\mathcal{R}}_2) f(\mathbf{r}) = \hat{\mathcal{R}}_1[\hat{\mathcal{R}}_2 f(\mathbf{r})] = \hat{\mathcal{R}}_1 g_2(\mathbf{r}) = g_2\left(R_1^{-1}\mathbf{r}\right) = \hat{\mathcal{R}}_2 f\left(R_1^{-1}\mathbf{r}\right) = f\left(R_2^{-1}R_1^{-1}\mathbf{r}\right)$.

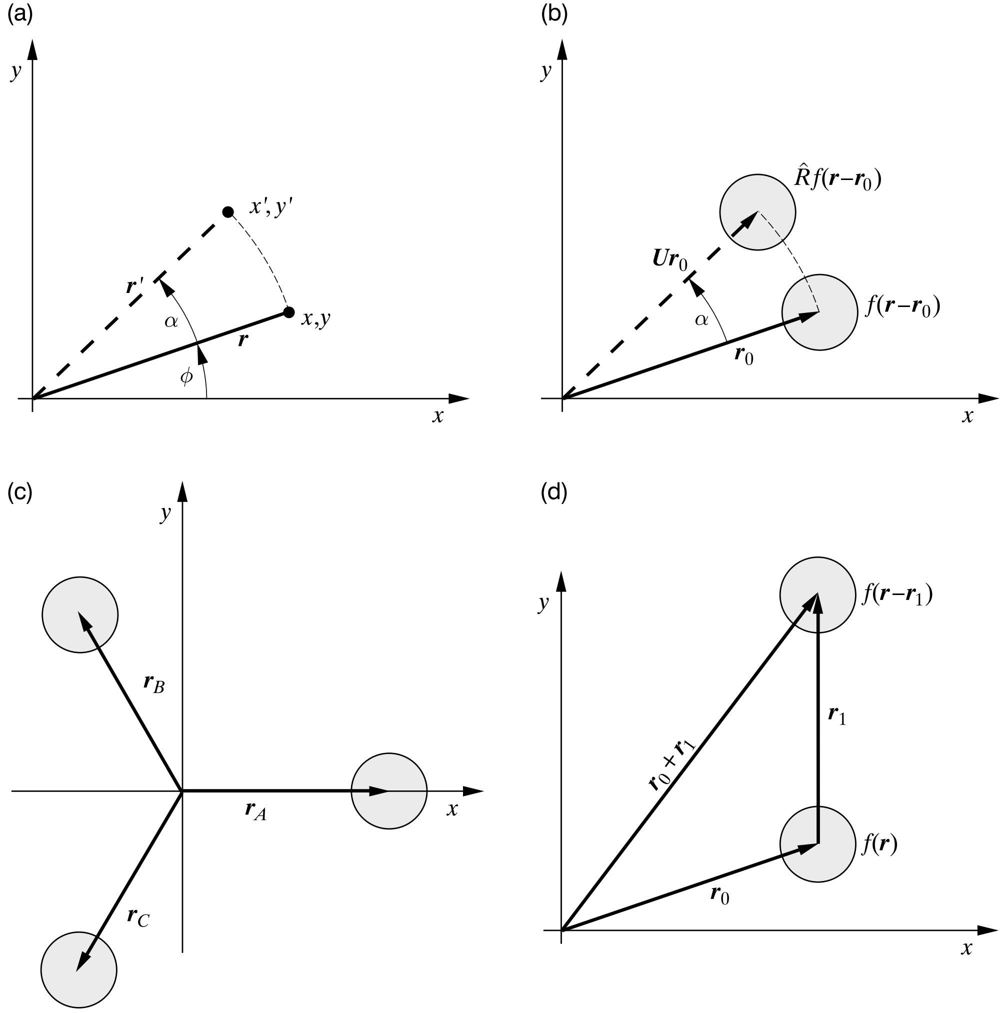


Fig. C.1. Examples of an isometric operation. (a) Unitary operation: rotation of a point by angle α about the z axis. The old position of the point is indicated by the vector $\mathbf{r}$, the new position by $\mathbf{r}'$ (of the same length). (b) Unitary operation: rotation of the function $f(\mathbf{r}-\mathbf{r}_0)$ by the angle α about the z axis. As a result we have the function $f(\mathbf{r}-\mathbf{U}\mathbf{r}_0)$, which in general represents a function that differs from $f(\mathbf{r}-\mathbf{r}_0)$. (c) The unitary operation that represents a symmetry operation: rotation by the angle $\alpha = 120°$ of the function $f(\mathbf{r}) = \exp\left[-|\mathbf{r}-\mathbf{r}_A|^2\right] + \exp\left[-|\mathbf{r}-\mathbf{r}_B|^2\right] + \exp\left[-|\mathbf{r}-\mathbf{r}_C|^2\right]$, where the vectors $\mathbf{r}_A, \mathbf{r}_B, \mathbf{r}_C$ are of the same length and form the Mercedes trademark (the angle $120°$). The new function is identical to the old one. (d) Translational operator by the vector $\mathbf{r}_1$: $\hat{\mathcal{R}}(\mathbf{r}_1)$ applied to the Gaussian function $f(\mathbf{r}) = \exp\left[-|\mathbf{r}-\mathbf{r}_0|^2\right]$ gives $\hat{\mathcal{R}}(\mathbf{r}_1)\, f(\mathbf{r}) = f(\hat{R}^{-1}\mathbf{r}) = \exp\left[-\left|\hat{R}^{-1}\mathbf{r}-\mathbf{r}_0\right|^2\right] = \exp\left[-|\mathbf{r}-\mathbf{r}_1-\mathbf{r}_0|^2\right] = \exp\left[-|\mathbf{r}-(\mathbf{r}_1+\mathbf{r}_0)|^2\right] = f(\mathbf{r}-\mathbf{r}_1)$, i.e., the function shifted in space by the vector $\mathbf{r}_1$ with respect to the original function.

Therefore, the corresponding transformation matrix of the old to the new coordinates is

$$\mathbf{U} = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We obtain the same new coordinates if the point stays still, while the coordinate system rotates in the opposite direction (i.e., by the angle $-\alpha$).

Example 7 (Rotation of an atomic orbital). Let us construct a single spherically symmetric Gaussian orbital $f(\mathbf{r}) = \exp\left[-|\mathbf{r} - \mathbf{r}_0|^2\right]$ in the Hilbert space for one electron. Let the atomic orbital be centered in the point indicated by the vector $\mathbf{r}_0$. The operator $\hat{\mathcal{R}}(\alpha; z)$ has to carry out rotation of a *function*[9] by the angle α about the z axis (Fig. C.1b), which corresponds to a rotation in the Hilbert space.[10] According to the definition of a rotation, what we need is $\hat{\mathcal{R}} f(\mathbf{r}) = f\left(\hat{R}^{-1}\mathbf{r}\right)$. Since the operator $\hat{R}$ corresponds to the matrix $\mathbf{U}$, $\hat{R}^{-1}$ corresponds to $\mathbf{U}^{-1}$. The last matrix is simply

$$\mathbf{U}^{-1} = \mathbf{U}^T = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We obtain the following chain of transformations:

$$f(\hat{R}^{-1}\mathbf{r}) = \exp\left[-\left|\hat{R}^{-1}\mathbf{r} - \mathbf{r}_0\right|^2\right] = \exp\left[-\left|\hat{R}^{-1}\mathbf{r} - \hat{R}^{-1}\hat{R}\mathbf{r}_0\right|^2\right] =$$

$$\exp\left[-\left\langle \hat{R}^{-1}\mathbf{r} - \hat{R}^{-1}\hat{R}\mathbf{r}_0 | \hat{R}^{-1}\mathbf{r} - \hat{R}^{-1}\hat{R}\mathbf{r}_0\right\rangle\right] = \exp\left[-\left\langle \hat{R}\hat{R}^{-1}\mathbf{r} - \hat{R}\hat{R}^{-1}\hat{R}\mathbf{r}_0 | \mathbf{r} - \hat{R}\mathbf{r}_0\right\rangle\right] =$$

$$\exp\left[-\left\langle \mathbf{r} - \hat{R}\mathbf{r}_0 | \mathbf{r} - \hat{R}\mathbf{r}_0\right\rangle\right] = \exp\left[-\left|\mathbf{r} - \hat{R}\mathbf{r}_0\right|^2\right].$$

Thus, the center of the orbital underwent the rotation and therefore $\hat{\mathcal{R}} f(\mathbf{r})$ represents indeed the spherically symmetric orbital[11] displaced by angle α.

---

[9] This orbital represents our *object* to rotate by α. The coordinate system rests unchanged, while the object moves. The job will be done by the operator $\hat{\mathcal{R}}(\alpha; z)$.

[10] We will obtain another (because centered differently) function.

[11] The definition $\hat{\mathcal{R}} f(\mathbf{r}) = f(\hat{R}^{-1}\mathbf{r})$ can transform anything: from the spherically symmetric Gaussian orbital through a molecular orbital (please recall it can be represented by the LCAO expansion) till the Statue of Liberty! Indeed, do you want to rotate the Statue of Liberty? If not, then leave the Statue in peace, but transform (in the opposite way) your Cartesian coordinate system!
More general transformations, allowing deformation of objects, could also be described by the formula $\hat{\mathcal{R}} f(\mathbf{r}) = f(\hat{R}^{-1}\mathbf{r})$, but the operator $\hat{R}$ would be nonunitary.

Since *in general* for any value of angle α function $\exp\left[-|\mathbf{r}-\mathbf{U}\mathbf{r}_0|^2\right]$ is not equal to $\exp\left[-|\mathbf{r}-\mathbf{r}_0|^2\right]$, unitary operation $\hat{\mathcal{R}}$ *is not* a symmetry operation of the object.

If, however, $\alpha = 2\pi n, n = 0, \pm 1, \pm 2, ...,$ then $\hat{\mathcal{R}} f(\mathbf{r}) = f(\mathbf{r})$ and $\hat{\mathcal{R}}(2\pi n; z)$ *is*[12] a symmetry operation.

Example 8 (Rotation of a particular sum of atomic orbitals). Let us take an example of a *sum* of three spherically symmetric Gaussian orbitals:

$$f(\mathbf{r}) = \exp\left[-|\mathbf{r}-\mathbf{r}_A|^2\right] + \exp\left[-|\mathbf{r}-\mathbf{r}_B|^2\right] + \exp\left[-|\mathbf{r}-\mathbf{r}_C|^2\right],$$

where vectors $\mathbf{r}_A, \mathbf{r}_B, \mathbf{r}_C$ are of the same length and form the Mercedes sign (angles equal to 120°, Fig. C.1c). Let us take operator $\hat{\mathcal{R}}\left(\alpha = 120^0; z\right)$ corresponding to the matrix $\mathbf{U}$ of rotation by 120°. Application of $\hat{\mathcal{R}}$ to function $f(\mathbf{r})$ is equivalent to[13]

$$f(\hat{R}^{-1}\mathbf{r}) = \exp\left[-\left|\hat{R}^{-1}\mathbf{r}-\mathbf{r}_A\right|^2\right] + \exp\left[-\left|\hat{R}^{-1}\mathbf{r}-\mathbf{r}_B\right|^2\right] + \exp\left[-\left|\hat{R}^{-1}\mathbf{r}-\mathbf{r}_C\right|^2\right] =$$
$$\exp\left[-\left|\mathbf{r}-\hat{R}\mathbf{r}_A\right|^2\right] + \exp\left[-\left|\mathbf{r}-\hat{R}\mathbf{r}_B\right|^2\right] + \exp\left[-\left|\mathbf{r}-\hat{R}\mathbf{r}_C\right|^2\right].$$

From the figure (or from the matrix) we have $\hat{R}\mathbf{r}_A = \mathbf{r}_B$; $\hat{R}\mathbf{r}_B = \mathbf{r}_C$; $\hat{R}\mathbf{r}_C = \mathbf{r}_A$. This gives

$$\hat{\mathcal{R}} f(\mathbf{r}) = \exp\left[-|\mathbf{r}-\mathbf{r}_B|^2\right] + \exp\left[-|\mathbf{r}-\mathbf{r}_C|^2\right] + \exp\left[-|\mathbf{r}-\mathbf{r}_A|^2\right] = f(\mathbf{r}).$$

We have obtained our old function; $\hat{\mathcal{R}}(\alpha = 120°; z)$ *is* therefore the *symmetry* operation[14] $f(\mathbf{r})$, i.e.,

$\hat{\mathcal{R}}(\alpha = 120°; z)$ represents a symmetry operation not only for the function f, but also for other objects, that would have the *symmetry of the equilateral triangle*

---

[12] The transformed and nontransformed orbitals coincide.

[13] We use the result from the last example.

[14] Note that, e.g., if one of the 1*s* orbitals had the opposite sign, then the function $f(\mathbf{r})$ would not have the symmetry of the equilateral triangle, although it would be invariant too with respect to *some* of the operations of the equilateral triangle.

Example 9 (Rotation of a many-electron wave function). If in the last example we took a three-electronic function as a *product* of the Gaussian orbitals $f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \exp\left[-|\mathbf{r}_1 - \mathbf{r}_A|^2\right] \cdot \exp\left[-|\mathbf{r}_2 - \mathbf{r}_B|^2\right] \cdot \exp\left[-|\mathbf{r}_3 - \mathbf{r}_C|^2\right]$, then after applying $\hat{\mathcal{R}}\left(\alpha = 120^0; z\right)$ to f we would obtain using an almost identical procedure, $\hat{\mathcal{R}} f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = f(\hat{R}^{-1}\mathbf{r}_1, \hat{R}^{-1}\mathbf{r}_2, \hat{R}^{-1}\mathbf{r}_3) = \exp\left[-|\mathbf{r}_1 - \mathbf{r}_B|^2\right] \cdot \exp\left[-|\mathbf{r}_2 - \mathbf{r}_C|^2\right] \cdot \exp\left[-|\mathbf{r}_3 - \mathbf{r}_A|^2\right]$, which represents a completely different function than the original one $f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$! Thus, $\hat{\mathcal{R}}$ does not represent any symmetry operation for $f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$. If, however, we took a symmetrized function, e.g., $\tilde{f}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \sum_P \hat{P} f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$, where $\hat{P}$ permutes the centers A, B, and C and the summation goes over all permutations, we would obtain $\tilde{f}$ that would turn out to be symmetric with respect to $\hat{\mathcal{R}}(\alpha = 120°; z)$.

Example 10 (Translation). Translation cannot be represented as a matrix transformation (C.1). It is however an *isometric operation*, i.e., it preserves the distances among the points of the transformed object. This is sufficient for us. Let us enlarge the set of the allowed operations in the three-dimensional Euclidean space by isometry. Similarly as in the case of rotations, let us define a shift of the *function* $f(\mathbf{r})$. A shift of the *function* $f(\mathbf{r})$ by vector $\mathbf{r}_1$ is such a transformation $\hat{\mathcal{R}}(\mathbf{r}_1)$ (in the Hilbert space) that the new function $\tilde{f}(\mathbf{r}) = f(\mathbf{r} - \mathbf{r}_1)$. As an example let us take the function $f(\mathbf{r}) = \exp\left[-|\mathbf{r} - \mathbf{r}_0|^2\right]$ and let us shift it by the vector $\mathbf{r}_1$. Translations obey the known relation (C.2): $\hat{\mathcal{R}}(\mathbf{r}_1) f(\mathbf{r}) = f(\hat{R}^{-1}\mathbf{r}) = \exp\left[-\left|\hat{R}^{-1}\mathbf{r} - \mathbf{r}_0\right|^2\right] = \exp\left[-|\mathbf{r} - \mathbf{r}_1 - \mathbf{r}_0|^2\right] = \exp\left[-|\mathbf{r} - (\mathbf{r}_1 + \mathbf{r}_0)|^2\right] = f(\mathbf{r} - \mathbf{r}_1)$. The function $f(\mathbf{r})$ was concentrated around the point $\mathbf{r}_0$, while the new function $\hat{\mathcal{R}}(\mathbf{r}_1) f(\mathbf{r})$ is concentrated around the point indicated by the vector $\mathbf{r}_1 + \mathbf{r}_0$, i.e., the function has been shifted by $\mathbf{r}_1$ (Fig. C.1d). This transformation is (similarly as in case of rotations) a unitary one, because the scalar product between two functions f_1 and f_2 shifted by the same operation is preserved: $\langle f_1(\mathbf{r}) | f_2(\mathbf{r})\rangle = \langle f_1(\mathbf{r} - \mathbf{r}_1) | f_2(\mathbf{r} - \mathbf{r}_1)\rangle$.

Symmetry group of the ammonia molecule

Imagine a model of the NH_3 molecule (trigonal pyramid) (Fig. C.2). A student sitting at the table plays with the model. We look at the model, then close our eyes for a second, and open them again. We see the student smiles, but the coordinate system, the model, and its position with respect to the coordinate system look exactly the same as before. Could the student change the model position? Yes, he could. He could for example rotate the model about the z axis (perpendicular to the table) by 120°, he might exchange two N–H bonds in the model, he may also do nothing. Whatever the student could do is called a *symmetry operation*.

Let us make a list of all the symmetry operations allowed for the ammonia molecule. To this end let us label the vortices of the triangle by a, b, c and locate it in such a way as to coincide the triangle center with the origin of the coordinate system, and the y axis indicates the vortex a.

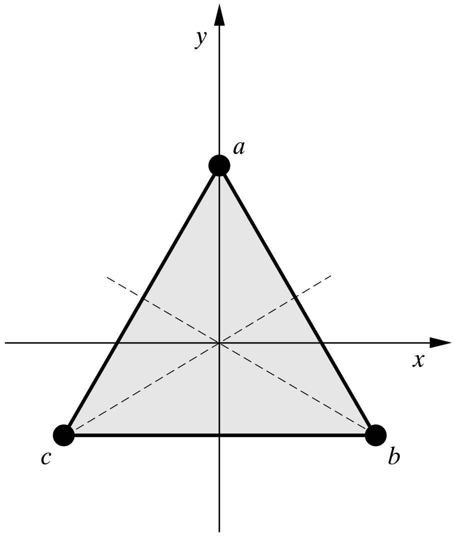


Fig. C.2. The equilateral triangle and the coordinate system. Positions *a*, *b*, *c* are occupied by hydrogen atoms, the nitrogen atom is (symmetrically) above the plane.

Table C.1. Symmetry operations of the ammonia molecule (the reflections pertain to the mirror planes perpendicular to the triangle (Fig. C.2) and going through the center of the triangle).

Symbol	Description	Symbolic explanation
$\hat{E}$	do nothing	$\hat{\mathrm{E}}\begin{bmatrix} & a & \\ c & & b \end{bmatrix} = \begin{bmatrix} & a & \\ c & & b \end{bmatrix}$
$\hat{A}$	reflection in the plane going through point *a* shown in the figure	$\hat{\mathrm{A}}\begin{bmatrix} & a & \\ c & & b \end{bmatrix} = \begin{bmatrix} & a & \\ b & & c \end{bmatrix}$
$\hat{B}$	reflection in the plane going through point *b* shown in the figure	$\hat{\mathrm{B}}\begin{bmatrix} & a & \\ c & & b \end{bmatrix} = \begin{bmatrix} & c & \\ a & & b \end{bmatrix}$
$\hat{C}$	reflection in the plane going through point *c* shown in the figure	$\hat{\mathrm{C}}\begin{bmatrix} & a & \\ c & & b \end{bmatrix} = \begin{bmatrix} & b & \\ c & & a \end{bmatrix}$
$\hat{D}$	rotation by 120° clockwise	$\hat{\mathrm{D}}\begin{bmatrix} & a & \\ c & & b \end{bmatrix} = \begin{bmatrix} & c & \\ b & & a \end{bmatrix}$
$\hat{F}$	rotation by 120° counterclockwise	$\hat{\mathrm{F}}\begin{bmatrix} & a & \\ c & & b \end{bmatrix} = \begin{bmatrix} & b & \\ a & & c \end{bmatrix}$

Now, let us check whether the operations given in Table C.1 form a group. Four conditions have to be satisfied. The first condition requires the existence of a "multiplication" in the group, and that the product of any two elements gives an element of the group: $\hat{R}_i\hat{R}_j = \hat{R}_k$. The elements will be the symmetry operations of the equilateral triangle. The product $\hat{R}_i\hat{R}_j = \hat{R}_k$ means that the operation $\hat{R}_k$ gives the same result as applying to the triangle the operation $\hat{R}_j$ *first*, and *then* the result is subject to the operation $\hat{R}_i$. In this way the "multiplication table," Table C.2, can be obtained.

Table C.2. Group multiplication table.

		Second in the product					
	$\hat{R}_j$	$\hat{E}$	$\hat{A}$	$\hat{B}$	$\hat{C}$	$\hat{D}$	$\hat{F}$
First in the product	$\hat{R}_i$						
	$\hat{E}$	$\hat{E}$	$\hat{A}$	$\hat{B}$	$\hat{C}$	$\hat{D}$	$\hat{F}$
	$\hat{A}$	$\hat{A}$	$\hat{E}$	$\hat{D}$	$\hat{F}$	$\hat{B}$	$\hat{C}$
	$\hat{B}$	$\hat{B}$	$\hat{F}$	$\hat{E}$	$\hat{D}$	$\hat{C}$	$\hat{A}$
	$\hat{C}$	$\hat{C}$	$\hat{D}$	$\hat{F}$	$\hat{E}$	$\hat{A}$	$\hat{B}$
	$\hat{D}$	$\hat{D}$	$\hat{C}$	$\hat{A}$	$\hat{B}$	$\hat{F}$	$\hat{E}$
	$\hat{F}$	$\hat{F}$	$\hat{B}$	$\hat{C}$	$\hat{A}$	$\hat{E}$	$\hat{D}$

Further, using the table we may check whether the operation is associative. For example, we check whether $\hat{A}(\hat{B}\hat{C}) = (\hat{A}\hat{B})\hat{C}$. The left-hand side gives $\hat{A}(\hat{B}\hat{C}) = \hat{A}\hat{D} = \hat{B}$. The right-hand side is $(\hat{A}\hat{B})\hat{C} = \hat{D}\hat{C} = \hat{B}$. It agrees, and it will agree for all the other entries in the table.

The unit operation is $\hat{E}$, as seen from the table, because multiplying by $\hat{E}$ does not change anything: $\hat{E}\hat{R}_i = \hat{R}_i\hat{E} = \hat{R}_i$. Also, using the table again, we can find the inverse element to any of the elements. Indeed, $\hat{E}^{-1} = \hat{E}$, because $\hat{E}$ times just $\hat{E}$ equals $\hat{E}$. Further, $\hat{A}^{-1} = \hat{A}$, because $\hat{A}$ times $\hat{A}$ equals $\hat{E}$, $\hat{B}^{-1} = \hat{B}$, $\hat{C}^{-1} = \hat{C}$, $\hat{D}^{-1} = \hat{F}$, $\hat{F}^{-1} = \hat{D}$, etc.

Thus, all the requirements are fulfilled and all these operations form a group of order $g = 6$. Note that in this group the operations do not necessarily commute, e.g., $\hat{C}\hat{D} = \hat{A}$, but $\hat{D}\hat{C} = \hat{B}$ (the group is not Abelian).

Classes

The group elements can be all divided into disjoint sets called classes. A class (to put it first in a simplified way) represents a set of the operations that are similar in a common sense, e.g., three reflection operations $\hat{A}$, $\hat{B}$, and $\hat{C}$ constitute one class, the rotations $\hat{D}$ and $\hat{F}$ form the second class, the third class is simply the element $\hat{E}$. Now, the precise definition.

CLASS

A class is the set of elements that are conjugate one to another. An element $\hat{R}_i$ is conjugate with $\hat{R}_j$ if we can find in the group G such an element (let us denote it by $\hat{X}$) that $\hat{X}^{-1}\hat{R}_j\hat{X} = \hat{R}_i$.

Then, of course, the element $\hat{R}_j$ is a conjugate to $\hat{R}_i$ as well. We check that by multiplying $\hat{R}_i$ from the left by $\hat{X} = \hat{Y}^{-1}$ and from the right by $\hat{X}^{-1} = \hat{Y}$ (which yields $\hat{Y}^{-1}\hat{R}_i\hat{Y} = \hat{X}\hat{R}_i\hat{X}^{-1} = \hat{X}\hat{X}^{-1}\hat{R}_j\hat{X}\hat{X}^{-1} = \hat{E}\hat{R}_j\hat{E} = \hat{R}_j$).

Let us make a little exercise using our table. We have $\hat{X}^{-1}\hat{E}\hat{X} = \hat{X}^{-1}\hat{X}\hat{E} = \hat{E}\hat{E} = \hat{E}$ for each $\hat{X} \in G$, i.e., $\hat{E}$ represents alone a class. Further, making $\hat{X}^{-1}\hat{A}\hat{X}$ for all possible $\hat{X}$ gives

$$\hat{E}^{-1}\hat{A}\hat{E} = \hat{E}\hat{A}\hat{E} = \hat{A}\hat{E} = \hat{A},\ \hat{A}^{-1}\hat{A}\hat{A} = \hat{A}\hat{A}\hat{A} = \hat{E}\hat{A} = \hat{A},\ \hat{B}^{-1}\hat{A}\hat{B} = \hat{B}\hat{A}\hat{B} = \hat{F}\hat{B} = \hat{C},$$
$$\hat{C}^{-1}\hat{A}\hat{C} = \hat{C}\hat{A}\hat{C} = \hat{D}\hat{C} = \hat{B},\ \hat{D}^{-1}\hat{A}\hat{D} = \hat{F}\hat{A}\hat{D} = \hat{B}\hat{D} = \hat{C},\ \hat{F}^{-1}\hat{A}\hat{F} = \hat{D}\hat{A}\hat{F} = \hat{B}\hat{F} = \hat{C}.$$

This means that $\hat{A}$ belongs to the same class as $\hat{B}$ and $\hat{C}$. Now, we will create some conjugate elements to $\hat{D}$ and $\hat{F}$:

$$\hat{A}^{-1}\hat{D}\hat{A} = \hat{A}\hat{D}\hat{A} = \hat{B}\hat{A} = \hat{F},\ \hat{B}^{-1}\hat{D}\hat{B} = \hat{B}\hat{D}\hat{B} = \hat{C}\hat{B} = \hat{F},\ \hat{C}^{-1}\hat{D}\hat{C} = \hat{C}\hat{D}\hat{C} = \hat{A}\hat{C} = \hat{F},$$

etc. Thus, $\hat{D}$ and $\hat{F}$ make a class. Therefore, the group under consideration consists of the following classes: $\{\hat{E}\}\{\hat{A}, \hat{B}, \hat{C}\}\{\hat{D}, \hat{F}\}$.

It is always like that: the *group is a sum of the disjoint classes*.

Representations

A representation of the group is such a g-element sequence of the square matrices (of the same dimension; each element of the group is associated to a matrix) that the matrices have the multiplication table consistent with the multiplication table of the group.

By consistency we mean the following. To each element of the group one assigns a square matrix (of the same dimension for all elements). If the multiplication table for the group says that $\hat{R}_i\hat{R}_j = \hat{R}_k$, then the matrix corresponding to $\hat{R}_i$ times the matrix that corresponds to $\hat{R}_j$ is the matrix that corresponds to $\hat{R}_k$. If this agrees for all $\hat{R}$, then we say that the matrices form a representation.[15]

We may create many group representations (see Table C.3).

The easiest thing is to see that Γ_1 satisfies the criterion of being a representation (the matrices have dimension 1, i.e., they are numbers). After a while of looking at Γ_2 we will say the same.

---

[15] More formally: a representation is a *homomorphism of the group into the above set of matrices.*

Table C.3. Several representations of the equilateral triangle symmetry group.

Repr.	Group elements					
	$\hat{E}$	$\hat{A}$	$\hat{B}$	$\hat{C}$	$\hat{D}$	$\hat{F}$
Γ_1	1	1	1	1	1	1
Γ_2	1	-1	-1	-1	1	1
Γ_3	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$
Γ_4	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}$

Multiplying the corresponding matrices we will prove that for Γ_3 and Γ_4. For example, for Γ_3 the product of the matrices $\hat{B}$ and $\hat{C}$ gives the matrix corresponding to the operation $\hat{D}$,

$$\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix},$$

i.e., the same as for the operations themselves. If we had more patience, we would show that equally easily for the whole multiplication table of the group. Note that

> there are many representations of a group.

Note also another interesting thing. Let us take a point with the coordinates $(x, y, 0)$ and see what will happen to it when the symmetry operations are applied (the coordinate system rests, while the point itself moves). The identity operation $\hat{E}$ leads to the following transformation matrix:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The results of the other operations are characterized by the following transformation matrices (you may check that step by step):

$$\hat{A}: \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \hat{B}: \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}, \quad \hat{C}: \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix},$$

$$\hat{D}: \begin{bmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}, \quad \hat{F}: \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}.$$

Note that the matrices obtained are identical to those of the representation Γ_3. Thus, by transforming the coordinates of a point, we have generated a representation of the symmetry.

> By transforming "anything" (coordinates of a point, vectors, functions) using the symmetry operations and collecting the results in the form of matrices, we obtain always a representation of the group.

Characters of representation

For any representation Γ we may define the vector $\boldsymbol{\chi}^{(\Gamma)}$ of dimension g, having as elements the traces of the representation matrices $\Gamma(\hat{R}_i)$,

$$\mathrm{Tr}\Gamma = \sum_i \Gamma_{ii}, \tag{C.3}$$

$$\boldsymbol{\chi}^{(\Gamma)} \equiv \begin{bmatrix} \mathrm{Tr}\Gamma\left(\hat{R}_1\right) \\ \mathrm{Tr}\Gamma\left(\hat{R}_2\right) \\ \dots \\ \mathrm{Tr}\Gamma\left(\hat{R}_g\right) \end{bmatrix} \equiv \begin{bmatrix} \chi^{(\Gamma)}\left(\hat{R}_1\right) \\ \chi^{(\Gamma)}\left(\hat{R}_2\right) \\ \dots \\ \chi^{(\Gamma)}\left(\hat{R}_g\right) \end{bmatrix}. \tag{C.4}$$

The number $\chi^{(\Gamma)}\left(\hat{R}_i\right)$ is called the *character of the representation* Γ that *corresponds to the operation* $\hat{R}_i$. The characters of representations will play the most important role in the application of group theory to spectroscopy.

Irreducible representations

To tell what an irreducible representation is, let us define first the so-called *reducible representations*.

A representation is called *reducible* if its matrices can be transformed into the so-called *block form* by using the transformation $\mathbf{P}^{-1}\Gamma(\hat{R}_i)\mathbf{P}$ for every matrix $\Gamma(\hat{R}_i)$, where $\mathbf{P}$ is a nonsingular matrix.

In a block form the nonzero elements can be only in the square blocks located on the diagonal (Fig. C.3).

If using the same $\mathbf{P}$ we can transform each of the matrices $\Gamma(\hat{R}_i)$ and obtain the same block form, then the representation is called *reducible.*

If we do not find such a matrix (because it does not exist), then the representation is irreducible. If we carry out the transformation $\mathbf{P}^{-1}\Gamma(\hat{R}_i)\mathbf{P}$ (*similarity transformation*) for $i = 1, 2, ..., g$ of a representation, the new matrices also form a representation Γ' called *equivalent* to Γ.

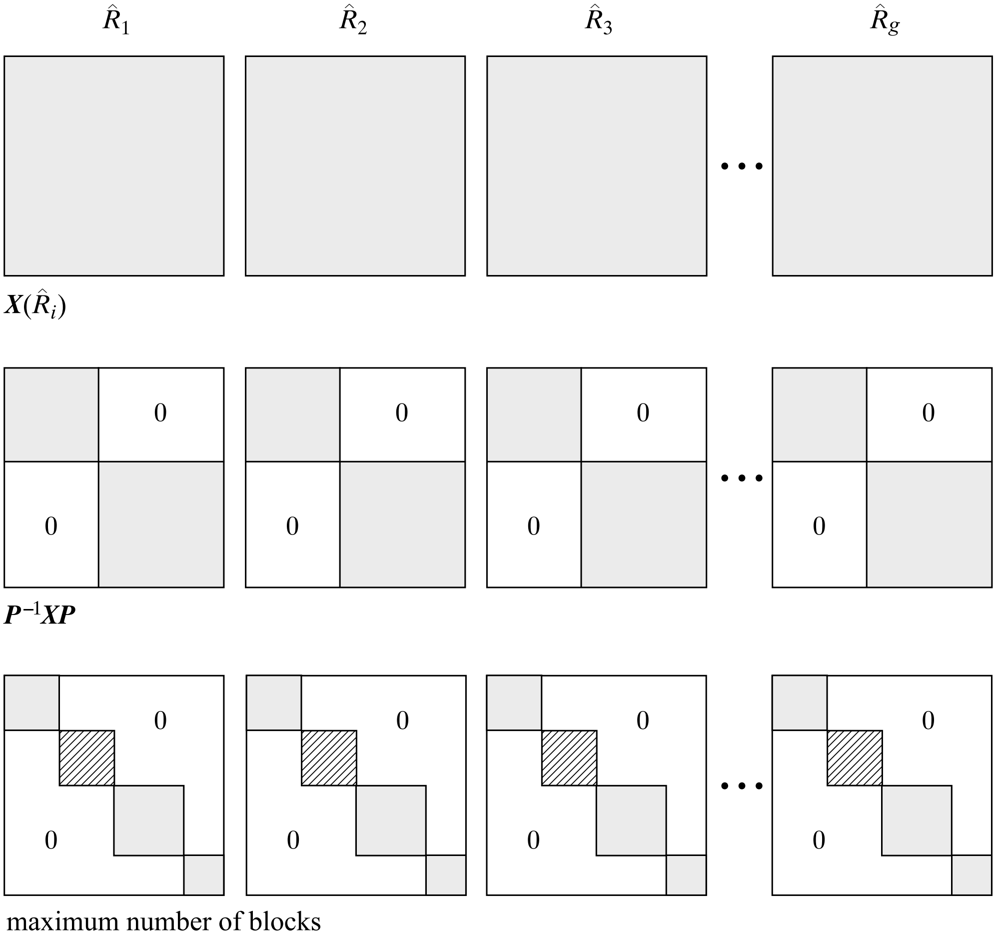


Fig. C.3. Reducible representation, block form, and irreducible representation. In the first row the matrices $\Gamma(\hat{R}_i)$ are displayed that form a *reducible representation* (each matrix corresponds to the symmetry operation $\hat{R}_i$); the matrix elements are in general nonzero. The central row shows a representation Γ' equivalent to the first one, i.e., related by a similarity transformation (with matrix $\mathbf{P}$); the new representation exhibits a block form, i.e., in this particular case each matrix has two blocks of zeros, identical in all matrices. The last row shows an equivalent representation Γ'' that corresponds to the smallest square blocks (of nonzeros), i.e., the maximum number of the blocks, of the form identical in all the matrices. Not only Γ, Γ', and Γ'' are representations of the group, but also any sequence of individual blocks (as that shadowed) is a representation. Thus, Γ'' is decomposed into the four irreducible representations.

This is easy to show. Indeed, the group operations $\hat{R}_i$ and $\hat{R}_j$ correspond to the matrices $\Gamma(\hat{R}_i)$ and $\Gamma(\hat{R}_j)$ in the original representation and to $\Gamma'(\hat{R}_i) = \mathbf{P}^{-1}\Gamma(\hat{R}_i)\mathbf{P}$ and $\Gamma'(\hat{R}_j) = \mathbf{P}^{-1}\Gamma(\hat{R}_j)\mathbf{P}$ in the equivalent representation (we will check in a moment whether indeed this is a representation). The product $\Gamma'(\hat{R}_i)\Gamma'(\hat{R}_j)$ is equal to $\mathbf{P}^{-1}\Gamma(\hat{R}_i)\mathbf{P}\mathbf{P}^{-1}\Gamma(\hat{R}_j)\mathbf{P} = \mathbf{P}^{-1}\Gamma(\hat{R}_i)\Gamma(\hat{R}_j)\mathbf{P}$, i.e., to the matrix $\Gamma(\hat{R}_i)\Gamma(\hat{R}_j)$ transformed by similarity transformation, therefore everything goes with the same multiplication table. Thus, the matrices $\Gamma'(\hat{R}_i)$ form also a representation (Γ'). This means that we can create as many representations as we wish; it is sufficient to change the matrix **P**, and this is easy (since what we want is its nonsingularity, i.e., the $\mathbf{P}^{-1}$ matrix has to exist).

The blocks are square matrices. It turns out the set of the first blocks $\Gamma_1(\hat{R}_1), \Gamma_1(\hat{R}_2), \ldots \Gamma_1(\hat{R}_g)$ (each block for one operation) is a representation, the set of the second blocks $\Gamma_2(\hat{R}_1)$, $\Gamma_2(\hat{R}_2), \ldots \Gamma_2(\hat{R}_g)$ forms a representation as well, etc. This is evident; it is sufficient to see what happens when we multiply two matrices in the same block form. The matrix product has the same block form and a particular block results from multiplication of the corresponding blocks of the matrices that are being multiplied. This is sufficient to treat each set of the blocks as a representation.[16]

In particular, the maximum decomposition into blocks leads, of course, to the blocks that are not decomposable anymore, and represent therefore the irreducible representations.

---

[16] Let us explain this taking an example. We have two square matrices of dimension 4: **A** and **B**, both having the block form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{bmatrix} \quad \text{with}$$
$$\mathbf{A}_1 = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Let us check that $\mathbf{C} = \mathbf{AB}$ has *the same* block form

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix}$$

and that (what is particularly important for us) $\mathbf{C}_1 = \mathbf{A}_1\mathbf{B}_1$ and $\mathbf{C}_2 = \mathbf{A}_2\mathbf{B}_2$. Indeed, multiplying **AB** we have

$$\mathbf{C} = \begin{bmatrix} 6 & 11 & 0 & 0 \\ 7 & 7 & 0 & 0 \\ 0 & 0 & 6 & 6 \\ 0 & 0 & 7 & 8 \end{bmatrix}, \quad \text{i.e.,} \quad \begin{bmatrix} 6 & 11 \\ 7 & 7 \end{bmatrix} = \mathbf{C}_1, \quad \begin{bmatrix} 6 & 6 \\ 7 & 8 \end{bmatrix} = \mathbf{C}_2.$$

Hence, indeed $\mathbf{C}_1 = \mathbf{A}_1\mathbf{B}_1$ and $\mathbf{C}_2 = \mathbf{A}_2\mathbf{B}_2$.

Properties of the irreducible representations

For two irreducible representations α and β the following *group orthogonality theorem* is satisfied[17]:

$$\sum_i [\Gamma^{(\alpha)}(\hat{R}_i)]_{mn} [\Gamma^{(\beta)}(\hat{R}_i)]^*_{m'n'} = \frac{g}{n_\alpha} \delta_{\alpha\beta} \delta_{mm'} \delta_{nn'}, \tag{C.5}$$

where $\Gamma^{(\alpha)}(\hat{R})$ and $\Gamma^{(\beta)}(\hat{R})$ denote the matrices that correspond to the group element $\hat{R}$ (m, n and m', n' determine the elements of the matrices), the summation goes over all the group elements, and n_α is the dimension of the irreducible representation α, i.e., the dimension of the matrices that form the representation. The symbol $*$ means the complex conjugation.[18] The symbols δ denote the Kronecker deltas, i.e., $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$ and $\delta_{\alpha\beta} = 0$ if $\alpha \neq \beta$. The "orthogonality" in the name of the theorem is related to the following observation. We create two g-dimensional vectors: one composed of the components $[\Gamma^{(\alpha)}(\hat{R}_i)]_{mn}$, the other vector from $[\Gamma^{(\beta)}(\hat{R}_i)]^*_{m'n'}$, $i = 1, 2, ..., g$. The group orthogonality theorem says that:

- if $\alpha \neq \beta$, then the vectors are orthogonal,
- if $m \neq m'$ *or* $n \neq n'$, then again the two vectors are orthogonal. The formula kills everything, except the two irreducible representations are identical and we choose as the vector components *the same* elements.

Characters of irreducible representations

The most important consequence of the group orthogonality theorem is the equation

$$\boxed{\sum_i \chi^{(\alpha)}(\hat{R}_i) \chi^{(\beta)}(\hat{R}_i)^* = g \delta_{\alpha\beta},} \tag{C.6}$$

where $\chi^{(\alpha)}(\hat{R}_i)$ is a character of the *irreducible representation* α corresponding to the symmetry operation $\hat{R}_i$. Eq. (C.6) in view of Eq. (C.3) may be rewritten as a scalar product in a unitary space (Appendix B), i.e.,

$$\left\langle \boldsymbol{\chi}^{(\beta)} | \boldsymbol{\chi}^{(\alpha)} \right\rangle = g \delta_{\alpha\beta}. \tag{C.7}$$

Eq. (C.7) can be obtained from the group orthogonality theorem after setting $m = n$ and $m' = n'$, and then summing up over m and m', i.e.,

---

[17] For the proof see H. Eyring, J. Walter, G.E. Kimball, "*Quantum Chemistry*," New York, Wiley, 1944.

[18] It is important only for the complex representations Γ.

$$\left\langle \chi^{(\beta)} | \chi^{(\alpha)} \right\rangle = \sum_i \sum_m \sum_{m'} [\Gamma^{(\alpha)}(\hat{R}_i)]_{mm} [\Gamma^{(\beta)}(\hat{R}_i)]^*_{m'm'} =$$

$$\frac{g}{n_\alpha} \delta_{\alpha\beta} \sum_m \sum_{m'} (\delta_{mm'})^2 = \frac{g}{n_\alpha} \delta_{\alpha\beta} n_\alpha = g \delta_{\alpha\beta}.$$

Decomposing reducible representation into irreducible ones

It is important that

equivalent representations have identical characters,

because the trace of a matrix is invariant with respect to any similarity transformation. Indeed, for two equivalent representations Γ and Γ' for any $\hat{R}_i$ we have $\Gamma'(\hat{R}_i) = \mathbf{P}^{-1}\Gamma(\hat{R}_i)\mathbf{P}$, which gives

$$\chi^{(\Gamma')}(\hat{R}_i) = \sum_m (\mathbf{P}^{-1}\Gamma(\hat{R}_i)\mathbf{P})_{mm} = \sum_{mkl} P^{-1}_{mk} \Gamma_{kl} P_{lm} = \sum_{kl} \Gamma_{kl} \sum_m P_{lm} P^{-1}_{mk} =$$

$$= \sum_{kl} \Gamma_{kl} (\mathbf{P}\mathbf{P}^{-1})_{lk} = \sum_{kl} \Gamma_{kl} \delta_{lk} = \sum_k \Gamma kk = \chi^{(\Gamma)}(\hat{R}_i).$$

In particular the character of a representation is the same as its block form (with the maximum number of blocks that correspond to the irreducible representations):

$$\chi(\hat{R}_i) = \sum_\alpha \mathrm{a}(\alpha) \chi^{(\alpha)}(\hat{R}_i), \tag{C.8}$$

or, in other words,

$$\chi = \sum_\alpha \mathrm{a}(\alpha) \chi^{(\alpha)}, \tag{C.9}$$

where a(α) *is a natural number telling us how many times the irreducible representation* α *appears in the block form*. The above formula comes from the very definition of the trace (a sum of the diagonal elements).

We will need another property of the characters. Namely,

the characters corresponding to the elements of a class are equal.

Indeed, two elements of the group $\hat{R}_i$ and $\hat{R}_j$ that belong to the same class are related to one another by the relation $\hat{R}_i = X^{-1}\hat{R}_j X$, where X is an element of the group. The same multiplication table is valid for the representations (from the definition of the representation), thus

$$\Gamma(\hat{R}_i) = \Gamma(X^{-1})\Gamma(\hat{R}_j)\Gamma(X) = [\Gamma(X)]^{-1}\Gamma(\hat{R}_j)\Gamma(X). \tag{C.10}$$

This concludes the proof, because in such a case the matrices $\Gamma(\hat{R}_i)$ and $\Gamma(\hat{R}_j)$ are related by a similarity transformation, and therefore have identical characters. From now on we can write $\chi(C)$ instead of $\chi(\hat{R})$, where C denotes a class to which operation $\hat{R}_i$ belongs.

Eq. (C.7) can be now modified appropriately. It can be rewritten as

$$\left\langle \boldsymbol{\chi}^{(\beta)} | \boldsymbol{\chi}^{(\alpha)} \right\rangle = \sum_C n_C \chi^{\alpha}(C)\chi^{\beta}(C)^* = \sum_C [\sqrt{n_C}\chi^{(\alpha)}(C)][\sqrt{n_C}\chi^{(\beta)}(C)^*] = g\delta_{\alpha\beta}, \tag{C.11}$$

where C stands for the class and n_C tells us how many operations belong to the class. Such a notation reminds us that the numbers $[\sqrt{n_C}\chi^{(\alpha)}(C)]$ for a fixed α and changing class C may be treated as the components of a vector (its dimension is equal to the number of classes) and that the vectors that correspond to different irreducible representations are orthogonal. The dimension of the vectors is equal to the number of classes, say, k. Since the number of the orthogonal vectors, each of dimension k, cannot exceed k, the number of the different irreducible representations is equal to the number of classes.

In future applications it will be of key importance to find such a natural number $\mathrm{a}(\alpha)$ that tells us how many times the irreducible representation α is encountered in a reducible representation. The formula for $\mathrm{a}(\alpha)$ is the following:

$$\mathrm{a}(\alpha) = \frac{1}{g}\sum_C n_C \chi(C)\chi^{(\alpha)}(C)^*. \tag{C.12}$$

The proof is simple. From the scalar product of both sides of Eq. (C.9) with the vector $\boldsymbol{\chi}^{(\beta)}$ after using Eq. (C.7) one obtains $\langle \boldsymbol{\chi}^{(\beta)} | \boldsymbol{\chi} \rangle = \sum_\alpha \mathrm{a}(\alpha) \langle \boldsymbol{\chi}^{(\beta)} | \boldsymbol{\chi}^{(\alpha)} \rangle = \sum_\alpha \mathrm{a}(\alpha) g\delta_{\alpha\beta} = \mathrm{a}(\beta) g$ or $\mathrm{a}(\alpha) = \frac{1}{g}\langle \boldsymbol{\chi}^{(\alpha)} | \boldsymbol{\chi} \rangle$. This is the formula sought, because the characters are the same for all operations of the same class.

Note that

to find $\mathrm{a}(\alpha)$ it is sufficient to know the *characters* of the representations; the representations themselves are not necessary.

Tables of characters of the irreducible representations

Any textbook on application of group theory in molecular spectroscopy contains tables of characters of irreducible representations that correspond to various symmetry groups of molecules.[19]

In order to apply group theory to a particular molecule we have to find first the abovementioned table of characters. To this end:

- the Born–Oppenheimer approximation is used; therefore, the positions of the nuclei are fixed in space ("geometry");
- looking at the geometry we make a list of all the symmetry operations that transform it into itself;
- we identify the corresponding symmetry group.[20]

In order to find the proper table we may use the Schoenflies notation for the symmetry[21] (there are also some other notations):

$\hat{E}$ means a symbol of the identity operation (i.e., does nothing),
$\hat{C}_\mathrm{n}$ rotation by angle $\frac{2\pi}{\mathrm{n}}$ about the n-fold symmetry axis,
$\hat{C}_\mathrm{n}^\mathrm{m}$ rotation by $\frac{2\pi \mathrm{m}}{\mathrm{n}}$ about the n-fold symmetry axis,
$\hat{\sigma}_v$ reflection in the plane going through the axis of the highest symmetry,
$\hat{\sigma}_h$ reflection in the plane perpendicular to the axis of the highest symmetry,
$\hat{\imath}$ inversion with respect to the center of symmetry,
$\hat{S}_\mathrm{n}$ rotation by angle $\frac{2\pi}{\mathrm{n}}$ about the n-fold symmetry axis with subsequent reflection in the plane perpendicular to it,
$\hat{S}_n^m$ rotation by angle $\frac{2\pi m}{\mathrm{n}}$ about the n-fold symmetry axis with subsequent reflection in the plane perpendicular to it.

The set of the symmetry operations forms the symmetry group. The symmetry groups also have their special symbols. In Table C.4 the Schoenflies notation of the symmetry groups of some simple molecules is given (in their geometry corresponding to the energy minimum).

A molecule may be much more complicated, but often its symmetry is identical to that of a simple molecule (e.g., one of those reported in the table).

---

[19] The tables have been constructed by considering possible symmetries (symmetry groups), creating suitable matrix representations, using similarity transformations to find the irreducible representations, and summing up the diagonal elements; we end up with the character tables in question.

[20] This may be done by using a flow chart, e.g., as given in P.W. Atkins, "*Physical Chemistry*," sixth edition, Oxford University Press, Oxford, 1998.

[21] Artur Moritz Schoenflies (1853–1928), German mathematician, professor at the universities in Göttingen, Königsberg, and Frankfurt am Main. Schoenflies proved (independently of J.S. Fiodorow and W. Barlow) the existence of the complete set of 230 space groups of crystals.

Table C.4. Examples of the symmetry group (for a few molecules in their ground-state optimum geometry).

Molecule	**Group**
H_2O	C_{2v}
NH_3	C_{3v}
CH_4	T_d
benzene	D_{6h}
naphthalene	D_{2h}

When we finally identify the table of characters suitable for the molecule under consideration, it is time to look at it carefully. For example, for the ammonia molecule we find the table of characters Table C.5.

Table C.5. C_{3v} group. Table of characters.

C_{3v}	$\hat{E}$	$3\hat{\sigma}_v$	$2\hat{C}_3$		
A_1	1	1	1	z	x^2+y^2, z^2
A_2	1	-1	1	R_z	
E	2	0	-1	$(x,y)(R_x,R_y)$	$(x^2-y^2,xy)(xz,yz)$

In the upper left corner the name of the group is displayed (C_{3v}). In the same row the symmetry operations are listed (in this case $\hat{E}$, $\hat{\sigma}_v$, $\hat{C}_3$).[22] The operations are collected in classes (columns), and the number of such operations in the class is given: the identity operation ($\hat{E}$) forms the first class, the *three* reflection operations (hence $3\hat{\sigma}_v$, called before $\hat{A}$, $\hat{B}$, $\hat{C}$) corresponding to the planes that contain the three-fold symmetry axis, *two* rotation operations (hence, $2\hat{C}_3$, called before $\hat{D}$ and $\hat{F}$) about the three-fold symmetry axis (by 120° and by 240°, or $-120°$; rotation by 360° is identical to $\hat{E}$).

In the second and further rows we have information about the irreducible representations, one row for each representation. The number of the irreducible representations is equal to the number of classes (three in our case), i.e., the table of characters is square. On the left-hand side we have the symbol of the representation informing us about its dimension (if the symbol is A, then the dimension is 1; if it is E, then the dimension is 2; if T, then 3). Thus, the letter E unfortunately plays in the table[23] a double role: as the identity operation $\hat{E}$ and as E – the symbol of an irreducible representation. In a given row (irreducible representation) the number below the

---

[22] The same symmetry operations as discussed on p. 613.

[23] This unfortunate traditional notation will not lead to trouble.

symbol of class is the corresponding character. For the identity operation $\hat{E}$ the corresponding matrices are unit matrices, the calculated character is therefore equal to the dimension of the irreducible representation.

Of great importance is the simplest representation possible, just all the characters equal to 1 (in our case A_1). It will be called the fully symmetric representation.

Example 11 (Decomposition of a reducible representation). Let us find how the reducible representation Γ_4 from p. 616 may be decomposed into the irreducible representations. First of all we see from Eq. (C.12) that what one needs are characters rather than the representations themselves. The characters $\chi^{(\Gamma_4)}$ are calculated by summing up the diagonals of the matrix representations for the corresponding classes, $\chi^{(\Gamma_4)}$: 3 (class $\hat{E}$), -1 (class $\hat{\sigma}_v$), 0 (class $\hat{C}_3$). Let us first ask how many times (a_{A_1}) the irreducible representation A_1 is encountered in Γ_4. The characters of A_1 (Table C.5) are 1, 1, 1 for the corresponding classes. The number of the operations in the classes is respectively n_C: 1, 3, 2. From (C.12) we find $a(A_1) = \frac{1}{6}(1 \cdot 3 \cdot 1 + 3 \cdot (-1) \cdot 1 + 2 \cdot 0 \cdot 1) = 0$. Similarly, we find $a(A_2) = \frac{1}{6}(1 \cdot 3 \cdot 1 + 3 \cdot (-1) \cdot (-1) + 2 \cdot 0 \cdot 1) = 1$ and $a(E) = \frac{1}{6}(1 \cdot 3 \cdot 2 + 3 \cdot (-1) \cdot 0 + 2 \cdot 0 \cdot (-1)) = 1$. Thus, we may write $\Gamma_4 = A_2 + E$. This exercise will be of great help when the selection rules in spectroscopy will be considered.

Projection operator on an irreducible representation

We will soon need the information whether a particular function exhibits certain symmetry properties of the system under consideration. To this end we will need certain projection operators, such as the following:

$$\hat{P}^{(\alpha)} = \frac{n_\alpha}{g} \sum_i \chi^{(\alpha)*}\left(\hat{R}_i\right) \hat{R}_i \tag{C.13}$$

represents the projection operator that projects on the space of such functions that transform according to the irreducible representation $\Gamma^{(\alpha)}$.

This means that either $\hat{P}^{(\alpha)} f$ *transforms according to the irreducible representation* $\Gamma^{(\alpha)}$ or we obtain zero. In order to be a projection operator, $\hat{P}^{(\alpha)}$ has to satisfy[24]

$$\hat{P}^{(\alpha)} \hat{P}^{(\beta)} = \delta_{\alpha\beta} \hat{P}^{(\alpha)}. \tag{C.14}$$

---

[24] This means that two functions that transform according to *different* irreducible representations are orthogonal, and that a projection of an already projected function changes nothing. Here is the proof. After noting that

We can also prove that

$$\sum_{\alpha} \hat{P}^{(\alpha)} = 1, \tag{C.15}$$

where the summation goes over all irreducible representations of the group.

$\hat{R}\hat{S} = \hat{Q}$ or $\hat{S} = \hat{R}^{-1}\hat{Q}$, we have

$$\begin{aligned}
\hat{P}^{(\alpha)}\hat{P}^{(\beta)} &= \frac{n_\alpha n_\beta}{g^2} \sum_{\hat{R},S} \chi^{(\alpha)*}\left(\hat{R}\right) \chi^{(\beta)*}\left(\hat{S}\right) \hat{R}\hat{S} \\
&= \frac{n_\alpha n_\beta}{g^2} \sum_{Q} \hat{Q} \sum_{\hat{R},} \chi^{(\alpha)*}\left(\hat{R}\right) \chi^{(\beta)*}\left(\hat{R}^{-1}\hat{Q}\right).
\end{aligned}$$

Note that

$$\chi^{(\beta)*}\left(\hat{R}^{-1}\hat{Q}\right) = \sum_k \Gamma_{kk}^{(\beta)*}\left(\hat{R}^{-1}\hat{Q}\right) = \sum_k \sum_l \Gamma_{kl}^{(\beta)*}\left(\hat{R}^{-1}\right) \Gamma_{lk}^{(\beta)*}\left(\hat{Q}\right).$$

After inserting this result we have

$$\begin{aligned}
\hat{P}^{(\alpha)}\hat{P}^{(\beta)} &= \frac{n_\alpha n_\beta}{g^2} \sum_{Q} \hat{Q} \sum_{\hat{R}} \sum_{m} \Gamma_{mm}^{(\alpha)*}\left(\hat{R}\right) \sum_k \sum_l \Gamma_{kl}^{(\beta)*}\left(\hat{R}^{-1}\right) \Gamma_{lk}^{(\beta)*}\left(\hat{Q}\right) \\
&= \frac{n_\alpha n_\beta}{g^2} \sum_{Q} \hat{Q} \sum_{\hat{R}} \sum_{k,l,m} \Gamma_{mm}^{(\alpha)*}\left(\hat{R}\right) \Gamma_{lk}^{(\beta)}\left(\hat{R}\right) \Gamma_{lk}^{(\beta)*}\left(\hat{Q}\right) \\
&= \frac{n_\alpha n_\beta}{g^2} \sum_{Q} \hat{Q} \sum_{k,l,m} \Gamma_{lk}^{(\beta)*}\left(\hat{Q}\right) \sum_{\hat{R}} [\Gamma_{mm}^{(\alpha)*}\left(\hat{R}\right) \Gamma_{lk}^{(\beta)}\left(\hat{R}\right)],
\end{aligned}$$

because from the unitary character of the representation matrices $\Gamma^{(\beta)}\left(\hat{R}^{-1}\right)$ and $\Gamma^{(\beta)}\left(\hat{R}\right)$ we have $\Gamma_{kl}^{(\beta)*}\left(\hat{R}^{-1}\right) = \Gamma_{lk}^{(\beta)}\left(\hat{R}\right)$. From the group theorem of orthogonality (Eq. (C.5)) we have

$$\begin{aligned}
\hat{P}^{(\alpha)}\hat{P}^{(\beta)} &= \frac{n_\alpha n_\beta}{g^2} \frac{g}{n_\alpha} \sum_{Q} \hat{Q} \sum_{k,l,m} \Gamma_{lk}^{(\beta)*}\left(\hat{Q}\right) \delta_{ml}\delta_{mk}\delta_{\alpha\beta} \\
&= \delta_{\alpha\beta} \frac{n_\alpha}{g} \sum_{Q} \hat{Q} \sum_{m} \Gamma_{mm}^{(\alpha)*}\left(\hat{Q}\right) \\
&= \delta_{\alpha\beta} \frac{n_\alpha}{g} \sum_{Q} \chi^{(\alpha)*}\left(\hat{Q}\right) \hat{Q} = \delta_{\alpha\beta} \hat{P}^{(\alpha)},
\end{aligned}$$

which is what we wanted to show (Eq. (C.13)).

Transformation of a function according to irreducible representation

The right-hand part of the character table, like, e.g., Table C.5, contains the symbols $x, y, z, (x^2 - y^2, xy)$ R_x, R_y, R_z. These symbols will be needed to establish the selection rules in spectroscopy (UV-VIS, IR, Raman). They pertain to the coordinate system (z axis coincides with the axis of the highest symmetry). Let us leave alone for a while the symbols R_x, R_y, R_z.

We have some polynomials in the rows of the table. *The polynomials transform according to the irreducible representation that corresponds to the row.*[25] If a polynomial (displayed in a row of the table of characters) is subject to the projection $\hat{P}^{(\alpha)}$, then:

- if α does not correspond to the row, then we obtain 0;
- if α corresponds to the row, then we obtain either the polynomial itself (if the irreducible representation has dimension 1), or, if the dimension of the irreducible representation is greater than 1, a linear combination of the polynomials given in the same row (in parentheses).

If function f transforms according to a one-dimensional irreducible representation, the function is an eigenfunction of all the symmetry operators $\hat{\mathcal{R}}$, with the corresponding eigenvalues $\chi^{(\alpha)}(\hat{R})$.

Let us come back to R_x, R_y, R_z. Imagine R_x, R_y, R_z as oriented circles perpendicular to a rotation axis (i.e., x, y, or z) that symbolize rotations about these axes. For example, the operation $\hat{E}$ and the two rotations $\hat{C}_3$ leave the circle R_z unchanged, while the operations $\hat{\sigma}_v$ change its orientation to the opposite one, hence R_z transforms according to the irreducible representation A_2. It turns out that R_x and R_y transform under the symmetry operations into their linear combinations and therefore correspond to a two-dimensional irreducible representation (E).

Group theory and quantum mechanics

Representation basis

If in a molecule we have two equivalent[26] nuclei, then this *always* results from a molecular symmetry, i.e., at least one symmetry operation exchanges the positions of these two nuclei.

---

[25] Please recall the definition of the symmetry operation given on p. 611: $\hat{\mathcal{R}}f(\mathbf{r}) = f(\mathbf{r})$, where $\hat{\mathcal{R}}f(\mathbf{r}) = f(\hat{R}^{-1}\mathbf{r})$.

[26] With respect to physical and chemical properties.

There is no reason at all that electrons like one of such nuclei more than the other one.[27] Let us focus on molecular orbitals calculated for a fully symmetric Fock operator.[28] Therefore,

> each molecular orbital has to be such that when taking the square of it, the electron density is the same on the equivalent nuclei.

What will happen however with the *molecular orbital itself*? Taking the square of it removes information about its sign. The signs at both atoms may be the same (symmetric orbital), but they may also be opposite[29] (antisymmetric orbital). For example, the bonding orbital for the hydrogen molecule is symmetric with respect to reflection in the plane perpendicular to the internuclear axis[30] and going through its center, while the antibonding orbital is antisymmetric with respect to the operation.

---

[27] This may be not true for nonstationary states. The reason is simple. Imagine a long polymer molecule with two equivalent atoms at its ends. If one of them is touched by the tip of the tunnel microscope and one electron is transferred to the polymer, a nonstationary asymmetric electron state is created.

[28] Limiting ourselves to molecular orbitals is not essential.

[29] This pertains to nondegenerate orbital levels. For a degenerate level any linear combination of the eigenfunctions (associated to the same level) is also an eigenfunction as good as those that entered the linear combination. A symmetry operation acting on an orbital gives another orbital corresponding to the same energy. In such a case the squares of both orbitals in general do not exhibit the symmetry of the molecule. However, we can find such a linear combination of both that its square preserves the symmetry.

[30] Let us see what it really means in a very *formal* way (it may help us in more complicated cases). The coordinate system is located in the middle of the internuclear distance (on the x axis, the internuclear distance is equal to $2A$). The bonding orbital $\varphi_1 = N_1(a+b)$ and the antibonding orbital $\varphi_2 = N_2(a-b)$, where N are the normalization constants, the $1s$ atomic orbitals have the following form:

$$a \equiv \frac{1}{\sqrt{\pi}} \exp[-|\mathbf{r}-\mathbf{A}|] = \frac{1}{\sqrt{\pi}} \exp\left[-\sqrt{(x-A)^2+y^2+z^2}\right] =$$

$$b \equiv \frac{1}{\sqrt{\pi}} \exp[-|\mathbf{r}+\mathbf{A}|] = \frac{1}{\sqrt{\pi}} \exp\left[-\sqrt{(x+A)^2+y^2+z^2}\right],$$

$$\mathbf{A} = (A, 0, 0).$$

The operator $\hat{\sigma}$ of the reflection in the plane $x=0$ corresponds to the unitary transformation matrix of the coordinates $\mathbf{U} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Therefore, the inverse matrix $\mathbf{U}^{-1} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, i.e., the transformation $\mathbf{U}^{-1}\mathbf{r}$ means
$x \to -x$, $y \to y$, $z \to z$, which transforms $a \to b$ and $b \to a$. Hence,

$$\hat{\sigma}(a+b) = (b+a) = (a+b),$$

$$\hat{\sigma}(a-b) = (b-a) = -(a-b).$$

In both cases the molecular orbital represents an eigenfunction of the symmetry operator with the eigenvalue $+1$ and -1, respectively.

We know how to apply the symmetry operations on molecular orbitals (p. 611) and transform them to other functions.

> Under such a symmetry operation the orbital either remains unchanged (as the above-mentioned bonding one) or changes sign (as the antibonding one).

If the orbital level is degenerate, we may obtain another function. This function corresponds to the same energy, because applying any symmetry operation we only exchange equivalent nuclei, which otherwise are treated on equal footing in the Hamiltonian.

If we obtain another orbital (φ_2), then we may begin to play with it by applying all the symmetry operations. Some operations will lead to the same (new) orbital, sometimes with the opposite sign, after some other operations we obtain the old orbital φ_1, sometimes with the opposite sign, and sometimes these operations will lead to the *third* orbital φ_3. Then, we apply the symmetry operations to the third orbital, etc. We do it until the final set of orbitals is obtained which transforms into themselves when subjected to the symmetry operations. We may treat the set of such linearly independent orbitals φ_i, $i = 1, ..., n$, as the basis set in a vector space.

All the results of application the operation $\hat{\mathcal{R}}_i$ on the orbitals φ_i are collected in a transformation matrix $\mathbf{R}_i$, where $\boldsymbol{\varphi} = \begin{bmatrix} \varphi_1 \\ \ldots \\ \varphi_n \end{bmatrix}$:

$$\hat{\mathcal{R}}_i \boldsymbol{\varphi} = \mathbf{R}_i^T \boldsymbol{\varphi}. \tag{C.16}$$

> The matrices $\mathbf{R}_i$, $i = 1, 2, ..., g$, form the n-dimensional representation (in general reducible) of the symmetry group of the molecule.

Indeed, let us see what happens if we apply the operation $\hat{\mathcal{T}} = \hat{\mathcal{R}}_1 \hat{\mathcal{R}}_2$ to the function φ_i. We obtain

$$(\hat{\mathcal{R}}_1 \hat{\mathcal{R}}_2)\varphi_i = \hat{\mathcal{R}}_1 \mathbf{R}_2^T \boldsymbol{\varphi} = \mathbf{R}_2^T \hat{\mathcal{R}}_1 \boldsymbol{\varphi} = \mathbf{R}_2^T \mathbf{R}_1^T \boldsymbol{\varphi} = (\mathbf{R}_1 \mathbf{R}_2)^T \boldsymbol{\varphi}.$$

This means that all the matrices $\mathbf{R}_i$ form a representation.

> BASIS OF A REPRESENTATION
>
> A set of the linearly independent functions φ_i that served to create the representation forms a *basis of the representation.*

The basis need not have been composed of the orbitals; it could be expressions like x, y, z or x^2, y^2, z^2, xy, xz, yz or any linearly independent functions, provided they transform into themselves under symmetry operations. We may begin from an atomic orbital, and after applying the symmetry operations soon will obtain a basis set that contains this orbital and all the equivalent ones.

Decomposition of a function into irreducible representation components

Let us take a function f belonging to a Hilbert space. Since (see Eq. (C.15)) $\sum_\alpha \hat{P}^{(\alpha)} = 1$, where α goes over all the irreducible representations of the group, f can be written as a sum of its components $f^{(\alpha)}$, each component (belonging to the corresponding subspace of the Hilbert space) transforming according to the irreducible representation α, i.e.,

$$f = 1 \cdot f = \sum_\alpha \hat{P}^{(\alpha)} f = \sum_\alpha f^{(\alpha)}. \tag{C.17}$$

In view of (C.14) the components $f^{(\alpha)}$ and $f^{(\beta)}$ are automatically orthogonal if $\alpha \neq \beta$.

Example 12 (Decomposition of a function). Let us take three hydrogen atoms in the configuration of an equilateral triangle, and assume we are in the simplest version of the molecular orbitals in the LCAO MO approximation, i.e., the atomic basis set is composed of the three 1s orbitals a, b, c centered on the three nuclei. Let us check whether the functions

$$\begin{aligned} u_1 &= a + b + c, \\ u_2 &= b - c, \\ u_3 &= a - c, \end{aligned}$$

form a basis to a (reducible) representation. If the symmetry operations are applied to a, b, c, then they transform into each other (cf. Fig. C.2), and the results obtained are easily shown as linear combinations of the functions u_1, u_2, u_3 (with $\mathbf{R}_i^T$ as transformation matrices). For example, $\hat{A}u_1 = a + b + c = u_1$, $\hat{A}u_2 = -b + c = -u_2$, $\hat{A}u_3 = a - b = -u_2 + u_3$. Hence,

$$\mathbf{A}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \tag{C.18}$$

In this way (see (C.16)) we obtain $\mathbf{R}_i$ as

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & -1 \end{bmatrix}, \tag{C.19}$$

$$\mathbf{C}=\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix},\quad \mathbf{D}=\begin{bmatrix}1&0&0\\0&0&1\\0&-1&-1\end{bmatrix},\quad \mathbf{F}=\begin{bmatrix}1&0&0\\0&-1&-1\\0&1&0\end{bmatrix}. \tag{C.20}$$

Let us check that $\mathbf{DF}=\mathbf{E}$ and $\mathbf{AD}=\mathbf{B}$, i.e., exactly as for the operations $\hat{D}\hat{F}=\hat{E}$, $\hat{A}\hat{D}=\hat{B}$, and so on. Thus, this is a representation; moreover, this is a representation already *in a block form*, because u_1 transforms always in itself, while u_2 and u_3 mix among themselves. It can be shown that this mixing cannot be avoided by any choice of u! Hence, u_1 alone represents a basis of a one-dimensional irreducible representation (A_1 – this is seen from the characters corresponding to the first block 1×1), while u_2 and u_3 form a basis of a two-dimensional irreducible representation (E). Note that from the mathematical form of the functions u it follows that u_2 and u_3 have to correspond to the same energy and this energy is different from that corresponding to u_1. The conclusion is that a, b, c form a basis for a reducible representation, while their linear combinations u_1 and $\{u_2,\ u_3\}$ form the basis sets of two irreducible representations: A_1 and E. Any function that is a linear combination of a, b, c can be represented as a linear combination of u_1, u_2, u_3 as well.

The same symmetry orbitals can be obtained by using the projection operators (C.13). Let us take any of functions a, b, c (the result does not depend on this choice), e.g., function a. In view of the beautiful equilateral triangle such a function is no doubt a really deformed object that does not take care of making the three vortices of the triangle equivalent. Let us see whether such a function has any component that transforms according to the irreducible representation A_1. To this end let us use the projection operator $\hat{P}^{(A_1)}$ (Eq. (C.13)) and the table of characters on p. 625): $\hat{P}^{(A_1)}a=\frac{1}{6}(a+b+c+a+b+c)=\frac{1}{3}(a+b+c)$. Thus, there *is* a fully symmetric component[31] in a. Now, let us use the same orbital a to obtain $\hat{P}^{(A_2)}a=\frac{1}{6}(a+b+c-a-b-c)=0$. This means that a does not contain anything that transforms according to A_2. Now it is the turn of the irreducible representation E: $\hat{P}^{(E)}a=\frac{2}{6}(2a-b-c+0\cdot a+0\cdot b+0\cdot c)=\frac{1}{3}[2(a-c)-(b-c)]$. What has been obtained is a linear combination of u_2 and u_3.

If the projections were made for the function b, then we would obtain a trivial repeating[32] for the irreducible representations A_1 and A_2 and a nontrivial result for the irreducible representation E: $\hat{P}^{(E)}b=\frac{2}{6}(2b-a-c+0\cdot a+0\cdot b+0\cdot c)=\frac{1}{3}[2(b-c)-(a-c)]$, which is just another linear combination of u_2 and u_3. These two functions are therefore inseparable and form a basis for a two-dimensional *irreducible* representation.

---

[31] This sentence carries a simple message: by mixing symmetric objects we may obtain an asymmetric one, e.g., the asymmetric function $a+2b$ can be represented by a linear combination u_1+u_2, both functions transforming according to the irreducible representation of the symmetry group.

[32] $\hat{P}^{(A_1)}b=\frac{1}{3}(a+b+c)$ and $\hat{P}^{(A_2)}b=0$.

DECOMPOSITION INTO IRREDUCIBLE REPRESENTATIONS

Any function that is a linear combination of the basis functions of a *reducible* representation can be decomposed into a linear combination of the basis functions of those *irreducible* representations that form the reducible representation.

Most important

THE MOST IMPORTANT...

The wave functions corresponding to an energy level

— *form a basis of an irreducible representation* of the symmetry group of the molecule, or in other words, transform according to this irreducible representation;

— the dimension of the representation is equal to the degeneracy of the energy level.

This is how it should be, because if a symmetry operation acts on an eigenfunction of the Hamiltonian, we will have only two possibilities: (1) we obtain the same function to the accuracy of the sign (in case of a one-dimensional representation, by definition irreducible), or (2) we obtain another function *corresponding to the same energy* (because of the same physical situation); acting on the function obtained and repeating the whole procedure we will arrive finally to a set of n linearly independent functions that correspond to the same energy (a basis of an n-dimensional irreducible representation).

This means (Fig. C.4) that

the energy levels may be labeled by tags, where each tag corresponds to a single irreducible representation.

This will be of fundamental importance when the selection rules in spectroscopy will be considered.

We usually have plenty of energy levels, while the number of the irreducible representation is small. Thus, there will be in general a lot of levels with the same labels. This result has been first obtained by Eugene Wigner. The group theory will not tell us how many levels correspond to a particular irreducible representation, or what energy they correspond to.

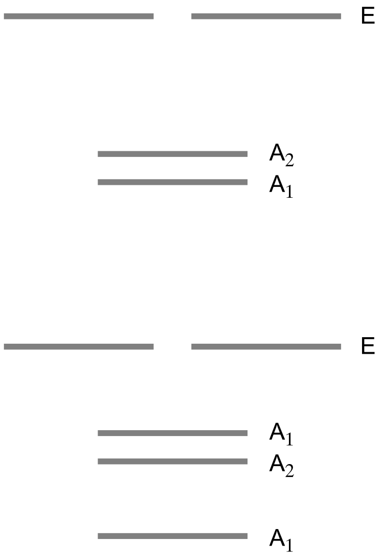


Fig. C.4. Each energy level corresponds to an irreducible representation of the symmetry group of the Hamiltonian. Its linearly independent eigenfunctions that correspond to a given level form a basis of the irreducible representation, or in other words, transform according to this representation. The number of the basis functions is equal to the degeneracy of the level.

Integrals important in spectroscopy

Direct product of irreducible representations

We are approaching application of group theory in optical transitions in spectroscopy. The most important issue will be the decision whether an integral is zero or nonzero. If the integral is zero, then the transition is forbidden; if it is nonzero, then it is allowed. To make such a decision we have to use what is known as the *direct product of irreducible representations*. Imagine basis functions $\{\varphi_i\}$ and $\{\psi_j\}$ that correspond to irreducible representations α and β of the symmetry group of a molecule. Let us make a set $\{\varphi_i\psi_j\}$ of all possible products of them (i.e., the Cartesian product).

DIRECT PRODUCT

The products $\{\varphi_i\psi_j\}$ when subject to symmetry operations lead (as usual) to a representation called the *direct product* $\Gamma^\alpha \times \Gamma^\beta$ *of the irreducible representations* Γ^α and Γ^β.

The functions $\{\varphi_i \psi_j\}$ form a basis set of a representation (reducible in general). We obtain the matrices of the representations as usual by applying symmetry operations:

$$\hat{\mathcal{R}}\left[\varphi_i(\mathbf{r})\,\psi_j(\mathbf{r})\right] = \varphi_i\left(\hat{R}^{-1}\mathbf{r}\right)\psi_j\left(\hat{R}^{-1}\mathbf{r}\right) = \sum_k U_{ik}^{(\alpha)}\varphi_k \sum_l U_{jl}^{(\beta)}\psi_l = \sum_{kl} U_{ik}^{(\alpha)} U_{jl}^{(\beta)}\varphi_k\psi_l =$$
$$= \sum_{kl} Z_{ij,kl}\varphi_k\psi_l,$$

where $U_{ik}^{(\gamma)}$ are the matrix elements of the irreducible representation γ, $Z_{ij,kl} = U_{ik}^{(\alpha)} U_{jl}^{(\beta)}$. Of course,

> the dimension of this representation is the product of the dimensions of the representations α and β, because this is the number of the functions $\varphi_k\psi_l$.

The characters of the representation can be easily obtained from the characters of the irreducible ones; we just have to multiply the latter ones, i.e.,

$$\chi^{(\alpha\times\beta)}(\hat{R}) = \chi^{(\alpha)}(\hat{R})\chi^{(\beta)}(\hat{R}). \qquad \text{(C.21)}$$

Indeed, the formula is justified by

$$\chi(\hat{R}) = \sum_{kl} Z_{kl,kl} = \sum_{kl} U_{kk}^{(\alpha)} U_{ll}^{(\beta)} = (\sum_k U_{kk}^{(\alpha)})(\sum_l U_{ll}^{(\beta)}) =$$
$$= \chi^{(\alpha)}(\hat{R})\chi^{(\beta)}(\hat{R}). \qquad \text{(C.22)}$$

This rule can be naturally generalized for larger numbers of the irreducible representations in the direct product (just multiply the characters of the irreducible representations). We will have in a while a product of three irreducible representations.

When is an integral bound to be zero?

Everybody knows how to calculate the integral

$$\int_{-1}^{+1} x\mathrm{d}x = \left[\frac{x^2}{2}\right]_{-1}^{+1} = \frac{1}{2} - \frac{1}{2} = 0.$$

Note, however, that we can tell what the value of the integral is without any calculation, just looking at the integrand. Indeed, the integrand is odd with respect to the transformation $x \to -x$, i.e., the plot of the integral is an antisymmetric function with respect to the reflection in the plane perpendicular to x at $x = 0$. The integration limits are symmetric with respect to that point. An integral means the area under the plot; therefore what we gain for $x > 0$, we lose for $x < 0$ and the integral will be exactly zero.

The force of group theory relies on the fact that even having a complicated integrand we are able to tell immediately whether the integral *is* equal to zero. This will allow us to predict whether an optical transition is allowed or forbidden.

We have to stress that these conclusions will be valid independently of the approximations used to compute the molecular wave functions. The reason is that they follow from the symmetry, which is identical for the exact and approximate wave functions.

The previous example can be generalized. Let us take the integral

$$\int f_\alpha f_\beta f_\gamma \ldots d\tau, \tag{C.23}$$

where $f_\alpha, f_\beta, f_\gamma, \ldots$ *transform according to the irreducible representations* $\Gamma^{(\alpha)}, \Gamma^{(\beta)}, \Gamma^{(\gamma)}$, ..., respectively, of a symmetry group, and the integration is over the whole space.

WHEN IS THE INTEGRAL EQUAL TO ZERO?

If a representation (in general reducible) being the direct product of the irreducible representations $\Gamma^{(\alpha)}, \Gamma^{(\beta)}, \Gamma^{(\gamma)} \ldots$ does not contain the fully symmetric representation (all its characters are equal to 1), then the integral is equal to zero.

This is precisely what we have aimed for in this appendix and why we have been working so hard with symmetry groups, operations, characters, etc. The essence of the theorem is very simple. The product $f_\alpha f_\beta f_\gamma \ldots$ transforms according to the (in general *reducible*) representation, which is the direct product of the irreducible representations $\Gamma^{(\alpha)}, \Gamma^{(\beta)}, \Gamma^{(\gamma)} \ldots$. This means that according to Eq. (C.17) the integrand $f_\alpha f_\beta f_\gamma \ldots$ can be represented as a linear combination of the basis functions of all the irreducible representations: $f_\alpha f_\beta f_\gamma \ldots = \sum_\mu g_\mu$, where g_μ transforms according to the irreducible representation $\Gamma^{(\mu)}$. Therefore, the integral (C.23) is a sum of the integrals

$$\int f_\alpha f_\beta f_\gamma \ldots d\tau = \sum_\mu \int g_\mu \, d\tau, \tag{C.24}$$

each with the integrand transforming to an *irreducible* representation $\Gamma^{(\mu)}$. Let us take one of such integrals: $\int g_\mu \, d\tau$. Note that the integration is over the whole space (i.e., the integration limits are symmetric). If the integrand g_μ were antisymmetric with respect to one or more symmetry operations, the integral would be automatically zero (the same argument as for $\int x dx$). *From this follows that all integrals in the sum would be zero except that single one that contains the integrand transforming according to the fully symmetric representation.*[33]

There are two (important for us) special cases of this theorem.

Two special cases

We have $\int f_\alpha f_\beta d\tau = \delta_{\alpha\beta} A$, i.e., in order to have the integral not vanish we have to have $\Gamma^{(\alpha)} = \Gamma^{(\beta)}$.

The proof is very simple and relies on the fact that the characters of the fully symmetric irreducible representation are equal to 1. The number of times, a (A), the fully symmetric representation A is present in the direct product $\Gamma^{(\alpha)} \times \Gamma^{(\beta)}$, we calculate from the formula:

$$\mathrm{a}\,(\mathrm{A}) = \frac{1}{g}\sum_i \chi^{(\alpha\times\beta)}(\hat{R}_i)\chi^{(\mathrm{A})}(\hat{R}_i)^* = \frac{1}{g}\sum_i \chi^{(\alpha\times\beta)}(\hat{R}_i) = \frac{1}{g}\sum_i \chi^{(\alpha)}(\hat{R}_i)\chi^{(\beta)}(\hat{R}_i)^* = \delta_{\alpha\beta}. \tag{C.25}$$

This means that the fully symmetric representation is always present in $\Gamma^{(\alpha)} \times \Gamma^{(\alpha)}$ and therefore the integral does not vanish.[34]

Let us take the integral

$$\int \mathrm{f}_\alpha \mathrm{f}_\beta \mathrm{f}_\gamma \mathrm{d}\tau, \tag{C.26}$$

where $f_\alpha, f_\beta, f_\gamma$ transform according to the irreducible representations α, β, γ. The integral, for the direct product $\Gamma^{(\alpha)} \times \Gamma^{(\beta)}$ not to vanish, has to contain the representation $\Gamma^{(\gamma)}$.

---

[33] Only for the fully symmetric representation all the characters are equal to 1, and therefore the corresponding function does not change under symmetry operations.

[34] It is easy to understand. What transforms according to $\Gamma^{(\alpha)} \times \Gamma^{(\alpha)}$ is a product of two (in general different) functions, each belonging to $\Gamma^{(\alpha)}$. This means that the function behaves in a very special way (typical for $\Gamma^{(\alpha)}$) under the symmetry operations, e.g., changes sign under $\hat{R}_1$, while other operations leave it unchanged. If we have a product of two such functions, then this means the product does not change at all under $\hat{R}_1$ (and, of course, the other operations), i.e., transforms according to the fully symmetric operation. This is why the fully symmetric representation is always present in $\Gamma^{(\alpha)} \times \Gamma^{(\alpha)}$.

This means that in order for integral (C.26) not to vanish, the function $f_\alpha f_\beta$ decomposes Eq. (C.17) in such a way, that there is a nonzero component belonging to $\Gamma^{(\gamma)}$. If this happens, according to the previous case, a component of the integrand will transform according to the fully symmetric representation, which will save the integral (C.26) from vanishing.

Selection rules for electronic transitions (UV-VIS)

The selection rules will be shown taking an example of pyrazine and its mono- and diprotonated ions (Fig. C.5).

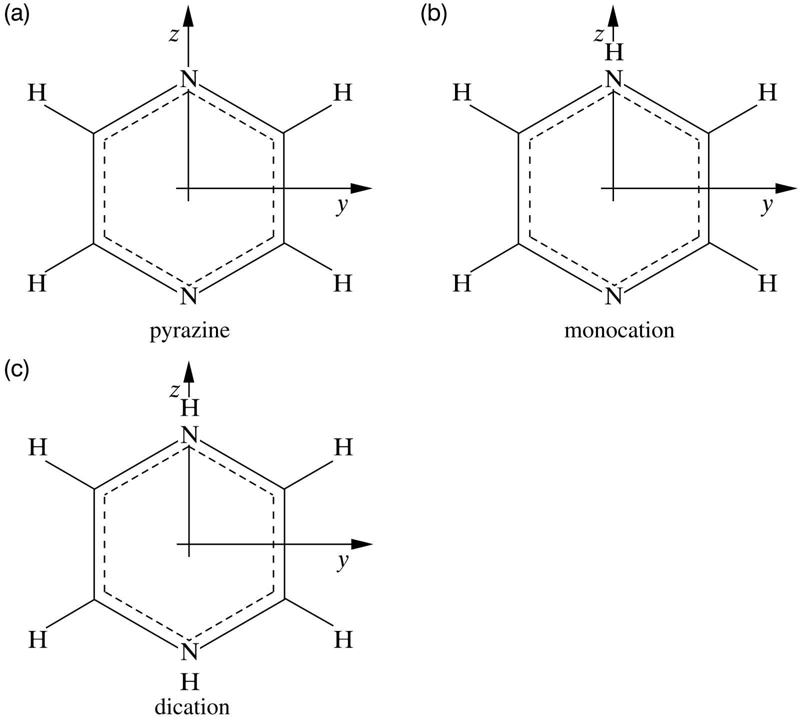


Fig. C.5. (a) Pyrazine and its (b) monoprotonated and (c) diprotonated derivatives. The x axis is perpendicular to the ring plane, the y axis is in the ring plane perpendicular to the NN axis, and z means the NN axis.

A glimpse on the chemical formulae is sufficient to tell that the monocation of pyrazine has the same symmetry as H_2O that corresponds to the symmetry group C_{2v} (see Table C.4), while pyrazine and its diprotonated derivative have the symmetry identical with that of naphthalene, i.e., D_{2h}. Let us focus first on the last case.

Example 13 (Pyrazine and its diprotonated derivative). Every book on group theory contains the table of characters of the symmetry group D_{2h} (Table C.6, the x axis perpendicular to the plane of the molecule, z goes through the nitrogen atoms).

Table C.6. D_{2h} group. Table of characters.

D_{2h}	$\hat{E}$	$\hat{C}_2(z)$	$\hat{C}_2(y)$	$\hat{C}_2(x)$	$\hat{\imath}$	$\hat{\sigma}(xy)$	$\hat{\sigma}(xz)$	$\hat{\sigma}(yz)$	
A_g	1	1	1	1	1	1	1	1	x^2, y^2, z^2
B_{1g}	1	1	-1	-1	1	1	-1	-1	R_z xy
B_{2g}	1	-1	1	-1	1	-1	1	-1	R_y xz
B_{3g}	1	-1	-1	1	1	-1	-1	1	R_x yz
A_u	1	1	1	1	-1	-1	-1	-1	
B_{1u}	1	1	-1	-1	-1	-1	1	1	z
B_{2u}	1	-1	1	-1	-1	1	-1	1	y
B_{3u}	1	-1	-1	1	-1	1	1	-1	x

From Table C.6 we see[35] that what we call irreducible representations is *distinct rhythms* of pluses and minuses, which after taking the square give the fully symmetric behavior. All the electronic states of pyrazine and its diprotonated derivative can be labeled by the irreducible representation labels: $\mathrm{A}_g, \mathrm{B}_{1g}, \mathrm{B}_{2g}, \mathrm{B}_{3g}, \mathrm{A}_u, \mathrm{B}_{1u}, \mathrm{B}_{2u}, \mathrm{B}_{3u}$.

We may ask what the selection rules are for the state-to-state optical transitions. Are all transitions allowed, or are some of them forbidden? From the theory of the electromagnetic field (cf. Chapters 7 and V2-4) it follows that the probability of the transition between the states k and l is proportional to $|\mu_{kl}(x)|^2$ or to $|\mu_{kl}(y)|^2$ or to $|\mu_{kl}(z)|^2$, respectively,[36] depending on the electromagnetic wave polarization along x or y or z axes, with

$$\begin{aligned} \mu_{kl}(x) &= \int \psi_k^* \hat{\mu}_x \psi_l \mathrm{d}\tau, \\ \mu_{kl}(y) &= \int \psi_k^* \hat{\mu}_y \psi_l \mathrm{d}\tau, \\ \mu_{kl}(z) &= \int \psi_k^* \hat{\mu}_z \psi_l \mathrm{d}\tau, \end{aligned} \tag{C.27}$$

where ψ stands for the electronic states k and l, $\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_z$ are the operators of the molecular dipole moment components,[37] e.g., $\hat{\mu}_z = \sum_i q_i z_i$, q_i is the electric charge of the particle

---

[35] Note that all the irreducible representations of the symmetry group of the molecules under consideration are one-dimensional, hence their energy levels are nondegenerate.

[36] From the equality $|\mu_{kl}(x)|^2 = |\mu_{lk}(x)|^2$ and similarly for y and z, it follows that the optical excitation and the corresponding deexcitation have the same probability.

[37] This may look alarming, because the operator depends on the choice of the coordinate system (cf. Appendix V2-G). Do not worry, everything is all right. Even if the dipole moment depends on such choice, any

(electron or nucleus) having its z component equal to z_i. Since we will decide, by using group theory, whether this integral[38] vanishes or not, what will count is that μ_x transforms exactly in the same way as the x coordinate does. The integrand $\psi_k^* \hat{\mu}_x \psi_l$ transforms as the direct product of the three irreducible representations: that of ψ_k, that of $\hat{\mu}_x$, and that of ψ_l.

Excitations from the ground-state

Suppose we:

- have a molecule in its ground state ψ_k (thus, belonging to the fully symmetric irreducible representation A_g),
- immobilize the molecule in space (say, in a crystal),
- introduce the coordinate system in the way described above,
- irradiate the molecule with light polarized along the x axis, and

ask to *which states the molecule can be excited*. The direct product of A_g and the irreducible representation to which x belongs decompose into some irreducible representations. If an optical transition is to be allowed to a state ψ_l we have to find among them an irreducible representation to which ψ_l belongs (just recall that $\int f_\alpha f_\beta d\tau = \delta_{\alpha\beta} A$). Only then will the integrand contain something that has a chance to transform according to the fully symmetric representation. The x coordinate belongs to the representation B_{3u} (see Table C.6, last column). Therefore, let us see what represents the direct product $A_g \times B_{3u}$. We have Eq. (C.12) for the integer $a(\alpha)$ that is a number of the irreducible representations α in a given reducible representation. Let us compute this number for the (in general reducible) representation being the direct product and all the irreducible representations α. In this particular case the direct product is[39] $A_g \times B_{3u}$. We have

$$\begin{aligned} a(A_g) = \frac{1}{8}[& 1 \times 1 + 1 \times (-1) + 1 \times (-1) + 1 \times 1 + 1 \times (-1) + 1 \times 1 + \\ & + 1 \times 1 + 1 \times (-1)] = 0, \end{aligned}$$

---

two choices give the dipole moments differing by a *constant vector*. This vector being a constant can be shifted outside the integral and the integral itself will become zero, because ψ_k and ψ_l are orthogonal. Thus, to our delight light absorption does not depend on the choice of the coordinate system! That is fine.

[38] The integration goes over all the electronic coordinates.

[39] The characters of $A_g \times B_{3u}$ are as follows (in the order of the symmetry operations in the table of characters): $1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \quad 1 \quad -1$, i.e., they are identical to those of the (it turned out...irreducible) representation B_{3u}. Such a product is a 100% banality to make. In the table of characters one finger just goes horizontally over the characters of A_g (all they are equal to 1), while the second finger moves similarly over the characters of B_{3u}, and we multiply what the first finger shows by that what the second one indicates. The result is the character of the direct product $A_g \times B_{3u}$, which in this case turns out to be exactly the character of B_{3u}. This is why we may expect that $a(\alpha)$ will all be zero except $a(B_{3u}) = 1$.

$$\begin{aligned}a(B_{1g}) = \frac{1}{8}[&1\times 1 + 1\times(-1) + (-1)\times(-1) + (-1)\times 1 +\\ &1\times(-1) + 1\times 1 + +(-1)\times 1\\ &+(-1)\times(-1)] = 0,\end{aligned}$$

etc., all zeros, and finally,

$$\begin{aligned}a(B_{3u}) = \frac{1}{8}(&1\times 1 + (-1)\times(-1) + (-1)\times(-1) + 1\times 1 + (-1)\times(-1) + 1\times 1 +\\ &+1\times 1 + (-1)\times(-1)] = 1,\end{aligned}$$

exactly as we expected. Thus, we can write[40]

$$A_g \times B_{3u} = B_{3u}.$$

Now only those ψ_l are allowed in optical transitions (from the ground state A_g) that are labeled by B_{3u}, because only the direct product $B_{3u} \times B_{3u}$ may contain the fully symmetric irreducible representation A_g. Thus, the transitions $A_g \Rightarrow B_{3u}$ and $B_{3u} \Rightarrow A_g$ are allowed if the light is polarized along x, i.e., perpendicularly to the ring of the molecule.

Now, let us take the light polarized along y, i.e., within the molecular plane, perpendicularly to the N–N line. This time we are interested in the irreducible representations that arise from $A_g \times B_{2u}$, because y transforms according to B_{2u}. Very similarly as before (by analyzing $a(\alpha)$) we find

$$A_g \times B_{2u} = B_{2u}.$$

This means that now the allowed states are of the B_{2u} type.

Similarly, for the polarization along z (z belongs to B_{1u}), i.e., along the nitrogen–nitrogen direction, we have

$$A_g \times B_{1u} = B_{1u}.$$

Thus, for polarization parallel to the NN axis of the molecule, the absorption may occur from the ground state to any state of the B_{1u} type (and *vice versa*).

Nothing more can be said when basing solely on group theory. One will not get any information about the energies of the transitions or about the corresponding intensities. In order to get this additional (and important) information we have to take pains and work hard to solve

---

[40] We may say that the fully symmetric representation plays the role of unity in multiplication of irreducible representations.

the Schrödinger equation, rather than to count on some easy profits obtained by primitive multiplication of integers (as in group theory). To obtain the intensities, we have to calculate the transition moment integrals μ_{kl}. However, group theory just by excluding from the spectrum a lot of transitions (forbidden ones) provides a lot of important information on the symmetry of the molecule. Table C.7 collects the calculated light frequencies[41] ($\bar{\nu}$ in wave numbers, or cm^{-1}, $\nu = c\bar{\nu}$, where ν is the usual frequency), the so-called oscillator strengths f_{kl} (in a.u.), i.e.,

$$f_{kl} = \frac{4\pi c}{3}\nu|\mu_{kl}|^2, \tag{C.28}$$

as well as the polarization of light for excitations from the electronic ground state for pyrazine and the pyrazine monocation. It is seen that the left-hand side of Table C.7 is consistent with the selection rules derived above. Indeed, a large f_{kl} corresponds only to those transitions from the ground state of pyrazine that have been predicted to be allowed (B_{1u}, B_{2u}, and B_{3u}). Also the predicted polarization agrees with the observed ones.

Table C.7. Wave numbers ($\bar{\nu}$), oscillator strengths (f_{kl}), and light polarization (in parentheses).

Pyrazine			Pyrazine monocation		
excited state	$\bar{\nu}$	f_{kl}	***excited state***	$\bar{\nu}$	f_{kl}
B_{3u}	28 960	0.015(x)	B_1	27 440	0.007(x)
B_{2u}	36 890	0.194(y)	B_2	34 130	0.280(y)
B_{2g}	38 890	0.0	A_2	45 100	0.0
A_u	41 710	0.0	A_1	49 720	0.126(z)
B_{1u}	49 800	0.183(z)	B_1	57 380	0.012(x)
B_{1g}	57 070	0.0	A_2	57 710	0.0
B_{1u}	57 420	0.426(z)	A_1	58 210	0.625(z)
A_u	60 170	0.0	A_2	59 830	0.0
B_{2g}	60 970	0.0	B_2	60 370	0.010(y)

Excitations from an excited state

Calculations for the absorption from the ground state were particularly simple. Now, let us see whether anything will be more complicated for the transitions from an excited state of the B_{2g} type of symmetry. We are going to calculate a(α) (for every α) for the following representations:

for polarization along x: $\mathrm{B}_{2g} \times \mathrm{B}_{3u}$,
for polarization along y: $\mathrm{B}_{2g} \times \mathrm{B}_{2u}$,
for polarization along z: $\mathrm{B}_{2g} \times \mathrm{B}_{1u}$.

---

[41] J. Koput, unpublished results.

The characters of the representation $\mathrm{B}_{2g} \times \mathrm{B}_{3u}$ are (Table C.6, the first finger goes along B_{2g}, the second along B_{3u}, etc.)

$$1 \quad -1 \quad -1 \quad -1 \quad -1 \quad 1 \quad 1$$

and are identical with the characters of B_{1u}. Hence, even without any calculation of $a(\alpha)$ we have $\mathrm{B}_{2g} \times \mathrm{B}_{3u} = \mathrm{B}_{1u}$. Thus, the transitions (for the polarization along x) are allowed only to the states labeled by B_{1u}, because otherwise there is no chance to obtain a fully symmetric integrand. Similarly, by multiplying B_{2g} and B_{2u} we obtain the following characters of $\mathrm{B}_{2g} \times \mathrm{B}_{2u}$:

$$1 \quad 1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1 \quad -1,$$

and this is identical to the characters of A_u; therefore $\mathrm{B}_{2g} \times \mathrm{B}_{2u} = \mathrm{A}_u$. If the polarization of the light is along y, then the only excitations (or deexcitations) possible are to the states belonging to A_u. Finally, for the polarization along z we find the following characters of $\mathrm{B}_{2g} \times \mathrm{B}_{1u}$:

$$1 \quad -1 \quad -1 \quad 1 \quad -1 \quad 1 \quad 1 \quad -1,$$

which turn out to be those of B_{3u}. This means that $\mathrm{B}_{2g} \times \mathrm{B}_{1u} = \mathrm{B}_{3u}$ and that the transitions are possible only to the states belonging to B_{3u}.

Example 14 (Pyrazine monocation). As to the selection rules nothing was said so far about the pyrazine monocation. We will be interested in excitations from the electronic ground state (as in Table C.7). The pyrazine monocation corresponds to the symmetry group C_{2v} (Table C.8).

Table C.8. C_{2v} group characters.

C_{2v}	E	C_2	$\sigma_v(xz)$	$\sigma_v(yz)$		
A_1	1	1	1	1	z	x^2, y^2, z^2
A_2	1	1	−1	−1	R_z	xy
B_1	1	−1	1	−1	x, R_y	xz
B_2	1	−1	−1	1	y, R_x	yz

The ground state belongs to the fully symmetric irreducible representation A_1. Since (as before) we begin by excitations from the ground state, let us see which irreducible representations arise from $\mathrm{A}_1 \times \mathrm{B}_1$ (for the x polarization of light, see Table C.8, x transforms according to B_1), $\mathrm{A}_1 \times \mathrm{B}_2$ (for the y polarization), and $\mathrm{A}_1 \times \mathrm{A}_1$ (for the z polarization). We calculate the characters of $\mathrm{A}_1 \times \mathrm{B}_1$ by multiplying 1 by

$$1 \quad -1 \quad 1 \quad -1,$$

and checking in Table C.8 that these correspond to B_1 (it has to be like that, because the characters of A_1 are all equal to 1), i.e., $A_1 \times B_1 = B_1$. Similarly, even without immediate checking we see that $A_1 \times B_2 = B_2$ and $A_1 \times A_1 = A_1$. In this way the following allowed transitions from the ground state (A_1) have been predicted:

$$\begin{aligned} &\text{for polarization along } x: && A_1 \to B_1, \\ &\text{for polarization along } y: && A_1 \to B_2, \\ &\text{for polarization along } z: && A_1 \to A_1. \end{aligned}$$

Now we are able to compare the spectrum for pyrazine and for its monocation (Table C.7). Attaching a proton to the pyrazine (creating its monocation) does not look as something that would ruin the UV-VIS spectrum. We might expect that the frequencies of the bands and even their intensities should be somehow similar in both molecules. As we can see from the table the frequencies are similar indeed, although in the middle of the table the deviations are quite significant. For both molecules there are forbidden ($f_{kl} = 0$) and allowed ($f_{kl} \neq 0$) transitions. Note that what is allowed for pyrazine is also allowed for its cation; the light polarization coincides, and even the values of f_{kl} are similar (we took into account that the transition to B_{1u} in pyrazine with the frequency of 49 800 cm^{-1} corresponds to the transition to A_1 in the monocation with the frequency of 49 720 cm^{-1}). In the monocation there are some additional transitions allowed: to B_1 and to B_{2g}. This is quite understandable, because the number of symmetry operations for the monocation is smaller, and forbids results from molecular symmetry. If a molecule had no symmetry operations at all (except of course the identity symmetry), all transitions would be allowed.

Thus, practically with zero effort we find the selection rules in UV-VIS for any molecule we want.

Selection rules in IR and Raman spectra

The selection rules derived above pertained to electronic transitions, where the positions of the nuclei were fixed in space. Now a change of the vibrational states of the molecule will be considered, while the electronic state is assumed to be unchanged. The vibrations of a molecule are related to its vibrational levels (each of them corresponding to an irreducible representation) and the corresponding vibrational wave functions, and the IR spectrum results from transitions between such levels. Fig. C.6 shows the energy levels of three normal modes.

In the harmonic approximation the problem of small amplitude vibrations (Chapters 6 and 7) reduces to the $(3N - 6)$ normal modes (N is the number of atoms in the molecule). Each

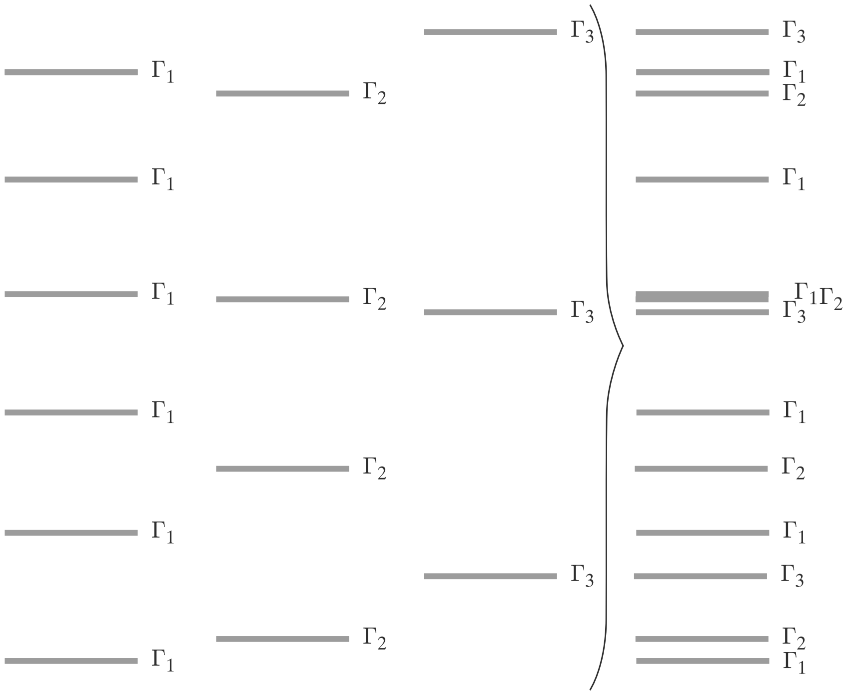


Fig. C.6. Small amplitude harmonic vibrations of a molecule (N atoms) are described by $(3N-6)$ independent harmonic oscillators (normal modes). Each normal mode is characterized by an irreducible representation. A scheme of the vibrational energy levels of three normal modes corresponds to the irreducible representations $\Gamma_1, \Gamma_2, \Gamma_3$. The modes have different frequencies, hence the interlevel separations are different for all of them (but equal for a given mode due to the harmonic potential). On the right-hand side all these levels are shown together.

of the normal modes may be treated as an independent harmonic oscillator. A normal mode moves all the atoms with a certain frequency about their equilibrium positions in a concerted motion (the same phase). The relative deviations (i.e., the ratios of the amplitudes) of the vibrating atoms from the equilibrium are characteristic for the mode, while the deviation itself is obtained from them by multiplication by the corresponding normal mode coordinate $Q \in (-\infty, \infty)$. The value $Q = 0$ corresponds to the equilibrium positions of all the atoms, Q and $-Q$ correspond to two opposite deviations of any atom from its equilibrium position.

Each normal mode belongs to an irreducible representation of the symmetry group of the molecule. What does it really mean? In any mode the displacements of the *equivalent* atoms from the equilibrium have the same absolute value, although they may differ by sign.

> We assume that small atomic deviations satisfy the symmetry requirements of the symmetry group of the molecule (valid for all atoms in the equilibrium positions) and transform according to the irreducible representation to which the normal mode belongs. Taking the square of the deviations destroys information about their signs, i.e., the absolute values of the deviations of the equivalent atoms are the same. This means that the squares of deviations transform according to the fully symmetric representation of the group.

To establish the vibrational selection rules let us define first the vibrational states of $(3N-6)$ harmonic oscillators (normal modes). The ground state of the system is no doubt the state in which *every* normal mode i is in its ground state, $\psi_{i,0}$. The ground-state wave function of the i-th normal mode reads as (p. 217)

$$\psi_{i,0} = N_0 \exp(-a_i Q_i^2), \tag{C.29}$$

where $a_i > 0$ is a constant and Q_i is the normal mode coordinate. Whatever this normal mode is, the wave function contains the *square* of Q_i, i.e., the sign of the deviations of the equivalent atoms *is irrelevant*.

> The squares of the deviations and therefore function $\psi_{i,0}$ itself transform independently of i.

Let us denote this fully symmetric irreducible representation by A_1. The wave function of the first *excited state* of a normal mode has the form

$$\psi_{i,1} = N_1 Q_i \exp(-a_i Q_i^2) \tag{C.30}$$

and we see that $\psi_{i,1}$ transforms exactly as the coordinate Q_i does, i.e., according to that irreducible representation to which the normal mode belongs (because Q_i^2 in the exponent and therefore the exponent itself both belong to the fully symmetric representation). In the harmonic approximation the total vibrational wave function of the system of $(3N-6)$ normal (i.e., independent) oscillators can be written as

$$\psi_0^{osc} = \psi_{1,0}\psi_{2,0}\psi_{3,0}\ldots\psi_{3N-6,0}, \tag{C.31}$$

where the zeros in the indices mean that all the modes are in their ground states. This means that ψ_0^{osc} transforms according to the representation being the direct product $A_1 \times A_1 \times A_1 \times \ldots A_1 = A_1$ (a banality; all the characters of A_1 are equal to 1). Now, let us focus on the excited states of the $(3N-6)$ vibrational modes. The excited states may be quite complex, but the most important (and the simplest) ones are those with all the normal modes in their ground states except a single mode that is in its first excited state. A transition from the many-oscillator ground state to such an excited state is called a *fundamental transition*. The intensities of the fundamental transitions are at least by one order of magnitude larger than others. This is why we will focus on the selection rules for such transitions. Let us take one of such singly excited states (with the first mode excited):

$$\psi_1^{osc} = \psi_{1,1}\psi_{2,0}\psi_{3,0}\ldots\psi_{3N-6,0}. \tag{C.32}$$

The function $\psi_{1,1}$ corresponding to the first excited state transforms according to the irreducible representation Γ to which the normal mode 1 belongs. Thus, ψ_1^{osc} transforms according to $\Gamma \times A_1 \times A_1 \times A_1 \times \ldots A_1 = \Gamma$, i.e., it belongs to the same irreducible representation as $\psi_{1,1}$ does. Of course, if the only excited mode were the i-th one, then the many-oscillator wave function would belong to the same irreducible representation as the wave function of the particular oscillator does. We will need this result in a while.

IR selection rules

Let us consider a molecule having a fixed position in a Cartesian coordinate system. To excite the molecule IR light (because the separation of the vibrational levels corresponds to the IR region) is used that is polarized along the x axis. The electromagnetic theory says that what decides about the intensity of the absorption is the transition integral[42]

$$\int \psi_0^{osc}\hat{\mu}_x\psi_1^{osc}\mathrm{d}\tau, \tag{C.33}$$

where $\hat{\mu}_x$ stands for the dipole moment component x. The selection rules mean to establish which integrals of that kind will be zero for symmetry reasons. To this end, what we need is information about the irreducible representations to which ψ_0^{osc}, $\hat{\mu}_x$, ψ_1^{osc} belong.[43] Since ψ_0^{osc} transforms according to A_1, the integral to survive the function ψ_1^{osc} has to belong to the same irreducible representation as $\hat{\mu}_x$ (and therefore x itself). It was shown a while before that ψ_1^{osc} belongs to the same irreducible representation to which the normal mode 1 belongs. In other words, the rule is:

SELECTION RULE IN IR

the transition from the ground state is allowed for those normal modes that transform as x, where x is the direction of the light polarization. The same will hold for the light polarization along y and z.

---

[42] The integration goes over the coordinates of the nuclei.

[43] We are going to analyze the direct product of these three representations. If it contains the fully symmetric representation, then the integral is not zero.

Raman selection rules

Chandrasekhar Venkata Raman (1888–1970), Indian physicist, professor at the University of Calcutta and at the Indian Scientific Institute in Bangalore. Raman discovered in 1928 light scattering that has been accompanied by a change of frequency (by frequency of molecular vibrations). In 1930 Raman received the Nobel Prize
"for his work on the scattering of light and for the discovery of the effect named after him."

The physics of the Raman spectra is different: rather than direct absorption this is a light scattering (in the UV-VIS region) on molecules. It turns out that besides the light the source is emitting we also detect quanta of the energy lower or higher by $h\nu$, where ν is a vibrational frequency of the molecule. For the Raman scattering to be nonzero, at least one of the following integrals should be nonzero:

$$\int \psi_0^{osc} \hat{\alpha}_{qq'} \psi_1^{osc} \mathrm{d}\tau, \tag{C.34}$$

where $\hat{\alpha}_{qq'}$ with $q, q' = x, y, z$ is a component of the polarizability tensor, which transforms as one of the following (cf. Eq. (V2-4.42), p. V2-281): $qq' = x^2, y^2, z^2, xy, xz, yz$, or their linear combinations (this information is available in the tables of characters). An identical reasoning as before leads to the conclusion that

> the normal mode excited in a fundamental transition has to belong to the same irreducible representation as the product qq'.

It remains to be seen to *which irreducible representations the normal modes belong*. The procedure consists of two stages.

Stage 1. First, the global Cartesian coordinate system is chosen. In this system we draw the equilibrium configuration of the molecule, the atoms numbered. On each atom a local Cartesian coordinate system is located with the axes parallel to those of the global one, Fig. C.7. For each atom we draw the arrows of its displacements along x, y, and z oriented towards the positive values (altogether $3N$ displacements), assuming that the displacements of equivalent atoms have to be the same. When symmetry operations are applied these displacements transform into themselves[44] and therefore form a basis set of a (reducible) representation Γ of the symmetry group of the molecule (in its equilibrium position). This representation will be decomposed into the irreducible representations.

---

[44] For example, a displacement of an atom along x under a symmetry operation turns out to be a displacement of another atom.

Stage 2. The reducible representation describes the genuine (internal) vibrations as well as the six apparent vibrations (three translations and three rotations). The apparent vibrations can be easily eliminated by throwing away (from the total reducible representation) those irreducible representations that correspond to x, y, z (translations) and R_x, R_y, R_z (rotations). We know what the latter ones are from the corresponding table of characters. Summing that up: the abovementioned reducible representation has to be decomposed into the irreducible ones. The decomposition yields $\Gamma = a(\Gamma_1)\Gamma_1 + a(\Gamma_2)\Gamma_2 + a(\Gamma_3)\Gamma_3 \ldots$. From these decompositions we have to subtract (in order to eliminate the apparent vibrations) all the irreducible representations the x, y, z, R_x, R_y, and R_z belong to.

After the two stages made we are left with a number of the irreducible representations that pertain to the genuine vibrations.[45] Only after that we can establish the vibrational selection rules according to the same procedure that has been used before. All this will be shown in a simple example of the carbonate anion CO_3^{2-} that in its equilibrium configuration corresponds to the D_{3h} symmetry group (Fig. C.7).

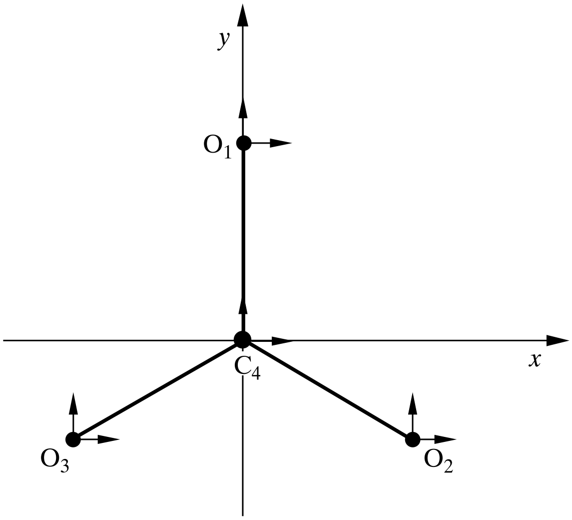


Fig. C.7. The carbonate anion CO_3^{2-}, the coordinate system used, and the versors describing the displacements of the atoms.

Example 15 (IR and Raman spectra of the carbonate anion). In order to decompose a reducible representation into the irreducible representations we do not need the reducible representation

---

[45] Rather internal motions. Note that some of these genuine vibrations may correspond to rotations of the functional groups in the molecule.

be given in details. It is sufficient to know its characters (p. 618). These characters are easy to deduce just considering what happens to the displacement vectors along x_i, y_i, z_i (for atom i) under all the symmetry operations. What will simplify our task greatly is that *only the diagonal* elements of the matrices of the reducible representation contribute to the characters. This is how it looks in practice.

Class	The character of the corresponding matrix
E	$\chi(E) = 12$ *Justification*: each versor transforms into itself. Hence, each diagonal element is equal to 1, and the number of them is equal to 3 times the number of atoms = 12.
$2C_3$	$\chi(C_3) = 0$ *Justification*: 0 from the oxygens, because they transform into other oxygens $+1(\text{from } z_4) + \cos 120°(\text{from } x_4) + \cos 120°(\text{from } y_4) = 0$.
$3C_2$	$\chi(C_2) = -2$ *Justification*: it is sufficient to consider only one of the operations of the class – other ones will have the same character. Let us take the rotation about the C_2 axis going through O_1 and C. Then, the only versors that transform into themselves (eventually changing sign – then the contribution to the character is -1) are those related to O_1 and C. We have $\chi(C_2) = -1(\text{from } z_4) + (-1)(\text{from } z_1) - 1(\text{from } x_1) - 1(\text{from } x_4) + 1(\text{from } y_1) + 1(\text{from } y_4) = -2$.
σ_h	$\chi(\sigma_h) = 4$ *Justification*: the contribution from each atom will be the same, i.e., χ will be equal to 4 times the contribution from a single atom, the latter one equals $-1(\text{from } z) + 1(\text{from } x) + 1(\text{from } y) = 1$.
$2S_3$	$\chi(S_3) = -2$ *Justification*: only C gives a contribution, which is equal to $-1(\text{from } z_4) - \frac{1}{2}(\text{from } x_4) - \frac{1}{2}(\text{from } y_4) = -2$.
$3\sigma_v$	$\chi(\sigma_v) = 2$ *Justification*: let us take only a single operation from the class – this one, which represents the reflection in the plane going through O_1 and C_4. Then the contributions to χ are the same for both these atoms, and one of them gives $-1(\text{from } x) + 1(\text{from } z) + 1(\text{from } y) = 1$.

Thus, the characters of the reducible representation have been found. In order to decompose the representation we have to know the table of characters for the D_{3h} symmetry group (Table C.9).

Table C.9. Characters of the irreducible representations of the symmetry group D_{3h}.

D_{3h}	$\hat{E}$	$2\hat{C}_3$	$3\hat{C}_2$	$\hat{\sigma}_h$	$2\hat{S}_3$	$3\hat{\sigma}_v$		
A_1'	1	1	1	1	1	1		x^2+y^2, z^2
A_2'	1	1	−1	1	1	−1	R_z	
E'	2	−1	0	2	−1	0	x, y	$x^2-y^2,\ xy$
A_1''	1	1	1	−1	−1	−1		
A_2''	1	1	−1	−1	−1	1	z	
E''	2	−1	0	−2	1	0	$R_x,\ R_y$	$xz,\ yz$

Let us write down (in the same order as in Table C.9) the characters of the reducible representation just found:

$$12 \quad 0 \quad -2 \quad 4 \quad -2 \quad 2.$$

Now, let us find (p. 623) how many times (a(α)) the irreducible representation α is present in the reducible representation (a sum over classes: number of operations in the class × the calculated character × the character of the irreducible representation):

$$\mathrm{a}(\mathrm{A}_1') = \frac{1}{12}[1\times 12\times 1 + 2\times 0\times 1 + 3\times(-2)\times 1 + 1\times 4\times 1 + 2\times(-2)\times 1 + 3\times 2\times 1] = 1.$$

Similarly, we find (knowing only how to multiply numbers such as 1, 2, 3) that

$$\mathrm{a}(\mathrm{A}_2') = 1,\ \mathrm{a}(\mathrm{E}') = 3,\ \mathrm{a}(\mathrm{A}_1'') = 0,\ \mathrm{a}(\mathrm{A}_2'') = 2,\ \mathrm{a}(\mathrm{E}'') = 1.$$

This means that the reducible representation in question decomposes into

$$\Gamma = \mathrm{A}_1' + \mathrm{A}_2' + 3\mathrm{E}' + 2\mathrm{A}_2'' + \mathrm{E}''. \tag{C.35}$$

From the table of characters we see that the apparent vibrations (see the irreducible representations corresponding to $x,\ y,\ z,\ R_x,\ R_y,\ R_z$) belong to A_2'', E', A_2', E''. After subtracting them from Γ we obtain the irreducible representations that correspond to the genuine vibrations:

$$\mathrm{A}_1', \quad \mathrm{A}_2'', \quad 2\mathrm{E}',$$

i.e., one vibration of symmetry A_1' (and a certain frequency ν_1), two vibrations (each doubly degenerate) of symmetry E' (they differ by frequency, $\nu_3 \neq \nu_4$), and one vibration of A_2'' symmetry (corresponding to frequency ν_2).

SELECTION RULES FOR IR

Therefore, we expect the following selection rules for the fundamental transitions in the IR spectrum for the CO_3^{2-} anion:

1. x and y belong to representation E′, and therefore frequencies ν_3 and ν_4 are active in IR;
2. z belongs to representation A_2'', and therefore frequency ν_2 is active in IR.

SELECTION RULES FOR RAMAN SPECTRA

For the Raman spectra we expect the following selection rules. Active will be the vibrations with the frequency:

1. ν_1, because $x^2 + y^2$ and z^2 belong to A_1';
2. ν_3 and ν_4, because $x^2 - y^2$ and xy belong to E′;

while the vibration of the frequency ν_2 will be inactive in the Raman spectroscopy, because none of the polarizability components (symbolized by x^2, y^2, etc.) belongs to A_2''.

The results are collected in Table C.10 (sign "+" = active vibration, sign "−" = inactive vibration, the polarization of the light is shown in parentheses).

Table C.10. Transitions in CO_3^{2-} active (+) in the IR and Raman spectra.

Representation	ν	IR (polarization)	Raman
A_1'	ν_1	−	+
A_2''	ν_2	+ (z)	−
E′	ν_3	+ (circular)	+
E′	ν_4	+ (circular)	+

As seen from Table C.10 in the case of the carbonate anion the vibration ν_1 is inactive in IR, but active in the Raman spectroscopy, while the opposite is true for ν_2. The vibrations with the frequencies ν_3 and ν_4 are active in both IR and Raman spectra.

EXCLUSION RULE

If the molecule under study has a center of symmetry, then the *exclusion rule* is valid, i.e., the vibrations that are active in IR are inactive in the Raman spectrum, and *vice versa*.

This follows from the fact that in that case x, y, z belong to different irreducible representations than $x^2, y^2, z^2, xy, xz, yz$. Indeed, the x, y, z are antisymmetric with respect to the inversion operation, whereas $x^2, y^2, z^2, xy, xz, yz$ or their combinations are symmetric with respect to inversion. This guarantees that they belong to *different* irreducible representations; therefore, for a molecule with the center of inversion the vibrations active in IR are inactive in Raman spectra and *vice versa*.

When do the conclusions drawn from group theory fail?

When deriving the selection rules the following assumptions have been made:

- the molecule is *isolated*;
- elements are represented by *the same isotope*;
- the molecule is in a *stationary state*;
- the vibrations have *small amplitudes*;
- the vibrations are *harmonic*;
- the electromagnetic field interacts with the molecule only through the *electric* field–molecule interaction;
- in the interaction of the molecule with the electromagnetic field only the *dipole transitions* are involved.[46]

However, in practice the molecule is never isolated. In particular, the interactions it undergoes in liquid or solid state are sufficiently strong to deform the molecule. As a result we have to do (especially in a liquid) with a population of molecules, each in a different geometry, usually devoid of *any particular symmetry* (for a single molecule this means a nonstationary state), although the molecule is not far away from the perfect symmetry (“broken symmetry”).

Suppose for a while that the molecule under consideration *is* indeed isolated. In a substance we usually have several isotopomers, with different distributions of the isotopes in the molecules. In most cases this means also a broken symmetry. *A broken symmetry means that the selection rules in principle are not applicable.*

> In practice, a broken symmetry means that the selection rules cause only a *small intensity of the forbidden transitions with respect to the allowed ones*.

---

[46] That is, the electric field of the electromagnetic wave within the molecule is assumed to be *uniform*. Then the only term in the Hamiltonian related to the light–molecule interaction is $-\hat{\boldsymbol{\mu}}\boldsymbol{\mathcal{E}}$, where $\hat{\boldsymbol{\mu}}$ stands for the dipole moment operator of the molecule and $\boldsymbol{\mathcal{E}}$ is the electric field intensity.

When considering electronic transitions we assumed that the molecule stays in its equilibrium geometry, often of high symmetry. This may be the most probable configuration,[47] but the vibrations and rotations deform it. An electronic excitation is fast and usually takes place at a molecular geometry that differs slightly from the most probable and most symmetric one. This will cause a transition forbidden for the perfectly symmetric geometry to have a nonnegligible intensity.

Deriving the selection rules for the IR and Raman spectra we assumed that equivalent atoms can differ only by the *sign* of the deviation from the equilibrium position, but its absolute value is the same. This is how it would be for a harmonic oscillator. An anharmonicity introduces therefore another reason why a (harmonically) forbidden transition will have a nonnegligible intensity.

The electromagnetic field has its electric and magnetic components. The selection rules we have derived did not take into account the presence of the magnetic field. Taking into account the magnetic field introduces some additional selection rules. Also, the wave length of an electromagnetic wave in the UV-VIS region is of the order of thousands of Å, whereas the length of the molecule is usually of the order of a few Å. This means that the assumption that the electric field of the electromagnetic wave is uniform looks good, but anyway the field is not perfectly uniform. The deviations will be small, but nonzero. Taking this into account by including further terms besides $-\hat{\boldsymbol{\mu}}\boldsymbol{\mathcal{E}}$ we obtain the interaction of the electric field gradient with the quadrupole moment of the molecule, as well as further terms. This also weakens the selection rules found.

Despite these complications, group theory allows for understanding the basic features of the molecular spectra. It works sometimes even if the molecule under study has no symmetry at all, because of a substituent that breaks it. Some electronic or vibrational excitations are of a local spatial character and pertain to a portion of the molecule that is (nearly) symmetric. Therefore some optical transitions that are absolutely allowed, because the molecule as a whole does not have any symmetry,[48] *will still have a very low intensity.*

---

[47] The maximum of the ground-state probability density for the *harmonic* oscillator corresponds indeed just to the equilibrium geometry. This is why the selection rules work at all (although in an approximate way).

[48] But they would be forbidden if the portion in question represented a separate molecule and were symmetric.

A Two-State Model

The Schrödinger equation $\hat{H}\psi = E\psi$ is usually solved[1] by expanding the unknown wave function ψ in a series of the complete basis set $\{\phi_i\}_{i=1}^{N}$ of states ϕ_i, where N in principle equals ∞ (instead in practice we end up with a chosen large value of N). The expansion gives $\hat{H}\sum_j c_j\phi_j = E\sum_j c_j\phi_j$, or $\sum_j c_j\left(\hat{H}\phi_j - E\phi_j\right) = 0$. By multiplying this equation successively by ϕ_i^*, $i = 1, 2, \ldots N$, and integrating we obtain a set of N linear equations for the unknown coefficients[2] c_i:

$$\sum_j c_j(H_{ij} - ES_{ij}) = 0,$$

where the Hamiltonian matrix elements $H_{ij} \equiv \left\langle \phi_i|\hat{H}\phi_j \right\rangle$, and the overlap integrals $S_{ij} \equiv \langle \phi_i|\phi_j \rangle$. The summation going to infinity makes any simple insight into the physics of the problem impossible. However, in many cases what matters most are only *two* states of comparable energies, while other states being far away in the energy scale practically do not count (have negligible c_j). If indeed only two states were in game (the two-state model), then the situation could be analyzed in detail. The conclusions drawn are of great conceptual (and smaller numerical) importance.

For the sake of simplicity in further analysis the functions ϕ_j will be assumed normalized and real[3]. Then for $N = 2$ we have $H_{12} = \left\langle \phi_1|\hat{H}\phi_2 \right\rangle = \left\langle \hat{H}\phi_1|\phi_2 \right\rangle = \left\langle \phi_2|\hat{H}\phi_1 \right\rangle = H_{21}$, where H_{11} and H_{22} are real numbers (in most practical applications $H_{12}, H_{11}, H_{22} \leq 0$). The overlap integral will be denoted by $S \equiv \langle \phi_1|\phi_2 \rangle = \langle \phi_2|\phi_1 \rangle$. After introducing the abbreviation $h \equiv H_{12}$ we have

$$\begin{aligned} c_1(H_{11} - E) + c_2(h - ES) &= 0, \\ c_1(h - ES) + c_2(H_{22} - E) &= 0. \end{aligned}$$

A nontrivial solution of these secular equations exists only if the secular determinant satisfies

$$\begin{vmatrix} H_{11} - E & h - ES \\ h - ES & H_{22} - E \end{vmatrix} = 0.$$

---

[1] As a few examples just recall the CI, VB (Chapter V2-2), and MO (Chapter 8) methods.

[2] The same set of equations ("secular equations") is obtained after using the Ritz method (Chapter 5).

[3] This pertains to almost all applications. For complex functions the equations are only slightly more complicated.

After expanding the determinant we obtain a quadratic equation for the unknown energy E, i.e.,

$$(H_{11}-E)(H_{22}-E)-(h-ES)^2=0$$

with its two solutions[4]

$$E_{\pm}=\frac{1}{1-S^2}\left\{\tfrac{H_{11}+H_{22}}{2}-hS\mp\sqrt{\left(\tfrac{H_{11}-H_{22}}{2}\right)^2+(h-S\sqrt{H_{11}H_{22}})^2+2hS\left(\sqrt{H_{11}H_{22}}-\tfrac{H_{11}+H_{22}}{2}\right)}\right\}.$$

After inserting the above energies into the secular equations we obtain the following two sets of solutions c_1 and c_2:

$$\left(\frac{c_1}{c_2}\right)_{\pm}=\frac{1}{(h-H_{11}S)}\left\{\frac{H_{11}-H_{22}}{2}\pm\sqrt{\left(\frac{H_{11}-H_{22}}{2}\right)^2+(h-H_{11}S)(h-H_{22}S)}\right\}.$$

Using the abbreviations

$$\Delta=\frac{H_{11}-H_{22}}{2}$$

and $E_{ar}=\frac{H_{11}+H_{22}}{2}$ for the arithmetic mean, as well as $E_{geom}=\sqrt{H_{11}H_{22}}$ for the geometric mean, we get a simpler formula for the energy,

$$E_{\pm}=\frac{1}{1-S^2}\left\{\ E_{ar}-hS\mp\sqrt{\Delta^2+(h-SE_{geom})^2+2hS\left(E_{geom}-E_{ar}\right)}\ \right\}.$$

Now, let us consider some important special cases.

Case I. $H_{11}=H_{22}$ and $S=0$ (ϕ_1 and ϕ_2 correspond to the same energy and do not overlap).

Then $\Delta=0$, $E_{ar}=E_{geom}=H_{11}$ and we have

$$E_{\pm}=H_{11}\pm h,$$
$$\left(\frac{c_1}{c_2}\right)_{\pm}=\pm 1.$$

---

[4] Most practical is to use the Mathematica coding

```
Solve[(H11-EdS)*(H22-EdS)-(h-EdS*S)^2==0,EdS]
Solve[(H11-EdS)*(H22-EdS)-(h-EdS*S)^2==0
&&c1*(H11-EdS)+c2*(h-EdS*S)==0
&&c1*(h-EdS*S)+c2*(H22-EdS)==0,{c1,c2},{EdS}]
```

For $h < 0$ this means that E_+ corresponds to stabilization (with respect to ϕ_1 or ϕ_2 states), while E_- corresponds to destabilization (by the same value of $|h|$). The wave functions contain equal contributions of ϕ_1 and ϕ_2 and (after normalization) are

$$\psi_+ = \frac{1}{\sqrt{2}}\left(\phi_1 + \phi_2\right),$$
$$\psi_- = \frac{1}{\sqrt{2}}\left(\phi_1 - \phi_2\right).$$

Case II. $H_{11} = H_{22}$ and $S \neq 0$ (ϕ_1 and ϕ_2 correspond to the same energy, but their overlap integral is nonzero).

Then

$$E_\pm = \frac{H_{11} \pm h}{1 \pm S},$$
$$\left(\frac{c_1}{c_2}\right)_\pm = \pm 1.$$

Here also E_+ corresponds to stabilization, while E_- corresponds to destabilization (because of the denominator this time the destabilization is larger than the stabilization). The wave functions have the same contributions of ϕ_1 and ϕ_2 and (after normalization) are equal to

$$\psi_+ = \frac{1}{\sqrt{2\left(1+S\right)}}\left(\phi_1 + \phi_2\right),$$
$$\psi_- = \frac{1}{\sqrt{2\left(1-S\right)}}\left(\phi_1 - \phi_2\right).$$

Case III. $H_{11} \neq H_{22}$ and $S = 0$ (ϕ_1 and ϕ_2 correspond to different energies and the overlap integral is equal to zero).

This time

$$E_\pm = E_{ar} \mp \sqrt{\Delta^2 + h^2},$$
$$\left(\frac{c_1}{c_2}\right)_\pm = \frac{1}{h}\left(\Delta \pm \sqrt{\Delta^2 + h^2}\right). \tag{D.1}$$

Here also the state with E_+ means stabilization, while E_- corresponds to destabilization (both effects are equal).

Let us consider a limit case when the mean energy in the state ϕ_1 is much lower than that in ϕ_2 ($H_{11} \ll H_{22}$), and in addition $\frac{\Delta}{h} \gg 0$. For the state with the energy E_+ we have $\frac{c_1}{c_2} \simeq \frac{2\Delta}{h}$, i.e.,

c_1 is very large, while c_2 is very small (this means that ψ_+ is very similar to ϕ_1). In the state ψ_-, the same ratio of the coefficients equals $\frac{c_1}{c_2} \simeq 0$, which means a domination of ϕ_2.

Thus, if two states differ very much by their energies (or h is small, which means the overlap integral is also small), they practically do not change (do not mix together).

This is why at the beginning of this appendix we admitted only ϕ_1 and ϕ_2 of comparable energies.

Dirac Delta Function

Paul Dirac introduced some useful formal tools (like for example his notation for integrals and operators), including an object then unknown to mathematicians (!), which turned out to be very useful in physics. This was the Dirac delta function $\delta(x)$. We may think of it as a function[1]

- which is nonzero only very close to $x = 0$, where its value is $+\infty$;
- the surface under its plot is equal to 1, which is highlighted by a *symbolic* equation

$$\int_{-\infty}^{\infty} \delta(x)\ dx = 1.$$

When we look at a straight thin reed protruding from a lake (the water level = 0), then we have to do with something similar to the Dirac delta function. The only task of the Dirac delta function is its specific behavior when integrating the product $f(x)\delta(x)$ and the integration *includes the point* $x = 0$, namely,

$$\int_{a}^{b} f(x)\ \delta(x)\ dx = f(0). \tag{E.1}$$

This result is well understandable: the integral means the surface under the curve $f(x)\delta(x)$, but since $\delta(x)$ is so concentrated at $x = 0$, it pays to take seriously only those $f(x)$ that are "extremely close" to $x = 0$. Over there $f(x)$ is equal to $f(0)$. The constant $f(0)$ can be extracted from the integral, which itself therefore has the form $\int_{-\infty}^{\infty} \delta(x)\ dx = 1$; this is why we get the right-hand side of the last equation. Of course, $\delta(x - c)$ represents the same narrow peak, but at $x = c$; therefore, for $a \le c \le b$ we have

$$\boxed{\int_{a}^{b} f(x)\ \delta(x-c)\ dx = f(c).} \tag{E.2}$$

---

[1] More precisely, this is not a function, but what is called a distribution. The theory of distributions has been developed by mathematicians only after Dirac.

Approximations to $\delta(x)$

The Dirac delta function $\delta(x)$ can be approximated by many functions that depend on a certain parameter and have the following properties:

- when the parameter tends to a limit, the values of the functions for x distant from 0 become smaller and smaller, while for x close to zero are getting larger and larger (a peak close to $x = 0$);
- the integral of the function tends to (or is close to) 1 when the parameter approaches its limit value.

Here are several functions that approximate the Dirac delta function:

- a rectangular function centered at $x = 0$ with the rectangle surface equal to 1 ($a \to 0$):

$$f_1(x; a) = \begin{cases} \frac{1}{a} & \text{for} - \frac{a}{2} \le x \le \frac{a}{2}, \\ 0 & \text{for other,} \end{cases}$$

- a (normalized to 1) Gaussian function[2] ($a \to \infty$):

$$f_2(x; a) = \sqrt{\frac{a}{\pi}} e^{-ax^2},$$

- a function:

$$f_3(x; a) = \frac{1}{\pi} \lim \frac{\sin ax}{x} \qquad \text{when} \quad a \to \infty,$$

- the last function is (we will use it when considering the interaction of matter with radiation)[3]:

---

[2] Let us see how an approximation $f_2 = \sqrt{\frac{a}{\pi}} e^{-ax^2}$ does the job of the Dirac delta function when $a \to \infty$. Let us take a function $f(x) = (x - 5)^2$ and consider the integral

$$\int_{-\infty}^{\infty} f(x) \quad f_2(x; a)\, dx = \sqrt{\frac{a}{\pi}} \int_{-\infty}^{\infty} (x-5)^2 e^{-ax^2} dx = \sqrt{\frac{a}{\pi}} \left(\frac{1}{4a} \sqrt{\frac{\pi}{a}} + 0 + 25 \sqrt{\frac{\pi}{a}} \right) = \frac{1}{4a} + 25.$$

When $a \to \infty$, the value of the integral tends to $25 = f(0)$, as it has to be for the Dirac delta function used instead of f_2.

[3] The function under the limit symbol may be treated as $A[\sin(ax)]^2$ with amplitude A *decaying* as $A = 1/x^2$ when $|x| \to \infty$. For small values of x the $\sin(ax)$ changes as ax (as seen from its Taylor expansion), hence for small x the function changes as a^2. This means that when $a \to \infty$, there will be a dominant peak close to $x = 0$, although there will be some smaller side peaks clustering around $x = 0$. The surface of the dominating peak may be approximated by a triangle of the base $2\pi/a$ and the height a^2; we obtain its surface equal to πa – hence the "approximate normalization factor" $1/(\pi a)$ in f_4.

$$f_4(x;a) = \frac{1}{\pi a} \lim \frac{\sin^2(ax)}{x^2} \qquad \text{when} \quad a \to \infty.$$

Properties of $\delta(x)$

Function $\delta(cx)$

Let us see what $\delta(cx)$ is:

$$\delta(cx) = \lim_{a\to\infty} \sqrt{\frac{a}{\pi}} \exp\left(-ac^2x^2\right) = \lim_{a\to\infty} \sqrt{\frac{ac^2}{\pi c^2}} \exp\left(-ac^2x^2\right) =$$

$$\frac{1}{|c|} \lim_{ac^2\to\infty} \sqrt{\frac{ac^2}{\pi}} \exp\left(-ac^2x^2\right) = \frac{1}{|c|}\delta(x).$$

Therefore,

$$\boxed{\delta(cx) = \frac{1}{|c|}\delta(x). \qquad \text{(E.3)}}$$

Dirac δ in 3D

The three-dimensional Dirac delta function is defined in the Cartesian coordinate system as

$$\delta(\mathbf{r}) = \delta(x)\,\delta(y)\,\delta(z),$$

where $\mathbf{r} = (x, y, z)$. Then $\delta(\mathbf{r})$ denotes a peak of infinite height at $\mathbf{r} = \mathbf{0}$, $\delta(\mathbf{r} - \mathbf{A})$ denotes an identical peak at the position shown by the vector $\mathbf{A}$ from the origin. Each of the peaks is normalized to 1, i.e., the integral over the whole three-dimensional space is equal to 1. This means that Eq. (E.1) is satisfied, but this time $x \in R^3$.

An application of the Dirac delta function

When may a concept such as the Dirac delta function be useful? Here is an example. Let us imagine that we have (in three-dimensional space) two molecular charge distributions, $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$. Each of the distributions consists of the electronic part and the nuclear part.

How can such charge distributions be represented mathematically? There is no problem in mathematical representation of the electronic parts; they are simply some functions of the position $\mathbf{r}$ in space, i.e., $-\rho_{el,A}(\mathbf{r})$ and $-\rho_{el,B}(\mathbf{r})$, respectively. The integrals of the corresponding electronic distributions yield, of course, $-N_A$ and $-N_B$ (in a.u.), or minus the number of the electrons (minus, because the electrons carry negative charge). How to write the nuclear charge distribution as a function of $\mathbf{r}$? There is no way to do this without the Dirac delta function. With the function our task gets simple:

$$\rho_{nucl,A}(\mathbf{r}) = \sum_{a \in A} Z_{A,a} \delta(\mathbf{r} - \mathbf{r}_a),$$
$$\rho_{nucl,B}(\mathbf{r}) = \sum_{b \in B} Z_{B,b} \delta(\mathbf{r} - \mathbf{r}_b).$$

Just at the nuclear positions we put delta functions with the "intensities" equal to the nuclear charges. For neutral molecules $\int \rho_{nucl,A}(\mathbf{r})\, d\mathbf{r}$ and $\int \rho_{nucl,B}(\mathbf{r})\, d\mathbf{r}$ have to give $+N_A$ and $+N_B$, respectively. Indeed, we have

$$\int \rho_{nucl,A}(\mathbf{r})\, d\mathbf{r} = \sum_{a \in A} Z_{A,a} \int \delta(\mathbf{r} - \mathbf{r}_a)\, d\mathbf{r} = \sum_{a \in A} Z_{A,a} = N_A,$$
$$\int \rho_{nucl,B}(\mathbf{r})\, d\mathbf{r} = \sum_{b \in B} Z_{B,b} \int \delta(\mathbf{r} - \mathbf{r}_b)\, d\mathbf{r} = \sum_{b \in B} Z_{B,b} = N_B.$$

Thus, the Dirac delta function enables us to write the total charge distributions and their interactions in an elegant way:

$$\rho_A(\mathbf{r}) = -\rho_{el,A}(\mathbf{r}) + \rho_{nucl,A}(\mathbf{r}),$$
$$\rho_B(\mathbf{r}) = -\rho_{el,B}(\mathbf{r}) + \rho_{nucl,B}(\mathbf{r}).$$

To demonstrate the difference, let us write the electrostatic interaction of the two charge distributions without and with the Dirac delta functions. The first way gives the following expression:

$$\begin{aligned} E_{inter} = {} & \sum_{a \in A} \sum_{b \in B} \frac{Z_A Z_B}{|\mathbf{r}_a - \mathbf{r}_b|} - \sum_{a \in A} \int \rho_{el,B}(\mathbf{r}) \frac{Z_A}{|\mathbf{r} - \mathbf{r}_a|}\, d\mathbf{r} + \\ & - \sum_{b \in B} \int \rho_{el,A}(\mathbf{r}) \frac{Z_B}{|\mathbf{r} - \mathbf{r}_b|} d\mathbf{r} + \int \int \frac{\rho_{el,A}(\mathbf{r}) \rho_{el,B}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}\, d\mathbf{r}'. \end{aligned}$$

The four terms mean the interactions of nuclei of A with nuclei of B, nuclei of A with electrons of B, electrons of A with nuclei of B, and electrons of A with electrons of B, respectively. With the Dirac delta function the same expression reads

$$E_{inter} = \int \frac{\rho_A(\mathbf{r})\,\rho_B\left(\mathbf{r}'\right)}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'.$$

The last expression comes from the definition of the Coulomb interaction and the definition of the integral.[4]

No matter how the charge distributions look, whether they are diffuse (like the electronic ones) or point-like (like those of the nuclei), the formula is always the same.

---

[4] Of course, the two notations are equivalent, because inserting the total charge distributions into the last integral, as well as using the properties of the Dirac delta function, gives the first expression for E_{inter}.

APPENDIX F

Translation Versus Momentum and Rotation Versus Angular Momentum

In Chapter 2 it has been shown that the Hamiltonian $\hat{H}$ commutes with any translation (p. 77) or rotation (p. 79) operator, denoted as $\hat{\mathcal{U}}$:

$$\left[\hat{H},\hat{\mathcal{U}}\right]=0. \tag{F.1}$$

The form of the $\hat{\mathcal{U}}$ operator

Below it will be demonstrated for $\boldsymbol{\kappa}$ meaning first a translation vector, and then a rotation angle about an axis in three-dimensional space, that operator $\hat{\mathcal{U}}$ is of the form

$$\hat{\mathcal{U}}=\exp(-\frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\mathbf{K}}), \tag{F.2}$$

where $\hat{\mathbf{K}}$ stands for a Hermitian operator (having the x, y, z components) acting on functions of points in the three-dimensional Cartesian space.

Translation and momentum operators

Translation of a function by a vector $\Delta\mathbf{r}$ is equivalent to the function f in the coordinate system translated in the opposite direction, i.e., $f(\mathbf{r}-\Delta\mathbf{r})$ (see Fig. 1.3 and p. 77). If the vector $\Delta\mathbf{r}$ is *infinitesimally small, then, in order to establish the relation between* $f(\mathbf{r}-\Delta\mathbf{r})$ *and* $f(\mathbf{r})$*, it is of course sufficient to know the gradient of* f (neglecting, obviously, the quadratic and higher terms in the Taylor expansion):

$$f(\mathbf{r}-\Delta\mathbf{r})=f(\mathbf{r})-\Delta\mathbf{r}\cdot\nabla f=(1-\Delta\mathbf{r}\cdot\nabla)f(\mathbf{r}). \tag{F.3}$$

We will compose a large translation of a function (by vector $\mathbf{T}$) from a number of small increments $\Delta\mathbf{r}=\frac{1}{N}\mathbf{T}$, where N is a very, very large natural number. Such a tiny translation will be repeated N times, thus recovering the translation of the function by $\mathbf{T}$. In order for the gradient formula to be exact, one has to ensure N tends to infinity. Recalling the definition $\exp(ax) = \lim_{N\to\infty}\left(1+\frac{a}{x}\right)^N$, we have

$$\hat{\mathcal{U}}(\mathbf{T})f(\mathbf{r}) = f(\mathbf{r}-\mathbf{T}) = \lim_{N\to\infty}(1-\frac{\mathbf{T}}{N}\nabla)^N f(r) = \exp(-\mathbf{T}\cdot\nabla)f = \exp(-\frac{i}{\hbar}\mathbf{T}\cdot\hat{\mathbf{p}})f(\mathbf{r}),$$

where $\hat{\mathbf{p}} = -i\hbar\nabla$ is the total momentum operator (see Chapter 1). Thus, for translations we have $\boldsymbol{\kappa} \equiv \mathbf{T}$ and $\hat{\mathbf{K}} \equiv \hat{\mathbf{p}}$.

Rotation and angular momentum operator

Imagine a function $f(\mathbf{r})$ of positions in the three-dimensional Cartesian space (think, e.g., about a probability density distribution centered somewhere in the space). Now, suppose the function is to be rotated about the z axis (the unit vector showing its direction is $\mathbf{e}$) by an angle α, so we obtain another function; let us denote it by $\hat{\mathcal{U}}(\alpha;\mathbf{e})f(\mathbf{r})$. What is the relation between $f(\mathbf{r})$ and $\hat{\mathcal{U}}(\alpha;\mathbf{e})f(\mathbf{r})$? This is what we want to establish. This relation corresponds to the opposite rotation (i.e., by the angle $-\alpha$, see Fig. 1.1 and p. 79) of the coordinate system, i.e.,

$$\hat{\mathcal{U}}(\alpha;\mathbf{e})f(\mathbf{r}) = f\left(\mathbf{U}^{-1}\mathbf{r}\right) = f(\mathbf{U}(-\alpha;\mathbf{e})\mathbf{r}),$$

where $\mathbf{U}$ is a 3×3 orthogonal matrix. The new coordinates $x(\alpha)$, $y(\alpha)$, $z(\alpha)$ are expressed by the old coordinates x, y, z through[1]

$$\mathbf{r}' \equiv \begin{pmatrix} x(\alpha) \\ y(\alpha) \\ z(\alpha) \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Therefore, the rotated function $\hat{\mathcal{U}}(\alpha;\mathbf{e})f(\mathbf{r}) = f(x(\alpha), y(\alpha), z(\alpha))$. The function can be expanded in the Taylor series about $\alpha = 0$ as follows:

$$\hat{\mathcal{U}}(\alpha;\mathbf{e})f(\mathbf{r}) = f(x(\alpha), y(\alpha), z(\alpha)) = f(x,y,z) + \alpha\left(\frac{\partial f}{\partial\alpha}\right)_{\alpha=0} + \ldots =$$

$$f(x,y,z) + \alpha\left(\frac{\partial x(\alpha)}{\partial\alpha}\frac{\partial f}{\partial x} + \frac{\partial y(\alpha)}{\partial\alpha}\frac{\partial f}{\partial y} + \frac{\partial z(\alpha)}{\partial\alpha}\frac{\partial f}{\partial z}\right)_{\alpha=0} + \ldots =$$

$$f(x,y,z) + \alpha\left[y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right]f + \ldots.$$

---

[1] A positive value of the rotation angle means a counterclockwise motion within the xy plane (x axis horizontal, y vertical, z axis pointing to us).

Now, instead of the large rotation angle α let us consider first an infinitesimally small rotation by angle $\varepsilon = \alpha/N$, where N is a very, very large natural number. In such a situation we retain only the first two terms in the previous equation:

$$\hat{\mathcal{U}}\left(\frac{\alpha}{N};\mathbf{e}\right) f(\mathbf{r}) = f(x,y,z) + \frac{\alpha}{N}\left[y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right] f(x,y,z) =$$
$$\left(1 + \frac{\alpha}{N}\frac{i\hbar}{i\hbar}\left[y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right]\right) f = \left(1 + \frac{\alpha}{N}\frac{1}{i\hbar}\left[x\hat{p}_y - y\hat{p}_x\right]\right) f = \left(1 - \frac{\alpha}{N}\frac{i}{\hbar}\hat{J}_z\right) f.$$

If such a rotation is repeated N times, then we recover the rotation of the function by a (possibly large) angle α (the limes ensure that ε is infinitesimally small):

$$\hat{\mathcal{U}}(\alpha;\mathbf{e}) f(\mathbf{r}) = \lim_{N\to\infty}\left[\hat{\mathcal{U}}\left(\frac{\alpha}{N};\mathbf{e}\right)\right]^N f(\mathbf{r}) = \lim_{N\to\infty}\left(1 - \frac{\alpha}{N}\frac{i}{\hbar}\hat{J}_z\right)^N f(\mathbf{r}) =$$
$$\exp\left(-i\frac{\alpha}{\hbar}\hat{J}_z\right) f = \exp\left(-\frac{i}{\hbar}\alpha\mathbf{e}\cdot\hat{\mathbf{J}}\right) f.$$

Thus, for rotations $\hat{\mathcal{U}}(\alpha;\mathbf{e}) = \exp\left(-\frac{i}{\hbar}\alpha\mathbf{e}\cdot\hat{\mathbf{J}}\right)$, and, therefore, we have $\boldsymbol{\kappa} \equiv \alpha\mathbf{e}$ and $\hat{\mathbf{K}} \equiv \hat{\mathbf{J}}$.

This means that in particular, for rotations about the x, y, z axes (with the corresponding unit vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$) we have, respectively,

$$\left[\hat{\mathcal{U}}(\alpha;\mathbf{x}), \hat{J}_x\right] = 0, \tag{F.4}$$
$$\left[\hat{\mathcal{U}}(\alpha;\mathbf{y}), \hat{J}_y\right] = 0, \tag{F.5}$$
$$\left[\hat{\mathcal{U}}(\alpha;\mathbf{z}), \hat{J}_z\right] = 0. \tag{F.6}$$

Useful relation

The relation (F.1) means that for any translation or rotation

$$\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1} = \hat{H}$$

and taking into account the general form of Eq. (F.2) we have for any such transformation a series containing nested commutators (valid for any $\boldsymbol{\kappa}$)

$$\hat{H} = \hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1} = \exp\left(-\frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\mathbf{K}}\right)\hat{H}\exp\left(\frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\mathbf{K}}\right) =$$
$$\left(1 - \frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\mathbf{K}} + \ldots\right)\hat{H}\left(1 + \frac{i}{\hbar}\boldsymbol{\kappa}\cdot\hat{\mathbf{K}} + \ldots\right)\hat{H} - \frac{i}{\hbar}\boldsymbol{\kappa}\cdot[\hat{\mathbf{K}},\hat{H}] - \frac{\kappa^2}{2\hbar^2}[[\hat{\mathbf{K}},\hat{H}],\hat{\mathbf{K}}] + \ldots,$$

where each term in "+..." contains $[\hat{\mathbf{K}}, \hat{H}]$. This means that to satisfy the equation we have necessarily

$$[\hat{\mathbf{K}}, \hat{H}] = \mathbf{0}. \tag{F.7}$$

Hamiltonian commutes with the total momentum operator

In particular this means $[\hat{\mathbf{p}}, \hat{H}] = \mathbf{0}$, i.e.,

$$[\hat{p}_\mu, \hat{H}] = 0, \tag{F.8}$$

for $\mu = x, y, z$. Of course, we have also $[\hat{p}_\mu, \hat{p}_\nu] = 0$ for $\mu, \nu = x, y, z$.

Since all these four operators mutually commute, the total wave function is simultaneously an eigenfunction of $\hat{H}$ and $\hat{p}_x, \hat{p}_y, \hat{p}_z$, i.e., the energy and the momentum of the center of mass can be both measured (without making any error) in a space-fixed coordinate system (see Appendix J). By definition the momentum of the center of mass is identical with the total momentum.[2]

Hamiltonian, $\hat{J}^2$, and $\hat{J}_z$ do commute

Eq. (F.7) for rotations means $[\hat{\mathbf{J}}, \hat{H}] = \mathbf{0}$, i.e., in particular

$$[\hat{J}_x, \hat{H}] = 0, \tag{F.9}$$

$$[\hat{J}_y, \hat{H}] = 0, \tag{F.10}$$

$$[\hat{J}_z, \hat{H}] = 0. \tag{F.11}$$

The components of the angular momentum operators satisfy the following commutation rules[3]:

$$[\hat{J}_x, \hat{J}_y] = i\hbar \hat{J}_z,$$

---

[2] Indeed, the position vector of the center of mass is defined as $\mathbf{R}_{CM} = \frac{\sum_i m_i \mathbf{r}_i}{\sum_i m_i}$, and after differentiation with respect to time $(\sum_i m_i)\dot{\mathbf{R}}_{CM} = \sum_i m_i \dot{\mathbf{r}}_i = \sum_i \mathbf{p}_i$. The right-hand side represents the momentum of all the particles (i.e., the total momentum), whereas the left-hand side is just the momentum of the center of mass.

[3] The commutation relations can be obtained by using directly the definitions of the operators involved: $\hat{J}_x = y\hat{p}_z - z\hat{p}_y$, etc.
For example, $[\hat{J}_x, \hat{J}_y]f = \left[\left(y\hat{p}_z - z\hat{p}_y\right)\left(z\hat{p}_x - x\hat{p}_z\right) - \left(z\hat{p}_x - x\hat{p}_z\right)\left(y\hat{p}_z - z\hat{p}_y\right)\right] f =$
$\left[\left(y\hat{p}_z z\hat{p}_x - z\hat{p}_x y\hat{p}_z\right) - \left(y\hat{p}_z x\hat{p}_z - x\hat{p}_z y\hat{p}_z\right) - \left(z\hat{p}_y z\hat{p}_x - z\hat{p}_x z\hat{p}_y\right) + \left(z\hat{p}_y x\hat{p}_z - x\hat{p}_z z\hat{p}_y\right)\right] f =$
$\left(y\hat{p}_z z\hat{p}_x - z\hat{p}_x y\hat{p}_z\right) f - \left(yx\hat{p}_z\hat{p}_z - yx\hat{p}_z\hat{p}_z\right) f - \left(z^2\hat{p}_y\hat{p}_x - z^2\hat{p}_x\hat{p}_y\right) f + \left(xz\hat{p}_y\hat{p}_z - x\hat{p}_z z\hat{p}_y\right) f =$
$\left(y\hat{p}_z z\hat{p}_x - yz\hat{p}_x\hat{p}_z\right) f - 0 - 0 + \left(xz\hat{p}_y\hat{p}_z - x\hat{p}_z z\hat{p}_y\right) f = (-i\hbar)^2\left[y\frac{\partial f}{\partial x} - x\frac{\partial f}{\partial y}\right] = i\hbar\hat{J}_z f.$

$$[\hat{J}_y, \hat{J}_z] = i\hbar\hat{J}_x, \tag{F.12}$$
$$[\hat{J}_z, \hat{J}_x] = i\hbar\hat{J}_y.$$

Eqs. (F.9)–(F.11) are not independent; e.g., from Eqs. (F.9) and (F.10), Eq. (F.11) can be derived. Indeed,

$$[\hat{J}_z, \hat{H}] = \hat{J}_z\hat{H} - \hat{H}\hat{J}_z = \frac{1}{i\hbar}[\hat{J}_x, \hat{J}_y]\hat{H} - \frac{1}{i\hbar}\hat{H}[\hat{J}_x, \hat{J}_y] = \frac{1}{i\hbar}[\hat{J}_x, \hat{J}_y]\hat{H} - \frac{1}{i\hbar}[\hat{J}_x, \hat{J}_y]\hat{H} = 0.$$

Also, from Eqs. (F.9), (F.10), and (F.11) it follows that

$$[\hat{J}^2, \hat{H}] = 0, \tag{F.13}$$

because from the Pythagoras theorem $\hat{J}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$.

Do $\hat{J}_x, \hat{J}_y, \hat{J}_z$ commute with $\hat{J}^2$? Let us check the commutator $[\hat{J}_z, \hat{J}^2]$:

$$[\hat{J}_z, \hat{J}^2] = [\hat{J}_z, \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2] = [\hat{J}_z, \hat{J}_x^2 + \hat{J}_y^2] = \hat{J}_z\hat{J}_x^2 - \hat{J}_x^2\hat{J}_z + \hat{J}_z\hat{J}_y^2 - \hat{J}_y^2\hat{J}_z =$$
$$(i\hbar\hat{J}_y + \hat{J}_x\hat{J}_z)\hat{J}_x - \hat{J}_x(-i\hbar\hat{J}_y + \hat{J}_z\hat{J}_x) + (-i\hbar\hat{J}_x + \hat{J}_y\hat{J}_z)\hat{J}_y - \hat{J}_y(i\hbar\hat{J}_x + \hat{J}_z\hat{J}_y) = 0.$$

Thus,

$$[\hat{J}_z, \hat{J}^2] = 0, \tag{F.14}$$

and also by the argument of symmetry (the space is isotropic)

$$[\hat{J}_x, \hat{J}^2] = 0, \tag{F.15}$$
$$[\hat{J}_y, \hat{J}^2] = 0. \tag{F.16}$$

Now, we need to determine the set of the operators that all mutually commute. Only then all the physical quantities to which the operators correspond can have definite values, when measured. Also the wave function can be an eigenfunction of all of these operators and it can be labeled by the quantum numbers, each corresponding to an eigenvalue of the operators in question. *We cannot choose as these operators the whole set of* $\hat{H}, \hat{J}_x, \hat{J}_y, \hat{J}_z, \hat{J}^2$, *because as shown before* $\hat{J}_x, \hat{J}_y, \hat{J}_z$ *do not commute among themselves* (although they do with $\hat{H}$ and $\hat{J}^2$).

> The only way is to choose as the set of the operators *either* $\hat{H}, \hat{J}_z, \hat{J}^2$ *or* $\hat{H}, \hat{J}_x, \hat{J}^2$ *or* $\hat{H}, \hat{J}_y, \hat{J}^2$. Traditionally, one chooses as the set of the mutually commuting operators $\hat{H}, \hat{J}_z, \hat{J}^2$ (z is known as *quantization axis*).

Rotation and translation operators do not commute

Now, we may think about adding the operators $\hat{p}_x, \hat{p}_y, \hat{p}_z$ to the above set of operators. The operators $\hat{H}, \hat{p}_x, \hat{p}_y, \hat{p}_z, \hat{J}^2$ and $\hat{J}_z$ *do not* represent a set of mutually commuting operators. The reason for this is that $[\hat{p}_\mu, \hat{J}_\nu] \neq 0$ for $\mu \neq \nu$, which is a consequence of the fact that in general rotation and translation operators do not commute, as shown in Fig. F.1.

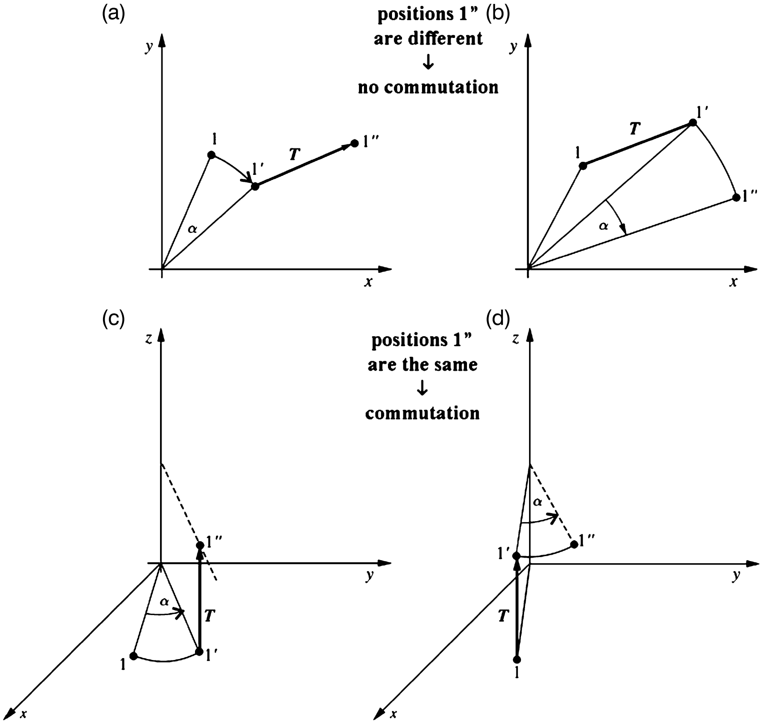


Fig. F.1. In general, translation $\hat{\mathcal{U}}(\mathbf{T})$ and rotation $\hat{\mathcal{U}}(\alpha; \mathbf{e})$ operators do not commute. The example shows what happens to a point belonging to the xy plane, (a) if a rotation $\hat{\mathcal{U}}(\alpha; \mathbf{z})$ by angle α about the z axis takes place first and then a translation $\hat{\mathcal{U}}(\mathbf{T})$ by a vector $\mathbf{T}$ (restricted to the xy plane) is carried out and (b) if the operations are applied in reverse order. As we can see the results are different (two points $1''$ have different positions in (a) and (b)), i.e., the two operators do not commute: $\hat{\mathcal{U}}(\mathbf{T})\hat{\mathcal{U}}(\alpha; \mathbf{z}) \neq \hat{\mathcal{U}}(\alpha; \mathbf{z})\hat{\mathcal{U}}(\mathbf{T})$. This, after expanding $\hat{\mathcal{U}}(\mathbf{T}) = \exp[-\frac{i}{\hbar}\left(T_x \hat{p}_x + T_y \hat{p}_y\right)]$ and $\hat{\mathcal{U}}(\alpha; \mathbf{z}) = \exp(-\frac{i}{\hbar}\alpha \hat{J}_z)$ in Taylor series and taking into account that T_x, T_y, α are arbitrary numbers, leads to the conclusion that $[\hat{J}_z, \hat{p}_x] \neq 0$ and $[\hat{J}_z, \hat{p}_y] \neq 0$. (c,d) Note that *some* translations and rotations do commute, e.g., $[\hat{J}_z, \hat{p}_z] = [\hat{J}_x, \hat{p}_x] = [\hat{J}_y, \hat{p}_y] = 0$, because we see by inspection that any translation by $\mathbf{T} = (0, 0, T_z)$ is independent of any rotation about the z axis, etc.

Conclusion

It is, therefore, impossible to make all the operators $\hat{H}$, $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$, $\hat{J}^2$ and $\hat{J}_z$ commute in a *space-fixed coordinate system*. What we are able to do though is to write down the total wave function Ψ_{pN} in the space-fixed coordinate system as a product of the plane wave $\exp(i\mathbf{p}_{CM} \cdot \mathbf{R}_{CM})$ depending on the center-of-mass variables and the wave function Ψ_{0N} depending on internal coordinates[4]

$$\Psi_{pN} = \Psi_{0N}\ \exp(i\mathbf{p}_{CM} \cdot \mathbf{R}_{CM}), \tag{F.17}$$

which is an eigenfunction of the total (i.e., center-of-mass) momentum operators: $\hat{p}_x = \hat{p}_{CM,x},\ \hat{p}_y = \hat{p}_{CM,y},\ \hat{p}_z = \hat{p}_{CM,z}$. The function Ψ_{0N} is the total wave function written in the center-of-mass coordinate system (a special body-fixed coordinate system, see Appendix J), in which the total angular momentum operators $\hat{J}^2$ and $\hat{J}_z$ are now defined. The three operators $\hat{H}$, $\hat{J}^2$, and $\hat{J}_z$ commute in any space-fixed or body-fixed coordinate system (including the center-of-mass coordinate system) and, therefore, the corresponding physical quantities (energy and angular momentum) have exact values. In this particular coordinate system $\hat{\mathbf{p}} = \hat{\mathbf{p}}_{CM} = \mathbf{0}$. We may say, therefore, that

> *in the center-of-mass coordinate system* $\hat{H}$, $\hat{p}_x$, $\hat{p}_y$, $\hat{p}_z$, $\hat{J}^2$, and $\hat{J}_z$ all do commute.

---

[4] See Chapter 2 and Appendix J, where the total Hamiltonian is split into a sum of the center-of-mass and internal coordinates Hamiltonians; N is the quantum number for the spectroscopic states.

APPENDIX G

Vector and Scalar Potentials

Maxwell equations

The electromagnetic field is described by two vector fields: the electric field intensity $\boldsymbol{\mathcal{E}}$ and the magnetic field intensity $\mathbf{H}$, both depending on the position in space (Cartesian coordinates x, y, z) and time t. The vectors $\boldsymbol{\mathcal{E}}$ and $\mathbf{H}$ are determined by the electric charges and their currents. The charges are defined by the charge density function $\rho(x, y, z, t)$ such that $\rho(x, y, z, t)\mathrm{d}V$ at time t represents the charge in the infinitesimal volume $\mathrm{d}V$ that contains the point (x, y, z). The velocity of the charge in position x, y, z measured at time t represents the vector field $\mathbf{v}(x, y, z, t)$, while the current in point x, y, z measured at t is equal to $\mathbf{i}(x, y, z, t) = \rho(x, y, z, t)\mathbf{v}(x, y, z, t)$.

It turned out (as shown by James Maxwell) that $\mathbf{H}$, $\boldsymbol{\mathcal{E}}$, ρ, and $\mathbf{i}$ are interrelated by the Maxwell equations (c stands for the speed of light)

$$\nabla \times \boldsymbol{\mathcal{E}} + \frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = \mathbf{0}, \tag{G.1}$$

$$\nabla \times \mathbf{H} - \frac{1}{c}\frac{\partial \boldsymbol{\mathcal{E}}}{\partial t} = \frac{4\pi}{c}\mathbf{i}, \tag{G.2}$$

$$\nabla \cdot \boldsymbol{\mathcal{E}} = 4\pi\rho, \tag{G.3}$$

$$\nabla \cdot \mathbf{H} = 0. \tag{G.4}$$

The Maxwell equations have an alternative notation, which involves two new quantities: the *scalar potential* ϕ and the *vector potential* $\mathbf{A}$, which replace $\boldsymbol{\mathcal{E}}$ and $\mathbf{H}$:

$$\boldsymbol{\mathcal{E}} = -\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}, \tag{G.5}$$

$$\mathbf{H} = \nabla \times \mathbf{A}. \tag{G.6}$$

After inserting $\boldsymbol{\mathcal{E}}$ and $\mathbf{H}$ from the last equation into the Maxwell equations, Eqs. (G.1) and (G.4), we obtain their automatic satisfaction (for smooth vector components):

$$\nabla \times \boldsymbol{\mathcal{E}} + \frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = \nabla \times (-\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}) + \frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \nabla\phi - \frac{1}{c}\frac{\partial \nabla \times \mathbf{A}}{\partial t} + \frac{1}{c}\frac{\partial \mathbf{H}}{\partial t} = 0$$

and

$$\nabla \cdot (\nabla \times \mathbf{A}) = \frac{\partial}{\partial x}(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}) + \frac{\partial}{\partial y}(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}) + \frac{\partial}{\partial z}(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}) = 0,$$

because[1] $\nabla \times \nabla\phi = 0$ and $\nabla \times \mathbf{A} = \mathbf{H}$, while Eqs. (G.2) and (G.3) transform into

$$\nabla \times (\nabla \times \mathbf{A}) + \frac{1}{c}\frac{\partial \nabla\phi}{\partial t} + \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = \frac{4\pi}{c}\mathbf{i},$$
$$-\nabla \cdot (\nabla\phi) - \frac{1}{c}\frac{\partial \nabla \cdot \mathbf{A}}{\partial t} = 4\pi\rho,$$

which in view of the identity $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \Delta\mathbf{A}$ and $\nabla \cdot (\nabla\phi) = \Delta\phi$ gives an alternative (with respect to Eqs. (G.1)–(G.4)) form of Maxwell equations (besides Eqs. (G.5) and (G.6)), i.e.,

$$\nabla(\nabla \cdot \mathbf{A} + \frac{1}{c}\frac{\partial \phi}{\partial t}) - \Delta\mathbf{A} + \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = \frac{4\pi}{c}\mathbf{i}, \tag{G.8}$$

$$\Delta\phi + \frac{1}{c}\nabla \cdot \frac{\partial \mathbf{A}}{\partial t} = -4\pi\rho. \tag{G.9}$$

To characterize the electromagnetic field we may use either $\boldsymbol{\mathcal{E}}$ and $\mathbf{H}$ or the two potentials ϕ and $\mathbf{A}$.

Arbitrariness of the potentials ϕ and $\mathbf{A}$

Potentials ϕ and $\mathbf{A}$ *are not defined uniquely, i.e., many different potentials lead to the same intensities of the electric and magnetic fields*. If we made in ϕ and $\mathbf{A}$ the following modifications:

$$\phi' = \phi - \frac{1}{c}\frac{\partial f}{\partial t}, \tag{G.10}$$

---

[1]

$$\nabla \times \nabla f = \left[\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial y}\right), \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial x}\right) - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z}\right), \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right] = [0, 0, 0] = \mathbf{0}.$$

$$\mathbf{A}' = \mathbf{A} + \nabla f, \tag{G.11}$$

where f is an *arbitrary differentiable function* (of x, y, z, t), then ϕ' and $\mathbf{A}'$ would lead to *the same* $\mathcal{E}$ and $\mathbf{H}$:

$$\mathcal{E}' = -\nabla\phi' - \frac{1}{c}\frac{\partial \mathbf{A}'}{\partial t} = \left(-\nabla\phi + \frac{1}{c}\nabla\frac{\partial f}{\partial t}\right) - \frac{1}{c}\left(\frac{\partial \mathbf{A}}{\partial t} + \frac{\partial}{\partial t}(\nabla f)\right) = \mathcal{E},$$
$$\mathbf{H}' = \nabla \times \mathbf{A}' = \nabla \times \mathbf{A} + \nabla \times \nabla f = \mathbf{H}.$$

Choice of potentials A *and* ϕ *for a uniform magnetic field*

From the second Maxwell equation (G.6) it is seen that if the magnetic field $\mathbf{H}$ is time-independent, then we get the time-independent $\mathbf{A}$. Profiting from the nonuniqueness of $\mathbf{A}$ we choose it in such a way as to satisfy (the so-called *Coulombic gauge*)[2]

$$\nabla \cdot \mathbf{A} = 0, \tag{G.12}$$

which diminishes the arbitrariness, but does not remove it.

Let us take an example of an atom in a uniform magnetic field $\mathbf{H}$. Let us locate the origin of the coordinate system on the nucleus, the choice being quite natural for an atom, and let us construct the vector potential at position $\mathbf{r} = (x, y, z)$ as

$$\mathbf{A}(\mathbf{r}) = \frac{1}{2}[\mathbf{H} \times \mathbf{r}]. \tag{G.13}$$

As shown above this is not a unique choice, there is an infinity of them. All the choices are equivalent from the mathematical and physical point of view, they differ however by a peanut: the economy of computations. It appears that this choice of $\mathbf{A}$ is at least a logical one. The choice is also consistent with the Coulombic gauge, Eq. (G.12), because

$$\nabla \cdot \mathbf{A} = \frac{1}{2}\nabla \cdot [\mathbf{H} \times \mathbf{r}] = \frac{1}{2}\nabla \cdot [\mathbf{H} \times \mathbf{r}] = \frac{1}{2}\nabla \cdot \left[H_y z - yH_z, H_z x - zH_x, H_x y - xH_y\right] =$$
$$\frac{1}{2}\left[\frac{\partial}{\partial x}(H_y z - yH_z) + \frac{\partial}{\partial y}(H_z x - zH_x) + \frac{\partial}{\partial z}(H_x y - xH_y)\right] = 0,$$

and also with the Maxwell equations (Eq. (G.6)), because

$$\nabla \times \mathbf{A} = \frac{1}{2}\nabla \times [\mathbf{H} \times \mathbf{r}] = \frac{1}{2}\nabla \cdot [\mathbf{H} \times \mathbf{r}] = \frac{1}{2}\nabla \times \left[H_y z - yH_z, H_z x - zH_x, H_x y - xH_y\right] =$$

---

[2] The Coulombic gauge, even if only one of possibilities, is almost exclusively used in molecular physics. The word "*gauge*" comes from the railway technology of measuring different rail widths.

$$\frac{1}{2}\left[\begin{array}{c}\frac{\partial}{\partial y}(H_x y - xH_y) - \frac{\partial}{\partial z}(H_z x - zH_x), \frac{\partial}{\partial z}(H_y z - yH_z) - \frac{\partial}{\partial x}(H_x y - xH_y), \\ \frac{\partial}{\partial x}(H_z x - zH_x) - \frac{\partial}{\partial y}(H_y z - yH_z))\end{array}\right] = \mathbf{H}.$$

Thus, this is a correct choice.

Practical importance of this choice

An example of possible choices of **A** gives Fig. G.1.

If we shifted the vector potential origin far away from the physical system under consideration (Fig. G.1b), then the values of $|\mathbf{A}|$ on all the particles of the system would be giant; **A** would be practically uniform *within* the atom or molecule. If we calculated $\nabla \times \mathbf{A} = \mathbf{H}$ on a particle of the system, we would obtain almost $\mathbf{0}$, because $\nabla \times \mathbf{A}$ means differentiation of **A**, and for a uniform field this yields zero. Thus, we are going to study the system in a magnetic field, but the field disappeared. A very high accuracy would be needed in order to calculate correctly $\nabla \times \mathbf{A}$ as differences of two large numbers, which numerically is always a risky business due to cancellation of accuracies. *It is therefore seen that the numerical results do depend critically on the choice of the origin of* **A** (arbitrary from the point of view of mathematics and physics). *It is always better to have the origin inside the system.*

Vector potential causes the wave function to change phase

The Schrödinger equation for a particle of mass m and charge q reads as

$$-\frac{\hbar^2}{2m}\Delta\Psi(\mathbf{r}) + V\Psi = E\Psi(\mathbf{r}),$$

where $V = q\phi$ with ϕ standing for the scalar electric potential.

The probability density of finding the particle at a given position depends on $|\Psi|$ rather than Ψ itself. This means that the wave function could be harmlessly multiplied by a phase factor $\Psi'(\mathbf{r}) = \Psi(\mathbf{r})\exp[-\frac{iq}{\hbar c}\chi(\mathbf{r})]$, where $\chi(\mathbf{r})$ could be any (smooth[3]) function of the particle's position $\mathbf{r}$. Then, we have $|\Psi| = |\Psi'|$ at any $\mathbf{r}$. If $\Psi'(\mathbf{r})$ is as good as Ψ is, it would be nice if it kindly satisfied the Schrödinger equation as Ψ does, of course with the same eigenvalue, i.e.,

---

[3] See Fig. 2.6.

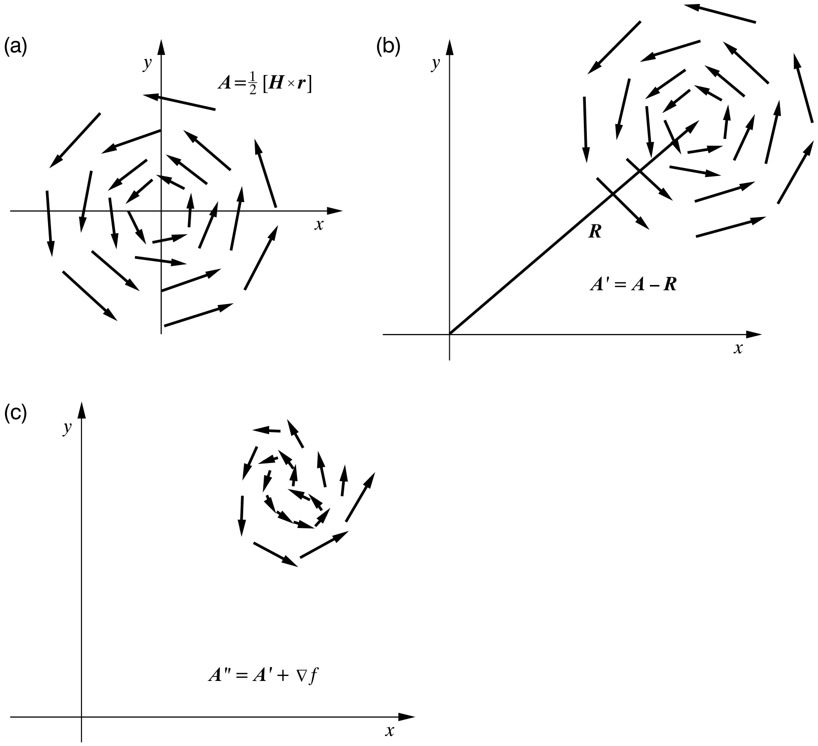


Fig. G.1. How to understand the arbitrariness of the vector potential $\mathbf{A}$? A schematic representation of three physically equivalent vector potentials $\mathbf{A}$. (a) A section by the plane $z = 0$ (the z axis protrudes towards the reader from the xy plane) of the vector field $\mathbf{A} = \frac{1}{2}(\mathbf{H} \times \mathbf{r})$ with $\mathbf{H} = (0, 0, H)$ and $H > 0$. It is seen that the vectors $\mathbf{A}$ become longer and longer when we are going out of the origin (where $\mathbf{A} = \mathbf{0}$); they "rotate" counterclockwise. Such $\mathbf{A}$ determines therefore $\mathbf{H}$ directed perpendicularly to the page and oriented towards the reader. (b) Note that any *shift* of the obtained potential should give the same magnetic field orthogonal to the drawing. This is what we get after adding, according to Eq. (G.11), the gradient of function $f = ax + by + c$ to potential $\mathbf{A}$, because $\mathbf{A} + \nabla f = \mathbf{A} + (\mathbf{i}a + \mathbf{j}b) = \mathbf{A} - \mathbf{R} = \mathbf{A}'$, where $\mathbf{R} = -(\mathbf{i}a + \mathbf{j}b) = \boldsymbol{const}$. The transformation is only one of possible transformations. If we took an arbitrary smooth function $f(x, y)$, e.g., with many maxima, minima, and saddle points (as in the mountains), we would deform (b) by expanding or shrinking it like a pancake. In this way we might obtain the situation shown in (c). The situations in (a), (b), and (c) are physically indistinguishable (if the scalar potential ϕ is changed appropriately).

$$-\frac{\hbar^2}{2m}\Delta\Psi'(\mathbf{r}) + V\Psi'(\mathbf{r}) = E\Psi'(\mathbf{r}).$$

Let us see how profound the consequences of this are. The left-hand side of the last equation can be transformed as follows:

$$
\begin{aligned}
& -\frac{\hbar^2}{2m}\Delta\Psi'(\mathbf{r}) + V\Psi'(\mathbf{r}) \\
&= -\frac{\hbar^2}{2m}\left[\exp(-\frac{iq}{\hbar c}\chi)\Delta\Psi + \Psi\Delta\exp(-\frac{iq}{\hbar c}\chi) + 2(\nabla\Psi)\left(\nabla\exp(-\frac{iq}{\hbar c}\chi)\right)\right] + \\
&\quad + V\exp(-\frac{iq}{\hbar c}\chi)\Psi \\
&= -\frac{\hbar^2}{2m}\left[\exp(-\frac{iq}{\hbar c}\chi)\Delta\Psi + \Psi\nabla[(-\frac{iq}{\hbar c})\exp(-\frac{iq}{\hbar c}\chi)\nabla\chi]\right. \\
&\quad \left. + 2(\nabla\Psi)[(-\frac{iq}{\hbar c})\exp(-\frac{iq}{\hbar c}\chi)\nabla\chi]\right] + +V\exp(-\frac{iq}{\hbar c}\chi)\Psi \\
&= -\frac{\hbar^2}{2m}\left[\exp(-\frac{iq}{\hbar c}\chi)\Delta\Psi + \Psi(-\frac{iq}{\hbar c})[(-\frac{iq}{\hbar c})\exp(-\frac{iq}{\hbar c}\chi)(\nabla\chi)^2 + \exp(-\frac{iq}{\hbar c}\chi)\Delta\chi]\right] + \\
&\quad - \frac{\hbar^2}{2m}2(\nabla\Psi)[(-\frac{iq}{\hbar c})\exp(-\frac{iq}{\hbar c}\chi)\nabla\chi] + V\exp(-\frac{iq}{\hbar c}\chi)\Psi.
\end{aligned}
$$

Dividing the Schrödinger equation by $\exp(-\frac{iq}{\hbar c}\chi)$ we obtain

$$
-\frac{\hbar^2}{2m}\left[\Delta\Psi + \Psi(-\frac{iq}{\hbar c})[(-\frac{iq}{\hbar c})(\nabla\chi)^2 + \Delta\chi] + 2(\nabla\Psi)[(-\frac{iq}{\hbar c})\nabla\chi]\right] + V\Psi = E\Psi(\mathbf{r}).
$$

Let us define a vector field $\mathbf{A}(\mathbf{r})$ by using the function $\chi(\mathbf{r})$,

$$
\mathbf{A}(\mathbf{r}) = \nabla\chi(\mathbf{r}). \tag{G.14}
$$

Hence, we have

$$
-\frac{\hbar^2}{2m}\left[\Delta\Psi + \Psi(-\frac{iq}{\hbar c})[(-\frac{iq}{\hbar c})\mathbf{A}^2 + \nabla\mathbf{A}] + 2(\nabla\Psi)[(-\frac{iq}{\hbar c})\mathbf{A}]\right] + V\Psi = E\Psi(\mathbf{r}),
$$

and introducing the momentum operator $\hat{\mathbf{p}} = -i\hbar\nabla$ we obtain

$$
\frac{1}{2m}\left[\hat{\mathbf{p}}^2\Psi + \Psi[(\frac{q}{c})^2\mathbf{A}^2 - (\frac{q}{c})\hat{\mathbf{p}}\mathbf{A}] - 2\left(\hat{\mathbf{p}}\Psi\right)(\frac{q}{c})\mathbf{A}\right] + V\Psi = E\Psi(\mathbf{r}),
$$

or finally

$$
\frac{1}{2m}\left(\hat{\mathbf{p}} - \frac{q}{c}\mathbf{A}\right)^2\Psi + V\Psi = E\Psi, \tag{G.15}
$$

which is the equation corresponding to the particle moving in an electromagnetic field with vector potential $\mathbf{A}$ (see p. V2-301).

Indeed, the last equation can be transformed in the following way:

$$\frac{1}{2m}[\hat{\mathbf{p}}^2\Psi + \left(\frac{q}{c}\right)^2 \mathbf{A}^2\Psi - \frac{q}{c}\hat{\mathbf{p}}(\mathbf{A}\Psi) - \frac{q}{c}\mathbf{A}\hat{\mathbf{p}}\Psi] + V\Psi = E\Psi,$$

which after using the equality[4] $\hat{\mathbf{p}}(\mathbf{A}\Psi) = \Psi\hat{\mathbf{p}}\mathbf{A} + \mathbf{A}\hat{\mathbf{p}}\Psi$ gives the expected result (Eq. (G.15)).

Therefore,

- from the fact that a wave function Ψ (with *a certain* phase) satisfies the Schrödinger equation,
- and requiring that a change of the phase coming from multiplication of Ψ by $\exp[-\frac{iq}{\hbar c}\chi(\mathbf{r})]$ does not destroy the satisfaction of the Schrödinger equation by $\Psi\exp[-\frac{iq}{\hbar c}\chi(\mathbf{r})]$ (in the text below $\chi(\mathbf{r})$ will be called the phase, although in reality the phase change is equal to $-\frac{q}{\hbar c}\chi(\mathbf{r})$),

one receives the Schrödinger equation for Ψ without this additional change of the phase, but with the modified momentum (and therefore also with the modified Hamiltonian) of the particle[5] by the so-called vector potential $\mathbf{A}(\mathbf{r}) : \hat{\mathbf{p}} \rightarrow (\hat{\mathbf{p}} - \frac{q}{c}\mathbf{A})$.

Hence, we have a link between the appearance of the phase (or rather its change χ) with the appearance of the vector potential $\mathbf{A}$ in the Schrödinger equation. This makes us believe that we got a more general form of the Schrödinger equation (this time with $\mathbf{A}$).

> A function with phase $\chi_1(\mathbf{r})$ corresponds to the vector potential $\mathbf{A}_1(\mathbf{r})$, while that with the phase $\chi_1(\mathbf{r}) + \chi(\mathbf{r})$ corresponds to $\mathbf{A}_2(\mathbf{r})$, where $\nabla\chi \equiv \mathbf{A}_2(\mathbf{r}) - \mathbf{A}_1(\mathbf{r})$. Therefore, these two vector potentials differ by the gradient of a function, and this is allowed (according to Eq. (G.11)) without any modification of physical phenomena.

Both functions, that with phase χ_1 as well as that with phase $\chi_1(\mathbf{r}) + \chi(\mathbf{r})$, give, of course, the same probability density.[6] Therefore, we are free to change $\mathbf{A}_1(\mathbf{r})$ to $\mathbf{A}_2(\mathbf{r})$, provided we compensate the new choice by changing the phase χ according to $\nabla\chi \equiv \mathbf{A}_2(\mathbf{r}) - \mathbf{A}_1(\mathbf{r})$.

Please note that the physically equivalent vector potentials may contain also a vector field component $\mathbf{A}$ (the same in both cases), *which is not a gradient (of any function)*. For example, they may contain a vortex-like field, which is not equivalent to any gradient field. The curl of such a field is nonzero, while the curl of a gradient of any function does equal zero. Thus, the $\mathbf{A}$ itself may contain an unknown admixture of the gradient of a function. Hence, any experimental

---

[4] Remember that $\hat{\mathbf{p}}$ is proportional to the first derivative operator.

[5] Please note that if we put $c = \infty$ (nonrelativistic approximation), no modification occurs.

[6] For any real χ_1 and χ, $\left|\exp[-\frac{iq}{\hbar c}\chi(\mathbf{r})]\right| = 1$ and $\left|\exp[-\frac{iq}{\hbar c}(\chi_1(\mathbf{r}) + \chi(\mathbf{r}))]\right| = 1$.

observation is determined solely by the nongradient component of the field. For example, for the magnetic field $\mathbf{H}_2$ for the vector potential $\mathbf{A}_2$,

$$\mathbf{H}_2 = \mathrm{curl}\mathbf{A}_2 \equiv \nabla \times \mathbf{A}_2 = \nabla \times (\mathbf{A}_1 + \nabla\chi) =$$
$$\nabla \times \mathbf{A}_1 + \nabla \times \nabla\chi = \nabla \times \mathbf{A}_1 + 0 = \mathrm{curl}\mathbf{A}_1 = \mathbf{H}_1,$$

both vector potentials ($\mathbf{A}_1$ and $\mathbf{A}_2$) correspond to the same physical situation, because $\mathbf{H}_1 = \mathbf{H}_2$. This result is obtained because $\mathrm{curl}(\mathrm{grad}\chi) \equiv \nabla \times \nabla\chi = 0$.

If a particle moves in a vector potential field **A** from $\mathbf{r}_0$ to **r**, then its wave function changes the phase by δ,

$$\delta = -\frac{q}{\hbar c}\int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{A}(\mathbf{r})\,\mathrm{d}\mathbf{r},$$

or, putting it in a different way, if the wave function underwent a phase change, then it would mean the particle moved in a vector potential of an electromagnetic field.

Incredible Aharonov–Bohm effect

In a small domain (say, in the center of the Brussels market place, where we like to locate the origin of the coordinate system) there is a magnetic field flux corresponding to the field intensity **H** directed along the z axis (perpendicular to the market place surface). Now, let us imagine a particle of the electric charge q enclosed in a three-dimensional box (say, a cube) of small dimensions located at a *very large distance* from the origin, and therefore from the magnetic flux, say, in Lisbon. Therefore, the *magnetic field in the box is equal to zero*. Now, we decide to make a travel with the box: Lisbon, Cairo, Ankara, St. Petersburg, Stockholm, Paris, and back to Lisbon. Did the wave function of the particle in the box change during the journey?

Let us see. The magnetic field **H** is related to the vector potential **A** through the relation $\nabla \times \mathbf{A} = \mathbf{H}$. This means that the particle was all the time subject to a huge vector potential field (see Fig. G.1), although the magnetic field was practically zero. Since the box is back to Lisbon, the phase acquired by the particle in the box[7] is an integral over the closed trajectory (loop)

$$\delta = -\frac{q}{\hbar c}\oint \mathbf{A}(\mathbf{r})\,\mathrm{d}\mathbf{r}.$$

---

[7] A nonzero δ requires a more general **A** than that satisfying Eq. (G.14).

However, from the Stokes equation we can replace the integral by an integral over surface enclosed by the loop

$$\int\int \mathbf{n}\cdot \mathrm{curl}\mathbf{A}\,\mathrm{d}S,$$

where $\mathbf{n}$ is a unit vector normal to the surface and $\mathrm{d}S$ is an infinitesimal element of the surface that is enclosed in the contour. We can also write this as

$$\delta = -\frac{q}{\hbar c}\int\int \mathbf{n}\cdot \mathbf{H}\,\mathrm{d}S = -\frac{q}{\hbar c}\Phi,$$

where Φ represents the *magnetic flux* (of the magnetic field $\mathbf{H}$) intersecting the loop surface that contains in particular the famous market place of Brussels. Thus, despite the fact that the particle could not feel the magnetic field $\mathbf{H}$ (because it was zero in the box), its wave function underwent a change of phase, which is detectable experimentally (in interference experiments).

Does the pair of potentials $\mathbf{A}$ and ϕ contain the same information as $\boldsymbol{\mathcal{E}}$ and $\mathbf{H}$ do? The Aharonov–Bohm effect (see also p. V2-458) suggests that $\mathbf{A}$ and ϕ are more important!

APPENDIX H

Optimal Wave Function for the Hydrogen-Like Atom

In several instances we encounter the problem of the mean value of the Hamiltonian for the hydrogen-like atom (atomic units are used throughout),

$$\hat{H} = -\frac{1}{2}\Delta - \frac{Z}{r}$$

with the normalized function

$$\Phi\left(r, \theta, \phi; c\right) = \sqrt{\frac{c^3}{\pi}} \exp\left(-cr\right),$$

where r, θ, ϕ are the spherical coordinates of the electron (the position of the nucleus of charge Z is fixed in the origin), $c > 0$ is a parameter to be found.

Calculation of the mean value of the Hamiltonian, i.e., the mean value of the energy

$$\varepsilon\left(\Phi\right) = \left\langle \Phi \left| \hat{H} \right| \Phi \right\rangle,$$

requires calculation of the mean value of the kinetic energy,

$$\bar{T} = \left\langle \Phi \left| -\frac{1}{2}\Delta \right| \Phi \right\rangle,$$

and the mean value of the potential energy (Coulombic attraction of the electron by the nucleus),

$$\bar{V} = -Z\left\langle \Phi \left| \frac{1}{r} \right| \Phi \right\rangle.$$

Therefore,

$$\varepsilon = \bar{T} + \bar{V}.$$

First, the Laplacian $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ may be expressed in the spherical coordinates

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

and in view of the fact that Φ is spherically symmetric (it depends on r only) we have

$$\left\langle \Phi \left| -\frac{1}{2}\Delta \right| \Phi \right\rangle = -\frac{1}{2}\left\langle \Phi \left| \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r} \right| \Phi \right\rangle =$$
$$-\frac{1}{2}\frac{c^3}{\pi}(-c)\left[\int_0^\infty r^2 \left[\frac{2}{r} - c\right] \exp(-2cr)\, dr \int_0^\pi \sin\theta d\theta \int_0^{2\pi} d\phi\right] =$$
$$\frac{1}{2}c^4 4\left[\int_0^\infty \left[2r - cr^2\right] \exp(-2cr)\, dr\right] =$$
$$2c^4\left[2\int_0^\infty r\exp(-2cr)\, dr - c\int_0^\infty r^2 \exp(-2cr)\, dr\right] =$$
$$4c^4(2c)^{-2} - 2c^5 2(2c)^{-3} = c^2 - \frac{1}{2}c^2 = \frac{1}{2}c^2,$$

where we have used that (this formula is often exploited throughout the book)

$$\int_0^\infty r^n \exp(-\beta r)\, dr = n!\beta^{-(n+1)}. \tag{H.2}$$

Quite similarly the second integral gives

$$-Z\left\langle \Phi \left| \frac{1}{r} \right| \Phi \right\rangle = -Z\frac{c^3}{\pi}\left[\int_0^\infty r\exp(-2cr)\, dr \int_0^\pi \sin\theta d\theta \int_0^{2\pi} d\phi\right] = -4Zc^3(2c)^{-2} = -Zc.$$

Therefore, finally,

$$\varepsilon = \frac{1}{2}c^2 - Zc. \tag{H.3}$$

We may want to use the variational method of finding the ground-state wave function. In this method we minimize the mean value of the Hamiltonian with respect to parameters in the variational function Φ. Hence, minimizing ε we force $\frac{\partial \varepsilon}{\partial c} = 0$, and therefore $c_{opt} = Z$. Note that in this particular case:

- Such value of c makes from the variational function the *exact* ground state of the hydrogen-like atom.
- The ground-state energy computed with $c_{opt} = Z$ gives $\varepsilon = \frac{1}{2}Z^2 - ZZ = -\frac{1}{2}Z^2$, which is the *exact* ground-state energy.
- The quantity $-\frac{\bar{V}}{\bar{T}} = \frac{Zc}{\frac{1}{2}c^2} = 2\frac{Z}{c}$. Only for the optimal value of $c = c_{opt} = Z$ we have what is called the *virial theorem* satisfied (see Chapter 5),

$$-\frac{\bar{V}}{\bar{T}} = 2. \tag{H.4}$$

APPENDIX I

The Virial Theorem

I.1 In classical mechanics

We consider N point-like particles of the time-independent masses m_i and radius vectors $\mathbf{r}_i$ in a chosen Cartesian coordinate system. Let us define two scalar quantities. The first one is the moment of inertia

$$I = \sum_{i=1}^{N} m_i r_i^2 \tag{I.1}$$

and the second quantity is defined by

$$G = \sum_{i=1}^{N} (\mathbf{r}_i \cdot \mathbf{p}_i) \, . \tag{I.2}$$

The quantities I and G are related by the following formula (t denotes time)[1]:

$$\frac{1}{2}\frac{\mathrm{d}I}{\mathrm{d}t} = G. \tag{I.3}$$

From Eq. (I.2) the time derivative of G is equal to[2]

$$\frac{\mathrm{d}G}{\mathrm{d}t} = 2T + \sum\nolimits_{i=1}^{N} (\mathbf{r}_i \cdot \mathbf{F}_i) \, , \tag{I.4}$$

where T stands for the total kinetic energy, $\mathbf{F}_i$ denotes the force acting on particle i, and the last term represents what is known as the virial ($\Sigma_i \, (\mathbf{r}_i \cdot \mathbf{F}_i)$).

---

[1] Indeed, $\frac{1}{2}\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\sum\nolimits_{i=1}^{N} m_i \mathbf{r}_i \cdot \mathbf{r}_i = \sum\nolimits_{i=1}^{N} m_i \mathbf{r}_i \cdot \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} = \sum\nolimits_{i=1}^{N} m_i \mathbf{r}_i \cdot \mathbf{p}_i = G$. Let us examine what happens to $I = \sum\nolimits_{i=1}^{N} m_i r_i^2$, when one shifts the origin by a fixed vector $\mathbf{R}$. We have $I' = \sum\nolimits_{i=1}^{N} m_i \, |\mathbf{r}_i - \mathbf{R}|^2 = \sum\nolimits_{i=1}^{N} m_i (r_i^2 + R^2 - 2\mathbf{r}_i \cdot \mathbf{R}) = I + R^2 M - 2\sum\nolimits_{i=1}^{N} m_i (\mathbf{r}_i \cdot \mathbf{R}) = I + R^2 M - 2M\mathbf{R} \cdot \mathbf{R}_{CM}$, where M denotes the total mass and $\mathbf{R}_{CM} = \frac{1}{M}\sum\nolimits_{i=1}^{N} m_i \mathbf{r}_i$ stands for the vector of the center of mass. What about G? $G' = G - \sum\nolimits_{i=1}^{N} (\mathbf{R} \cdot \mathbf{p}_i) = G - \mathbf{R} \cdot \mathbf{P}$, where $\mathbf{P} = \sum\nolimits_{i=1}^{N} \mathbf{p}_i$ stands for the total momentum. Therefore, since $\frac{1}{2}\frac{\mathrm{d}I'}{\mathrm{d}t} = G' + \mathbf{R} \cdot \mathbf{P} - M\frac{\mathrm{d}\mathbf{R}_{CM}}{\mathrm{d}t}\mathbf{R} = G'$, Eq. (I.3) is invariant with respect to any shift of the origin.

[2] We have $\frac{\mathrm{d}G}{\mathrm{d}t} = \Sigma_i (\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} \cdot p_i + \mathbf{r}_i \cdot \mathbf{F}_i) = \Sigma_i (m_i (\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t})^2 + \mathbf{r}_i \cdot \mathbf{F}_i) = 2T + \sum\nolimits_{i=1}^{N} (\mathbf{r}_i \cdot \mathbf{F}_i)$.

For pairwise interactions

Now, let us focus on the forces $\mathbf{F}_i$. We will limit ourselves to the forces that result from the time-independent potential energy $V(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N)$ through the relation

$$\mathbf{F}_i = -\nabla_i V. \tag{I.5}$$

If, in addition, they are pairwise additive and depend on the interparticle distances r_{ik} only, since V is a sum of pairwise particle–particle contributions, $V(\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N) = \sum_{i,j\,(i<k)}^{N} \mathcal{V}(r_{ij})$, we have

$$\mathbf{F}_i = \sum_{j(\neq i)}^{N} \mathbf{f}_{ij}(r_{ij}), \tag{I.6}$$

where $\mathbf{f}_{ij}$,

$$\mathbf{f}_{ij}(r_{ij}) = \frac{\mathrm{d}V}{\mathrm{d}r_{ij}} \frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}} = \frac{\mathrm{d}\mathcal{V}(r_{ij})}{\mathrm{d}r_{ij}} \frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}}, \tag{I.7}$$

represents the central force exerted on particle i by particle j (its direction shows the unit vector $\frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}}$) and the force depends uniquely on the interparticle distance $r_{ij} = \left|\mathbf{r}_j - \mathbf{r}_i\right|$.

For the pairwise interactions the virial from Eq. (I.4) takes the form[3]

$$\sum_{i=1}^{N} (\mathbf{r}_i \cdot \mathbf{F}_i) = -\sum_{i<j}^{N} r_{ij} \frac{\mathrm{d}\mathcal{V}(r_{ij})}{\mathrm{d}r_{ij}}.$$

Therefore, we have

$$\frac{\mathrm{d}G}{\mathrm{d}t} = 2T - \sum_{i<j}^{N} r_{ij} \frac{\mathrm{d}\mathcal{V}(r_{ij})}{\mathrm{d}r_{ij}}. \tag{I.8}$$

It often happens in physics that we use an approximation for $\mathcal{V}(r)$ in the form of power law

$$\mathcal{V}(r) = Ar^n, \tag{I.9}$$

---

[3] $\sum_{i=1}^{N} (\mathbf{r}_i \cdot \mathbf{F}_i) = \sum_{i=1}^{N} \sum_{j(j\neq i)}^{N} \left(\mathbf{r}_i \cdot \mathbf{f}_{ij}(r_{ij})\right) = \sum_{i=1}^{N} \sum_{j(j\neq i)}^{N} \frac{\mathrm{d}\mathcal{V}_{ij}}{\mathrm{d}r_{ij}} \left(\mathbf{r}_i \cdot \frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}}\right) =$ $\frac{1}{2}\left[\sum_{i=1}^{N} \sum_{j(j\neq i)}^{N} \frac{\mathrm{d}\mathcal{V}_{ij}}{\mathrm{d}r_{ij}} \left(\mathbf{r}_i \cdot \frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}}\right) - \sum_{i=1}^{N} \sum_{j(j\neq i)}^{N} \frac{\mathrm{d}\mathcal{V}_{ij}}{\mathrm{d}r_{ij}} \left(\mathbf{r}_j \cdot \frac{\mathbf{r}_j - \mathbf{r}_i}{r_{ij}}\right)\right] =$ $-\frac{1}{2} \sum_{i=1}^{N} \sum_{j(j\neq i)}^{N} \frac{\mathrm{d}\mathcal{V}_{ij}}{\mathrm{d}r_{ij}} \left(\frac{r_{ij}^2}{r_{ij}}\right) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j(j\neq i)}^{N} r_{ij} \frac{\mathrm{d}\mathcal{V}_{ij}}{\mathrm{d}r_{ij}} = -\sum_{i=1}^{N} \sum_{j(j\neq i)}^{N} r_{ij} \frac{\mathrm{d}\mathcal{V}_{ij}}{\mathrm{d}r_{ij}}.$

where n is an integer, while A is a real constant. This is the case for the electrostatic ($n=-1$) or gravitational ($n=-1$) interactions,[4] also for the harmonic interaction[5] ($n=2$). After Eq. (I.9) is inserted in Eq. (I.8) one gets

$$\frac{\mathrm{d}G}{\mathrm{d}t}=2T-An\sum_{i<j}^{N} r_{ij}r_{ij}^{n-1}=2T-n\sum_{i<j}^{N}\mathcal{V}(r_{ij})=2T-nV. \tag{I.10}$$

Time-averaging

Eq. (I.4) means that the sum of $2T$ and the virial $\sum_{i=1}^{N}(\mathbf{r}_i\cdot\mathbf{F}_i)$ is controlled by $\frac{\mathrm{d}G}{\mathrm{d}t}$, which is a function of time (this result has been obtained first by Lagrange). The presence of $\frac{\mathrm{d}G}{\mathrm{d}t}$ in Eq. (I.4) suggests that making a time average $\left\langle\frac{\mathrm{d}G}{\mathrm{d}t}\right\rangle_\tau$ of $\frac{\mathrm{d}G}{\mathrm{d}t}$ over a period of $t\in(0,\tau)$ one can get something interesting, since $\left\langle\frac{\mathrm{d}G}{\mathrm{d}t}\right\rangle_\tau=\frac{1}{\tau}\int_0^\tau\frac{\mathrm{d}G}{\mathrm{d}t}\mathrm{d}t=\frac{G(\tau)-G(0)}{\tau}$. This latter result means that if for whatever reason $G(\tau)-G(0)$ remains finite, for sufficiently large τ, we may use

$$2\langle T\rangle_\tau+\left\langle\sum_{i=1}^{N}(\mathbf{r}_i\cdot\mathbf{F}_i)\right\rangle_\tau=0.$$

A finite $G(\tau)-G(0)$ may happen, e.g., for a periodic motion. Recalling Eq. (I.2), $G=\sum_{i=1}^{N}(\mathbf{r}_i\cdot\mathbf{p}_i)$, a danger of infinite G happens for nonbound states with some particles escaping to infinity ($\mathbf{r}_i$ and $\mathbf{p}_i$ becoming colinear and $|\mathbf{r}_i\cdot\mathbf{p}_i|$ increasing in time). For the systems that keep together the virial theorem $2\langle T\rangle_\tau+\left\langle\sum_{i=1}^{N}(\mathbf{r}_i\cdot\mathbf{F}_i)\right\rangle_\tau=0$ holds. For the power law potentials we have from Eq. (I.10)

$$\boxed{\langle T\rangle_\tau=\frac{n}{2}\langle V\rangle_\tau\,.} \tag{I.11}$$

Example (Free particle (the virial theorem gives a trivial result)). A free particle of mass m moves along the x axis, its momentum is p_x. From Eqs. (I.1) and (I.2), $I=mx^2$ and $G=xp_x$ and we check that Eq. (I.3), $\frac{1}{2}\frac{\mathrm{d}I}{\mathrm{d}t}=G$, is satisfied. Then, since no forces act on the particle we have from Eq. (I.4) $\frac{\mathrm{d}G}{\mathrm{d}t}=2T=$ const. Thus, the virial theorem says something banal: the kinetic energy remains constant.

---

[4] This is in fact an approximation valid for point-like particles.

[5] Zero anharmonicity included.

Example (Classical oscillator (the virial theorem satisfied)). We have a one-dimensional harmonic oscillator (of mass m) with the force constant equal to k. The particle's position has the form $x = A\cos\omega t$, where $\omega = \sqrt{\frac{k}{m}}$ and the corresponding potential energy $V = \frac{1}{2}kx^2 = \frac{1}{2}kA^2(\cos\omega t)^2$, which means $n = 2$. The kinetic energy $T = \frac{mv^2}{2} = \frac{m(-A\omega\sin\omega t)^2}{2} = \frac{kA^2(\sin\omega t)^2}{2}$. Now, $\langle T\rangle_\tau = \frac{kA^2}{2}\left\langle \sin\omega t^2\right\rangle_\tau$. The virial theorem says that $2\langle T\rangle_\tau + \left\langle \sum_{i=1}^N (\mathbf{r}_i\cdot\mathbf{F}_i)\right\rangle_\tau = 0$, because $\left\langle \frac{\mathrm{d}G}{\mathrm{d}t}\right\rangle_\tau$ is equal to zero[6] as it has to be for a periodic motion. Let us check this. We have $2\langle T\rangle_\tau + \left\langle \sum_{i=1}^N (\mathbf{r}_i\cdot\mathbf{F}_i)\right\rangle_\tau = kA^2\left\langle(\sin\omega t)^2\right\rangle_\tau - kA^2\left\langle(\cos\omega t)^2\right\rangle_\tau = 0$, if τ is sufficiently large. Thus, Eq. (I.11) is satisfied and $\langle T\rangle_\tau = \langle V\rangle_\tau$.

I.2 In quantum mechanics

Let us recall Eq. (2.17), known also as the Ehrenfest theorem:

$$\frac{\mathrm{d}\left\langle \hat{A}\right\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle \left[\hat{H},\hat{A}\right]\right\rangle + \left\langle \frac{\partial}{\partial t}\hat{A}\right\rangle.$$

The equation tells us about the time dependence of the mean value of an arbitrary mechanical quantity A. Considering $\hat{A}$ that does not depend explicitly on time, i.e., $\frac{\partial}{\partial t}\hat{A} = 0$, we have

$$\frac{\mathrm{d}\left\langle \hat{A}\right\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle \left[\hat{H},\hat{A}\right]\right\rangle. \tag{I.12}$$

Let us take $A = p_x x$, where x is the x component of a particle's position vector $\mathbf{r}$, and the corresponding momentum component is p_x. We have from Eq. (I.12)

$$\frac{\mathrm{d}\left\langle \hat{p}_x x\right\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle \left[\hat{H},\hat{p}_x x\right]\right\rangle. \tag{I.13}$$

Let us transform a bit the commutator from the right-hand side. We obtain

---

[6] $\left\langle \frac{\mathrm{d}G}{\mathrm{d}t}\right\rangle_\tau = \left\langle \frac{\mathrm{d}}{\mathrm{d}t}xp_x\right\rangle_\tau = G(\tau) - G(0) = \frac{1}{\tau}(-mA^2\cos\omega\tau\,\sin\omega\tau - 0)$, which means that for very large τ we have $\left\langle \frac{\mathrm{d}G}{\mathrm{d}t}\right\rangle_\tau = 0$.

$$\left[\hat{H}, \hat{p}_x x\right] = \hat{H}\hat{p}_x x - \hat{p}_x x\hat{H} = \left(\hat{H}\hat{p}_x - \hat{p}_x\hat{H}\right)x - \hat{p}_x\left(x\hat{H} - \hat{H}x\right) =$$
$$\left[\hat{H}, \hat{p}_x\right]x - \hat{p}_x\left[x, \hat{H}\right].$$

Note that one has (f is an arbitrary function of class Q)[7] $\left[\hat{H}, \hat{p}_x\right]f = i\hbar\left(\frac{\partial V}{\partial x}\right)f$ and[8] $\left[x, \hat{H}\right]f = \frac{i\hbar}{m}\hat{p}_x f$.

Inserting the result into Eq. (I.13) one obtains[9]

$$\frac{\mathrm{d}\left\langle\hat{p}_x x\right\rangle}{\mathrm{d}t} = 2\left\langle\hat{T}\right\rangle - \left\langle x\frac{\partial V}{\partial x}\right\rangle. \tag{I.14}$$

When one takes N point-like particles and sums up all the equations of the above form one obtains

$$\frac{\mathrm{d}\left\langle\sum_{i=1}^{N}\mathbf{r}_i\cdot\hat{\mathbf{p}}_i\right\rangle}{\mathrm{d}t} = 2\left\langle\hat{T}\right\rangle + \left\langle\sum_{i=1}^{N}\left(\mathbf{r}_i\cdot\hat{\mathbf{F}}_i\right)\right\rangle, \tag{I.15}$$

where the operator of the force acting on the i-th particle is equal to $\hat{\mathbf{F}}_i = -\nabla_i V$ and is of pure electrostatic nature, because in nonrelativistic molecular quantum mechanics V is assumed to be uniquely of electrostatic character.

I.2.1 For stationary states

If the mean values calculated above pertain to *any stationary state*, one has $\frac{\mathrm{d}\left\langle\sum_i\mathbf{r}_i\cdot\mathbf{p}_i\right\rangle}{\mathrm{d}t} = 0$, because nothing changes in time in such a state. Thus, we obtain (for a stationary state)

$$2\left\langle\hat{T}\right\rangle + \left\langle\sum_{i=1}^{N}\left(\mathbf{r}_i\cdot\hat{\mathbf{F}}_i\right)\right\rangle = 0. \tag{I.16}$$

---

[7] Indeed, $\left[\hat{H}, \hat{p}_x\right]f = -i\hbar\left[\hat{T}+V, \frac{\partial}{\partial x}\right]f = -i\hbar\left[\hat{T}, \frac{\partial}{\partial x}\right]f - i\hbar\left[V, \frac{\partial}{\partial x}\right]f = 0 - i\hbar\left[V, \frac{\partial}{\partial x}\right]f = -i\hbar\left(V\frac{\partial f}{\partial x}\right) + i\hbar\frac{\partial}{\partial x}(Vf) = -i\hbar\left(V\frac{\partial f}{\partial x}\right) + i\hbar\left(\frac{\partial}{\partial x}V\right)f + i\hbar\left(V\frac{\partial f}{\partial x}\right) = i\hbar\left(\frac{\partial V}{\partial x}\right)f$.

[8] $\left[x, \hat{H}\right]f = \left[x, \hat{T}+V\right]f = \left[x, \hat{T}\right]f + [x, V]f = \left[x, \hat{T}\right]f + 0 = \left[x, \hat{T}\right]f = x\hat{T}f - \hat{T}(xf) = x\hat{T}f + \frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}(xf) = x\hat{T}f + \frac{\hbar^2}{2m}\frac{\partial}{\partial x}\left(f + x\frac{\partial f}{\partial x}\right) = x\hat{T}f + \frac{\hbar^2}{2m}\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial x} + x\frac{\partial^2 f}{\partial x^2}\right) = x\hat{T}f - x\hat{T}f + \frac{\hbar^2}{m}\frac{\partial f}{\partial x} = 0 + (-i)\,i\frac{\hbar^2}{m}\frac{\partial f}{\partial x} = \frac{i\hbar}{m}\hat{p}_x f$.

[9] $\frac{\mathrm{d}\left\langle\hat{p}_x x\right\rangle}{\mathrm{d}t} = \frac{i}{\hbar}\left\langle\left[\hat{H}, \hat{p}_x x\right]\right\rangle = \frac{i}{\hbar}\left\langle\left[\hat{H}, \hat{p}_x\right]x - \hat{p}_x\left[x, \hat{H}\right]\right\rangle = \frac{i}{\hbar}\left\langle i\hbar\left(\frac{\partial V}{\partial x}\right)x - \hat{p}_x\frac{i\hbar}{m}\hat{p}_x\right\rangle = -\left\langle\left(\frac{\partial V}{\partial x}\right)x\right\rangle$.

If the force operators acting on the point-like particles are pairwise additive and have the power form (as is the case for the Hamiltonian) of Eq. (I.9), one can use Eqs. (I.6)–(I.9) to find finally the quantum virial theorem for stationary solutions of the Schrödinger equation with the potential energy V:

$$\boxed{\frac{2\left\langle \hat{T} \right\rangle}{\langle V \rangle} = n. \qquad \text{(I.17)}}$$

APPENDIX J

Space- and Body-Fixed Coordinate Systems

Space-fixed coordinate system (SFCS)

A planetoid (or molecule) moves through empty Space, and we are observing it from our (inertial[1]) space ship. In order to carry out observations of the planetoid (molecule), we have to install some gear on our space ship and to fix on it a Cartesian coordinate system. It will enable us to describe the planetoid whatever will happen with it. This is the space-fixed coordinate system (SFCS); its orientation with respect to distant stars does not change in time.

If the molecule does not interact with anything, then with respect to the SFCS (see Chapter 2)

- its *total energy* remains invariant (because of the homogeneity of time);
- its *total momentum* remains invariant (because of the homogeneity of space);
- its *total angular momentum vector* remains invariant (because of the isotropy of space).

An observer on another space ship (inertial as well) will see the same phenomena in exactly the same way,[2] the energy, momentum, and angular momentum will be also invariant, but in general will be different from what was measured in the first space ship.

Let us introduce in the SFCS the vectors $\mathbf{r}_i = (x_i, y_i, z_i)$, showing from the origin the particles from which our molecule is composed (i.e., the electrons and the nuclei), $i = 1, 2, ...N$. Then, using the SFCS we write down the Hamiltonian of the system, the operators of the mechanical quantities we are interested in, we calculate all the wave functions we need, we compare with spectra measured in the SFCS, etc.

Body-fixed coordinate system (BFCS)

One day, however, we may feel that we do not like the SFCS, because of too many variables. Of course, this is not a sin, but only wasting our forces. Indeed, since in all inertial systems

---

[1] No rotation. We will convince ourselves that our SFCS is inertial by measuring how a point-like mass (assumed to be noninteracting with the rest of the space ship) moves. If it moves along a straight line with a constant velocity, then the SFCS is inertial. In a *noninertial* coordinate system the description of the physical phenomena in the molecule will be different.

[2] In the nonrelativistic approximation. The Doppler effect, the change of the electromagnetic wave frequency due to the motion (even uniform) of the emitting object, *is seen* in the experiment. The effect is of the relativistic character, i.e., vanishes, if we assume infinite velocity of light.

we have the same physics, we can separate the motion of the center of mass[3] (the total mass $M = \sum_i m_i$). The center of mass with position

$$\mathbf{R}_{CM} = \frac{\sum_i m_i \mathbf{r}_i}{M}$$

moves in the SFCS with a constant velocity along a straight line, which can be easily taken into account after the solution is obtained, and in most cases is irrelevant. This is why we decide to introduce the Cartesian coordinates $(X_{CM}, Y_{CM}, Z_{CM}) = \mathbf{R}_{CM}$ in the hope that in the future we will be able to get rid of them. Now, we need to introduce a coordinate system (of the lacking $(3N-3)$ variables) located on the molecule, the so called *body-fixed coordinate system* (BFCS). How to define this? Well, it should be a coordinate system that will define unambiguously any configuration of the particles in the molecule. There are a lot of such coordinate systems. Here we present some of the possibilities for the BFCS (*in all of them their axes are parallel to the corresponding axes of the SFCS*[4]). We may choose one of the following sets[5] of position vectors:

- $\mathbf{R}_{CM}$: then we locate the BFCS *on any of the particles* (say, this indicated by vector $\mathbf{r}_1$), and the BFCS positions of the other particles are shown by $\mathbf{r}'_i = \mathbf{r}_i - \mathbf{r}_1$ for $i = 2, 3, ...N$.
- $\mathbf{R}_{CM}$: the vector $\mathbf{R} = \mathbf{r}_2 - \mathbf{r}_1$ *indicates particle* 2 *from particle* 1, and the remaining particles are shown by the vectors that begin in the center of the section linking particles 1 and 2: $\mathbf{r}'_i = \mathbf{r}_i - \frac{(\mathbf{r}_1+\mathbf{r}_2)}{2}$ for $i = 3, 4, ...N$.
- $\mathbf{R}_{CM}$: all the vectors showing the particles (except particle 1): $\mathbf{r}'_i = \mathbf{r}_i - \mathbf{R}_{CM}$ for $i = 2, 3, ...N$. The position vector of particle 1 can be calculated from the coordinates already given.

Center-of-mass separation

After writing the Hamiltonian $\hat{\mathcal{H}}$ in the SFCS and then introducing *any of the above choices* of the coordinate system, we obtain $\hat{\mathcal{H}} = \hat{H}_{CM} + \hat{H}$.

At any of the choices the operator $\hat{H}$ is identical, but the mathematical formula for $\hat{H}$ will be different, because different coordinates are used.

---

[3] The exact separation of the center-of-mass motion in the SFCS and (not shown in this appendix) the exact separation of rotation of the molecule have been shown for the first time in the paper R.T. Pack, J.O. Hirschfelder, *J. Chem. Phys.*, 49(1968)4009.

[4] Only after introducing the axes of the coordinate system associated to the particles, and not with the SFCS, separation of rotation is possible.

[5] There are other choices possible.

Thus, the total Hamiltonian in the SFCS is

$$\hat{\mathcal{H}} = \hat{H}_{CM}\left(X_{CM}, Y_{CM}, Z_{CM}\right) + \hat{H}\left(\mathbf{r}\right),$$

where $\mathbf{r}$ *symbolizes*[6] all the other variables. The key result is that the two operators on the right-hand side *do depend on different variables*.

The goal of the above changes of the coordinate system was to show that the Schrödinger equation written in the SFCS, i.e., $\hat{\mathcal{H}}\Psi = \mathcal{E}\Psi$, splits into *two* Schrödinger equations ("separation of variables"):

- $\hat{H}_{CM}\psi_{CM} = E_{CM}\psi_{CM}$, describing the motion of a free "particle" of mass M and the coordinates X_{CM}, Y_{CM}, Z_{CM} (the "center-of-mass motion"), with $\psi_{CM} = \exp(i\mathbf{p}_{CM} \cdot \mathbf{R}_{CM})$, where $\mathbf{p}_{CM}$ stands for the total momentum of the system,
- $\hat{H}\psi = E\psi$,

where

$$\mathcal{E} = E + E_{CM},$$
$$\Psi(\mathbf{R}_{CM}, \mathbf{r}) = \psi_{CM}(\mathbf{R}_{CM}) \cdot \psi(\mathbf{r}).$$

The proof is simple. Let us check that the product wave function satisfies the Schrödinger equation. The left-hand side is

$$\begin{gathered}
\hat{\mathcal{H}}[\psi_{CM}(\mathbf{R}_{CM}) \cdot \psi(\mathbf{r})] = \\
\hat{H}_{CM}[\psi_{CM}(\mathbf{R}_{CM}) \cdot \psi(\mathbf{r})] + \hat{H}[\psi_{CM}(\mathbf{R}_{CM}) \cdot \psi(\mathbf{r})] = \\
\psi(\mathbf{r}) \cdot \hat{H}_{CM}\psi_{CM}(\mathbf{R}_{CM}) + \psi_{CM}(\mathbf{R}_{CM}) \cdot \hat{H}\psi(\mathbf{r}) = \\
\psi(\mathbf{r}) \cdot E_{CM}\psi_{CM}(\mathbf{R}_{CM}) + \psi_{CM}(\mathbf{R}_{CM}) \cdot E\psi(\mathbf{r}) = \\
(E + E_{CM})\left[\psi_{CM}(\mathbf{R}_{CM}) \cdot \psi(\mathbf{r})\right]
\end{gathered}$$

and this equals to the right-hand side, $\mathcal{E}\Psi$.

Example 1 (Center-of-mass separation for the first choice of the coordinates). We use the first choice of the coordinates for the system of *two* particles. In the SFCS $\hat{\mathcal{H}} = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 + V$. The new coordinates are

$$X_{CM} = \frac{\sum_i m_i x_i}{M}, \quad Y_{CM} = \frac{\sum_i m_i y_i}{M}, \quad Z_{CM} = \frac{\sum_i m_i z_i}{M},$$
$$x = x_2 - x_1,$$
$$y = y_2 - y_1,$$
$$z = z_2 - z_1.$$

---

[6] For the sake of brevity.

Then[7]

$$\frac{\partial}{\partial x_1} = \frac{\partial X_{CM}}{\partial x_1}\frac{\partial}{\partial X_{CM}} + \frac{\partial Y_{CM}}{\partial x_1}\frac{\partial}{\partial Y_{CM}} + \frac{\partial Z_{CM}}{\partial x_1}\frac{\partial}{\partial Z_{CM}} + \frac{\partial x}{\partial x_1}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_1}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_1}\frac{\partial}{\partial z} =$$

$$\frac{m_1}{M}\frac{\partial}{\partial X_{CM}} + 0 + 0 - \frac{\partial}{\partial x} + 0 + 0 = \frac{m_1}{M}\frac{\partial}{\partial X_{CM}} - \frac{\partial}{\partial x}$$

and similarly for y_1 and z_1. Further,

$$\frac{\partial}{\partial x_2} = \frac{\partial X_{CM}}{\partial x_2}\frac{\partial}{\partial X_{CM}} + \frac{\partial Y_{CM}}{\partial x_2}\frac{\partial}{\partial Y_{CM}} + \frac{\partial Z_{CM}}{\partial x_2}\frac{\partial}{\partial Z_{CM}} + \frac{\partial x}{\partial x_2}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_2}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_2}\frac{\partial}{\partial z} =$$

$$\frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + 0 + 0 + \frac{\partial}{\partial x} + 0 + 0 = \frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + \frac{\partial}{\partial x}$$

and similarly for y_2 and z_2.

Hence, the kinetic energy operator (after constructing the proper Laplacians from the above operators)

$$\hat{T} = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 =$$

$$-\frac{\hbar^2}{2m_1}\left[(\frac{m_1}{M})^2\frac{\partial^2}{\partial X_{CM}^2} + \frac{\partial^2}{\partial x^2} - 2\frac{m_1}{M}\frac{\partial^2}{\partial X_{CM}\partial x}\right] + (\text{similarly for } y \text{ and } z) +$$

$$-\frac{\hbar^2}{2m_2}\left[(\frac{m_2}{M})^2\frac{\partial^2}{\partial X_{CM}^2} + \frac{\partial^2}{\partial x^2} + 2\frac{m_2}{M}\frac{\partial^2}{\partial X_{CM}\partial x}\right] + (\text{similarly for } y \text{ and } z) =$$

$$-\frac{\hbar^2}{2M}\Delta_{CM} - \frac{\hbar^2}{2\mu}\Delta,$$

where for the reduced mass μ of the two particles $\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$, and $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$.

---

[7] According to the mathematical analysis we have to write down the contributions of all the differential operators $\frac{\partial}{\partial u}$ of the new coordinates u multiplied by their "coupling constants" $\frac{\partial u}{\partial x_1}$ with the coordinate x_1.

Our derivation is over, the operator $\hat{H}$ has been found. It turned out to be[8] (note that the new coordinates have to be introduced also in the potential energy V) of the form

$$\hat{H} = -\frac{\hbar^2}{2\mu}\Delta + V.$$

Example 2 (Center-of-mass separation for the third choice of the coordinates). Let us take again the same two particles, but this time use the third choice of the coordinate system. We have

$$X_{CM} = \frac{\sum_i m_i x_i}{M}, \qquad Y_{CM} = \frac{\sum_i m_i y_i}{M}, \qquad Z_{CM} = \frac{\sum_i m_i z_i}{M},$$
$$x = x_2 - X_{CM},$$
$$y = y_2 - Y_{CM},$$
$$z = z_2 - Y_{CM}.$$

Then

$$\frac{\partial}{\partial x_1} = \frac{\partial X_{CM}}{\partial x_1}\frac{\partial}{\partial X_{CM}} + \frac{\partial Y_{CM}}{\partial x_1}\frac{\partial}{\partial Y_{CM}} + \frac{\partial Z_{CM}}{\partial x_1}\frac{\partial}{\partial Z_{CM}} + \frac{\partial x}{\partial x_1}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_1}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_1}\frac{\partial}{\partial z} =$$
$$\frac{m_1}{M}\frac{\partial}{\partial X_{CM}} + 0 + 0 - \frac{m_1}{M}\frac{\partial}{\partial x} + 0 + 0 = \frac{m_1}{M}\left(\frac{\partial}{\partial X_{CM}} - \frac{\partial}{\partial x}\right)$$

and similarly for y_1 and z_1. Further,

$$\frac{\partial}{\partial x_2} = \frac{\partial X_{CM}}{\partial x_2}\frac{\partial}{\partial X_{CM}} + \frac{\partial Y_{CM}}{\partial x_2}\frac{\partial}{\partial Y_{CM}} + \frac{\partial Z_{CM}}{\partial x_2}\frac{\partial}{\partial Z_{CM}} + \frac{\partial x}{\partial x_2}\frac{\partial}{\partial x} + \frac{\partial y}{\partial x_2}\frac{\partial}{\partial y} + \frac{\partial z}{\partial x_2}\frac{\partial}{\partial z} =$$

---

[8] The kinetic energy operator has a quite interesting form. Particle 1 *rests* right in the origin of the BFCS ($x = 0, y = 0, z = 0$), and therefore its kinetic energy operator is absent in $\hat{H}$. There is the kinetic energy of particle 2, *but its mass is equal to* μ, not to m_2. The coordinates x, y, z (measured from the origin of the BFCS) correspond to particle 2. For example, for the hydrogen-like atom, if someone takes the nucleus as particle 1 and the electron as particle 2, then x, y, z show the *electron* from the Cartesian BFCS located on the nucleus. The potential energy operator $V = -\frac{Ze^2}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}} = -\frac{Ze^2}{\sqrt{x^2+y^2+z^2}}$ corresponds to the Coulombic interaction of the electron of charge $-e$ and the nucleus of charge Ze. After separation of the center of mass we are left with equation $\hat{H}\psi = E\psi$. The electron of mass μ is described by the wave function ψ. In the ground state $\psi = \frac{1}{\sqrt{\pi}}e^{-\sqrt{x^2+y^2+z^2}}$. This a description of the hydrogen-like atom *according to an observer sitting at the nucleus*.

If another observer puts his armchair (with the axes of the BFCS carved on it) at the *electron*, then he would see the hydrogen-like atom "according to the electron." Since in V there are squares of x, y, z, and in the kinetic energy operator there are the *second* derivatives with respect to x, y, z, we would obtain the same wave function as before, i.e., $\psi = \frac{1}{\sqrt{\pi}}e^{-\sqrt{x^2+y^2+z^2}}$, where the particle moving with respect to the electron is the *nucleus*, but with the mass equal to μ, i.e., the same as before. By the way, this μ is almost equal to the mass of the *electron*. *Thus, the two descriptions mean the same!*

$$\frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + 0 + 0 + \left(1 - \frac{m_2}{M}\right)\frac{\partial}{\partial x} + 0 + 0 = \frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + \left(1 - \frac{m_2}{M}\right)\frac{\partial}{\partial x} =$$
$$\frac{m_2}{M}\frac{\partial}{\partial X_{CM}} + \frac{m_1}{M}\frac{\partial}{\partial x}$$

and similarly for y_2 and z_2.

Thus, the kinetic energy operator takes the form (after inserting the squares of the corresponding operators)

$$\hat{T} = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 =$$
$$-\frac{\hbar^2}{2m_1}\left[(\frac{m_1}{M})^2\left(\frac{\partial^2}{\partial X_{CM}^2} + \frac{\partial^2}{\partial x^2} - 2\frac{\partial^2}{\partial X_{CM}\partial x}\right)\right] + (\text{similarly for } y \text{ and } z) +$$
$$-\frac{\hbar^2}{2m_2}\left[(\frac{m_2}{M})^2\frac{\partial^2}{\partial X_{CM}^2} + (\frac{m_1}{M})^2\frac{\partial^2}{\partial x^2} + 2\frac{m_1 m_2}{M^2}\frac{\partial^2}{\partial X_{CM}\partial x}\right] + (\text{similarly for } y \text{ and } z) =$$
$$-\frac{\hbar^2}{2M}\Delta_{CM} - \frac{\hbar^2}{2m_1}(\frac{m_1}{M})^2\Delta_{xyz} - \frac{\hbar^2}{2m_2}(\frac{m_1}{M})^2\Delta_{xyz} - \frac{\hbar^2}{2m_1}(\frac{m_1}{M})^2\left(-2\frac{\partial^2}{\partial X_{CM}\partial x}\right) + \ldots$$
$$-\frac{\hbar^2}{2m_2}2\frac{m_1 m_2}{M^2}\frac{\partial^2}{\partial X_{CM}\partial x} + \ldots = -\frac{\hbar^2}{2M}\Delta_{CM} - \frac{\hbar^2}{2m_1}(\frac{m_1}{M})^2\Delta_{xyz} - \frac{\hbar^2}{2m_2}(\frac{m_1}{M})^2\Delta_{xyz} =$$
$$-\frac{\hbar^2}{2M}\Delta_{CM} - \frac{\hbar^2}{2}(\frac{m_1}{m_2 M})\Delta_{xyz}.$$

It is seen that once again we have reached the situation allowing us to separate the motion of the center of mass in the Schrödinger equation. This time, however, the *form* of the operator $\hat{H}$ is different (e.g., Δ_{xyz} has only formally the same form as Δ), only because the variables are different (the operator remains *the same*). Once again this is the kinetic energy of a point-like particle[9] with the coordinates x, y, z (defined in *this* example) and the mass equal to $\frac{m_2 M}{m_1}$.

---

[9] Let us first denote the nucleus as particle 1 and the electron as particle 2. Then $\mathbf{R}_{CM}$ shows almost the position of the nucleus, and x, y, z are almost the coordinates of the electron measured from the nucleus, while $\frac{m_2 M}{m_1}$ is almost equal to the mass of the electron. Thus, we have a situation that resembles Example 1.
If the particles are chosen in the other way (the electron is particle 1 and the nucleus is particle 2), then the same physical situation looks completely different. The values of x, y, z are very close to 0, while the mass of the effective point-like particle becomes very large.
Note that the new coordinates describe the potential energy in a more complex way. We need the differences of the kind $x_2 - x_1$ in order to insert them into the Pythagoras formula for the distance. We have $x_1 = X_{CM}\frac{m_1+m_2}{m_1} - \frac{m_2}{m_1}x_2 = X_{CM}\frac{m_1+m_2}{m_1} - \frac{m_2}{m_1}(x + X_{CM}) = X_{CM} - \frac{m_2}{m_1}x$, $x_1 - x_2 = X_{CM} - \frac{m_2}{m_1}x - x - X_{CM} = -x\left(1 + \frac{m_2}{m_1}\right)$. This immediately gives (r stands for the distance between the electron and the center of mass) $V\,(\text{new}) = -\frac{Ze^2}{\left(1+\frac{m_2}{m_1}\right)r}$.

APPENDIX K

Orthogonalization

Schmidt orthogonalization

Two vectors

Imagine two vectors $\mathbf{u}$ and $\mathbf{v}$, each of length 1 (i.e., normalized), having the scalar product $\langle \mathbf{u}|\mathbf{v}\rangle = a$. If $a = 0$, then the two vectors are orthogonal. We are interested in the case $a \neq 0$. Can one make such linear combinations of $\mathbf{u}$ and $\mathbf{v}$ that the new vectors $\mathbf{u}'$ and $\mathbf{v}'$ will be orthogonal? We can do that in many ways; two of them are called Schmidt orthogonalization:

Case I.

$$\mathbf{u}' = \mathbf{u},$$

$$\mathbf{v}' = \mathbf{v} - \mathbf{u}\,\langle \mathbf{u}|\mathbf{v}\rangle.$$

Case II.

$$\mathbf{u}' = \mathbf{u} - \mathbf{v}\,\langle \mathbf{v}|\mathbf{u}\rangle,$$

$$\mathbf{v}' = \mathbf{v}.$$

It is seen that the Schmidt orthogonalization is based on a very simple idea. In Case I the first vector is left unchanged, while from the second vector one cuts out its component along the first one (Fig. K.1). In this way the two vectors are treated differently (hence, the two cases above).

In this book the vectors to orthogonalize will be the Hilbert space vectors (see Appendix B), i.e., the normalized wave functions. In the case of two such vectors ϕ_1 and ϕ_2 having the scalar product $\langle \phi_1|\phi_2\rangle$ we construct the new orthogonal wave functions $\psi_1 = \phi_1$, $\psi_2 = \phi_2 - \phi_1\,\langle \phi_1|\phi_2\rangle$ or $\psi_1 = \phi_1 - \phi_2\,\langle \phi_2|\phi_1\rangle$, $\psi_2 = \phi_2$ in analogy to the previous formulae.

More vectors

In case of many vectors the procedure is similar. First, we decide about the order of the vectors to be orthogonalized. Then, we begin the procedure by leaving the first vector unchanged. We continue, remembering that from a new vector we have to cut out all its components along the new vectors already found. Of course, the final set of vectors depends on the order chosen and are nonnormalized (the normalization can be done afterwards).

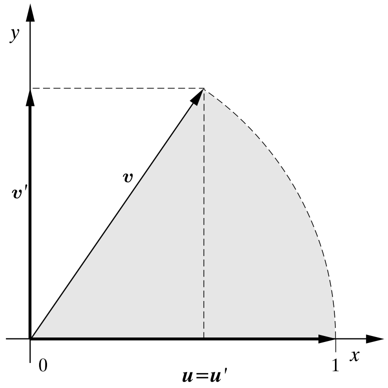


Fig. K.1. The Schmidt orthogonalization of the unit (i.e., normalized) vectors $\mathbf{u}$ and $\mathbf{v}$. The new vectors are $\mathbf{u}'$ and $\mathbf{v}'$. Vector $\mathbf{u}' \equiv \mathbf{u}$, while from vector $\mathbf{v}$ we subtract its component along $\mathbf{u}$. The new vectors are orthogonal, but not normalized.

Löwdin symmetric orthogonalization

Imagine the normalized but nonorthogonal basis set wave functions collected as the components of the vector $\boldsymbol{\phi}$. By making proper linear combinations of the wave functions we will get the orthogonal wave functions. The *symmetric* orthogonalization (as opposed to the Schmidt orthogonalization) treats all the wave functions on equal footing. Instead of the old nonorthogonal basis set $\boldsymbol{\phi}$ we construct a new basis set $\boldsymbol{\phi}'$ by a linear transformation $\boldsymbol{\phi}' = \mathbf{S}^{-\frac{1}{2}}\boldsymbol{\phi}$, where $\mathbf{S}$ is the overlap matrix with the elements $S_{ij} = \langle \phi_i | \phi_j \rangle$, and the square matrix $\mathbf{S}^{-\frac{1}{2}}$ and its cousin $\mathbf{S}^{\frac{1}{2}}$ are defined in the following way. First, we diagonalize $\mathbf{S}$ using a unitary matrix $\mathbf{U}$, i.e., $\mathbf{U}^\dagger\mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{1}$ (for real $\mathbf{S}$ the matrix $\mathbf{U}$ is orthogonal, $\mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{1}$),

$$\mathbf{S}_{diag} = \mathbf{U}^\dagger\mathbf{S}\mathbf{U}.$$

The eigenvalues of $\mathbf{S}$ are always positive; therefore, the diagonal elements of $\mathbf{S}_{diag}$ can be replaced by their square roots, thus producing the matrix denoted by *symbol* $\mathbf{S}^{\frac{1}{2}}_{diag}$. Using this matrix we define the matrices $\mathbf{S}^{\frac{1}{2}} = \mathbf{U}\mathbf{S}^{\frac{1}{2}}_{diag}\mathbf{U}^\dagger$ and $\mathbf{S}^{-\frac{1}{2}} = \left(\mathbf{S}^{\frac{1}{2}}\right)^{-1} = \mathbf{U}\mathbf{S}^{-\frac{1}{2}}_{diag}\mathbf{U}^\dagger$. Their *symbols* correspond to their properties: $\mathbf{S}^{\frac{1}{2}}\mathbf{S}^{\frac{1}{2}} = \mathbf{U}\mathbf{S}^{\frac{1}{2}}_{diag}\mathbf{U}^\dagger\mathbf{U}\mathbf{S}^{\frac{1}{2}}_{diag}\mathbf{U}^\dagger = \mathbf{U}\mathbf{S}^{\frac{1}{2}}_{diag}\mathbf{S}^{\frac{1}{2}}_{diag}\mathbf{U}^\dagger = \mathbf{U}\mathbf{S}_{diag}\mathbf{U}^\dagger = \mathbf{S}$, and similarly $\mathbf{S}^{-\frac{1}{2}}\mathbf{S}^{-\frac{1}{2}} = \mathbf{S}^{-1}$. Also, a straightforward calculation gives[1] $\mathbf{S}^{-\frac{1}{2}}\mathbf{S}^{\frac{1}{2}} = \mathbf{1}$.

---

[1] The matrix $\mathbf{S}^{-\frac{1}{2}}$ is not just a symbol anymore. Let us check whether the transformation $\boldsymbol{\phi}' = \mathbf{S}^{-\frac{1}{2}}\boldsymbol{\phi}$ gives indeed orthonormal wave functions (vectors). Remembering that $\boldsymbol{\phi}$ represents a vertical vector with the components ϕ_i

An important feature of the symmetric orthogonalization is[2] that among all possible orthogonalizations the symmetric orthogonalization ensures

$$\sum_i \left|\left|\phi_i - \phi_i'\right|\right|^2 = \min\text{imum},$$

where $\left|\left|\phi_i - \phi_i'\right|\right|^2 \equiv \left\langle \phi_i - \phi_i' | \phi_i - \phi_i' \right\rangle$. This means that

> the symmetrically orthogonalized functions ϕ_i' are the "least distant" from the original functions ϕ_i. Thus, a symmetric orthogonalization means the most gentle pushing the directions of the vectors in order to get them orthogonal.

Example. The symmetric orthogonalization will be shown taking an example of two nonorthogonal vectors $\mathbf{u}$ and $\mathbf{v}$ (instead of functions ϕ_1 and ϕ_2), each of length 1 with the scalar product[3] $\langle \mathbf{u}|\mathbf{v}\rangle = a \neq 0$. We decide to consider the vectors with real components, hence $a \in R$. First we have to construct matrix $\mathbf{S}^{-\frac{1}{2}}$. Here is how we arrive at this. Matrix $\mathbf{S}$ is equal to $\mathbf{S} = \begin{pmatrix} 1 & a \\ a & 1 \end{pmatrix}$, and as we see it is symmetric. First, let us diagonalize $\mathbf{S}$. To achieve that, we apply the orthogonal transformation $\mathbf{U}^\dagger\mathbf{S}\mathbf{U}$ (thus, in this case $\mathbf{U}^\dagger = \mathbf{U}^T$), where (to ensure the orthogonality of the transformation matrix) we choose $\mathbf{U} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, and therefore $\mathbf{U}^\dagger = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ with angle θ to be specified. After the transformation we have $\mathbf{U}^\dagger\mathbf{S}\mathbf{U} = \begin{pmatrix} 1 - a\sin 2\theta & a\cos 2\theta \\ a\cos 2\theta & 1 + a\sin 2\theta \end{pmatrix}$.

It is seen that if we chose $\theta = 45°$, then the matrix $\mathbf{U}^\dagger\mathbf{S}\mathbf{U}$ would be *diagonal*[4] (that is what we would like to have): $\mathbf{S}_{diag} = \begin{pmatrix} 1-a & 0 \\ 0 & 1+a \end{pmatrix}$. We construct then $\mathbf{S}_{diag}^{\frac{1}{2}} = \begin{pmatrix} \sqrt{1-a} & 0 \\ 0 & \sqrt{1+a} \end{pmatrix}$.

(being functions), $\int \boldsymbol{\phi}\boldsymbol{\phi}^T d\tau = \mathbf{S}$, while $\int \boldsymbol{\phi}'\boldsymbol{\phi}'^T d\tau = \int \mathbf{S}^{-\frac{1}{2}}\boldsymbol{\phi}\boldsymbol{\phi}^T\mathbf{S}^{-\frac{1}{2}} d\tau = \mathbf{S}^{-\frac{1}{2}}(\int \boldsymbol{\phi}\boldsymbol{\phi}^T d\tau)\mathbf{S}^{-\frac{1}{2}} = \mathbf{S}^{-\frac{1}{2}}\mathbf{S}\mathbf{S}^{-\frac{1}{2}} = \mathbf{1}$. This is what we wanted to show.

[2] G.W. Pratt, S.P. Neustadter, *Phys. Rev.*, 101(1956)1248.

[3] We have $-1 \le a \le 1$.

[4] In such a case the transformation matrix is

$$\mathbf{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Next, we form[5] $\mathbf{S}^{\frac{1}{2}} = \mathbf{U}\mathbf{S}_{diag}^{\frac{1}{2}}\mathbf{U}^{\dagger} = \frac{1}{2}\begin{pmatrix} \sqrt{1-a}+\sqrt{1+a} & \sqrt{1+a}-\sqrt{1-a} \\ \sqrt{1+a}-\sqrt{1-a} & \sqrt{1-a}+\sqrt{1+a} \end{pmatrix}$ and the matrix $\mathbf{S}^{-\frac{1}{2}}$ needed for the transformation is equal to

$$\mathbf{S}^{-\frac{1}{2}} = \mathbf{U}\mathbf{S}_{diag}^{-\frac{1}{2}}\mathbf{U}^{\dagger} = \mathbf{U}\begin{pmatrix} \frac{1}{\sqrt{1-a}} & 0 \\ 0 & \frac{1}{\sqrt{1+a}} \end{pmatrix}\mathbf{U}^{\dagger} = \frac{1}{2}\begin{pmatrix} \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} & \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} \\ \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} & \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} \end{pmatrix}.$$

Now we are ready to construct the orthogonalized vectors[6]:

$$\begin{pmatrix} \mathbf{u}' \\ \mathbf{v}' \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} & \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} \\ \frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}} & \frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}} \end{pmatrix}\begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix},$$

$$\mathbf{u}' = C\mathbf{u} + c\mathbf{v},$$

$$\mathbf{v}' = c\mathbf{u} + C\mathbf{v},$$

where the "large" coefficient $C = \frac{1}{2}\left(\frac{1}{\sqrt{1-a}}+\frac{1}{\sqrt{1+a}}\right)$, and a "small" admixture $c = \frac{1}{2}\left(\frac{1}{\sqrt{1+a}}-\frac{1}{\sqrt{1-a}}\right)$. As we can see the new (orthogonal) vectors are formed from the old ones (nonorthogonal) by an *identical* (hence the name "*symmetric orthogonalization*") admixture of the old vectors, i.e., the contribution of **u** and **v** in $\mathbf{u}'$ is the same as that of **v** and **u** in $\mathbf{v}'$.

The new vectors are obtained by correcting the directions of the old ones, each by the same angle.

---

[5] They are symmetric matrices. For example, $\left(\mathbf{S}^{\frac{1}{2}}\right)_{ij} = \left(\mathbf{U}\mathbf{S}_{diag}^{\frac{1}{2}}\mathbf{U}^{\dagger}\right)_{ij} = \sum_k \sum_l U_{ik}\left(\mathbf{S}_{diag}^{\frac{1}{2}}\right)_{kl} U_{jl} = \sum_k \sum_l U_{ik}\left(\mathbf{S}_{diag}^{\frac{1}{2}}\right)_{kl} \delta_{kl} U_{jl} = \sum_k U_{ik}\left(\mathbf{S}_{diag}^{\frac{1}{2}}\right)_{kk} U_{jk} = \left(\mathbf{S}^{\frac{1}{2}}\right)_{ji}$.

[6] It is seen that if the vectors **u** and **v** were already orthogonal, i.e., $a = 0$, then $\mathbf{u}' = \mathbf{u}$ and $\mathbf{v}' = \mathbf{v}$. Of course, we like this result.

This is illustrated in Fig. K.2.

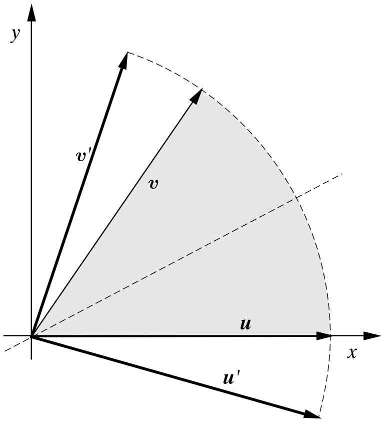


Fig. K.2. The symmetric (or Löwdin) orthogonalization of the normalized vectors $\mathbf{u}$ and $\mathbf{v}$. The vectors are just pushed off by the same angle in such a way as to ensure $\mathbf{u}'$ and $\mathbf{v}'$ become orthogonal.

The new vectors have automatically length 1, the same as the starting vectors.

APPENDIX L

Diagonalization of a Matrix

In quantum chemistry we often encounter the following mathematical problem.

We have a Hermitian[1] matrix $\mathbf{A}$ (of dimension n), i.e., $\mathbf{A}^\dagger = \mathbf{A}$, and we are interested in all numbers λ (called "eigenvalues"[2]) and the corresponding column vectors ("eigenvectors" of dimension n) $\mathbf{L}$ that satisfy the following equation:

$$(\mathbf{A} - \lambda\, \mathbf{1})\mathbf{L} = \mathbf{0}, \tag{L.1}$$

where $\mathbf{1}$ is the unit matrix (of dimension n). There are n solutions to the last equation: n eigenvalues of λ and also n eigenvectors $\mathbf{L}$. Some eigenvalues λ may be equal (degeneracy), i.e., two or more linearly independent eigenvectors $\mathbf{L}$ correspond to a single eigenvalue λ. From (L.1) it is shown that any vector $\mathbf{L}$ is determined only to the accuracy of a multiplicative factor.[3] This is why in the future it will be justified to normalize them to unity.

In quantum chemistry the eigenvalue problem is solved in two ways: one is easy for $n \le 2$, but more and more difficult for larger n, the second way (with computers) treats all cases uniformly.

- The first way sees the eigenvalue equation as a set of linear homogeneous equations for the unknown components of the vector $\mathbf{L}$. Then the condition for the nontrivial solution[4] to exist is $\det(\mathbf{A} - \lambda\, \mathbf{1}) = 0$. This condition can be fulfilled only for some particular values of λ, which are to be found by expanding the determinant and solving the resulting n-th-degree polynomial equation for λ. Then each solution λ is inserted into Eq. (L.1) and the components of the corresponding vector $\mathbf{L}$ are found by using any method applicable to linear equations. Thus, we end up with λ_k and $\mathbf{L}_k$ for $k = 1, 2, 3, \ldots n$.
- The second way is based on diagonalization of $\mathbf{A}$.

First, let us show that *the same* λ's satisfy the eigenvalue equation, Eq. (L.1), but with a much simpler matrix. To this end let us multiply (L.1) by an (at the moment) *arbitrary* nonsingular[5] square matrix[6] $\mathbf{B}$. We obtain the following chain of transformations: $\mathbf{B}^{-1}(\mathbf{A} \ - \lambda\, \mathbf{1})\ \mathbf{L} =$

---

[1] In practice matrix $\mathbf{A}$ is usually real, and therefore satisfies $\left(\mathbf{A}^T\right)^* = \mathbf{A}^T = \mathbf{A}$, i.e., $\mathbf{A}$ is symmetric.

[2] They are real.

[3] In other words an unnormalized wave function still satisfies the Schrödinger equation, or to any normal mode can be assigned an arbitrary amplitude.

[4] The trivial one is obviously $\mathbf{L} = \mathbf{0}$, which is however unacceptable, since the wave function cannot vanish everywhere, or atoms have to vibrate, etc.

[5] That is, its inverse matrix exists.

[6] To be found.

$\mathbf{B}^{-1}(\mathbf{ABB}^{-1} - \lambda\mathbf{1})\ \mathbf{L} = (\mathbf{B}^{-1}\mathbf{AB} - \lambda\ \mathbf{1})\ \mathbf{B}^{-1}\ \mathbf{L} = \left(\tilde{\mathbf{A}} - \lambda\ \mathbf{1}\right)\tilde{\mathbf{L}} = \mathbf{0}$, where[7] $\tilde{\mathbf{A}} = \mathbf{B}^{-1}\mathbf{AB}$ and $\tilde{\mathbf{L}} = \mathbf{B}^{-1}\mathbf{L}$. Thus, another matrix and other eigenvectors, but *the same* λ's! Now, let us choose such a special $\mathbf{B}$ that the resulting equation is as simple as possible, i.e., with a diagonal $\tilde{\mathbf{A}}$. Then, we will know[8] what the λ values have to be in order to satisfy the equation $\left(\tilde{\mathbf{A}} - \lambda\ \mathbf{1}\right)\tilde{\mathbf{L}} = \mathbf{0}$.

Indeed, if $\tilde{\mathbf{A}}$ were diagonal, then $\det(\tilde{\mathbf{A}} - \lambda\ \mathbf{1}) = \mathbf{\Pi}_{k=1}^{n}\left(\tilde{A}_{kk} - \lambda\right) = 0$, which gives the solutions $\lambda_k = \tilde{A}_{kk}$. Then, it is easy to find the corresponding vector $\tilde{\mathbf{L}}_k$. For example, $\tilde{\mathbf{L}}_1$ we find from equation $\left(\tilde{\mathbf{A}} - \lambda_1\ \mathbf{1}\right)\tilde{\mathbf{L}}_1 = \mathbf{0}$ in the following way[9]:

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & \tilde{A}_{22} - \lambda_1 & 0 & \dots & 0 \\ 0 & 0 & \tilde{A}_{33} - \lambda_1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \tilde{A}_{nn} - \lambda_1 \end{pmatrix} \begin{pmatrix} \tilde{L}_{1,1} \\ \tilde{L}_{1,2} \\ \tilde{L}_{1,3} \\ \dots \\ \tilde{L}_{1,n} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

which means that in order to get $\mathbf{0}$ on the right-hand side, one has to have an *arbitrary* $\tilde{L}_{1,1}$, while the other $\tilde{L}_{1,j} = 0$ for $j = 2, 3, \dots n$.

In order to have the length of $\tilde{\mathbf{L}}_1$ equal to 1, it is sufficient to put $\tilde{L}_{1,1} = 1$. Similarly, we find easily that the vectors $\tilde{\mathbf{L}}_k$ corresponding to λ_k represent nothing else but the column vectors with all components equal to 0 except the component k, which is equal to 1. We are interested in vectors $\mathbf{L}$, rather than $\tilde{\mathbf{L}}$. We get these vectors from $\mathbf{L} = \mathbf{B}\tilde{\mathbf{L}}$, and when taking into account the form of $\tilde{\mathbf{L}}$, this means that $\mathbf{L}_k$ is nothing else but the k-th column of matrix $\mathbf{B}$. Since $\mathbf{B}$ is known, because precisely this matrix led to the diagonalization, there is no problem with $\mathbf{L}$:

> the columns of $\mathbf{B}$ represent the eigenvectors $\mathbf{L}$ of the equation $(\mathbf{A} - \lambda\ \mathbf{1})\mathbf{L} = \mathbf{0}$.

Our task is over!

---

[7] Then, such a *unitary matrix* $\mathbf{B}$ (i.e., satisfying $\left(\mathbf{B}^T\right)^* = \mathbf{B}^{-1}$) can be found that $\mathbf{B}^{-1}\mathbf{AB}$ is *real and diagonal*. When (as is the case in most applications) we have to do with real matrices, instead of unitary and Hermitian matrices we have to do with orthogonal and symmetric matrices, respectively.

[8] Just by looking at it.

[9] The λ has been replaced by λ_1, because one is interested in getting $\tilde{\mathbf{L}}_1$.

Secular Equation $(\mathbf{H}-\varepsilon\mathbf{S})\,\mathbf{c}=\mathbf{0}$

A typical approach to solve an eigenvalue problem is its "algebraization," i.e., representation of the wave function as a linear combination of the known basis functions with the unknown coefficients. Then, instead of searching for a function, we try to find the expansion *coefficients* $\mathbf{c}$ from the secular equation[1] $(\mathbf{H}-\varepsilon\mathbf{S})\,\mathbf{c}=0$. Our goal is to reduce this task to the eigenvalue problem of a matrix. If the basis set used were orthonormal, then the goal would be immediately achieved, because the secular equation would be reduced to $(\mathbf{H}-\varepsilon\mathbf{1})\,\mathbf{c}=0$, i.e., the quite standard eigenvalue problem. However, in most cases the basis set used is not orthonormal. We may however orthonormalize the initial basis. We will achieve this by using the symmetric orthogonalization (see Appendix K, p. 697).

Instead of the old basis set (collected in the vector $\boldsymbol{\phi}$) in which the matrices $\mathbf{H}$ and $\mathbf{S}$ were calculated, $H_{ij}=\left\langle\phi_i|\hat{H}\phi_j\right\rangle$, $S_{ij}=\langle\phi_i|\phi_j\rangle$, we will use the orthogonal basis set $\boldsymbol{\phi}'=\mathbf{S}^{-\frac{1}{2}}\boldsymbol{\phi}$, where $\mathbf{S}^{-\frac{1}{2}}$ is computed as described in Appendix K. Then, we multiply the secular equation $(\mathbf{H}-\varepsilon\mathbf{S})\,\mathbf{c}=\mathbf{0}$ from the left-hand side by $\mathbf{S}^{-\frac{1}{2}}$ and make the following transformations:

$$\left(\mathbf{S}^{-\frac{1}{2}}\mathbf{H}-\varepsilon\mathbf{S}^{-\frac{1}{2}}\mathbf{S}\right)\mathbf{c}=0,$$
$$\left(\mathbf{S}^{-\frac{1}{2}}\mathbf{H}\mathbf{1}-\varepsilon\mathbf{S}^{-\frac{1}{2}}\mathbf{S}\right)\mathbf{c}=0,$$
$$\left(\mathbf{S}^{-\frac{1}{2}}\mathbf{H}\mathbf{S}^{-\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}-\varepsilon\mathbf{S}^{-\frac{1}{2}}\mathbf{S}\right)\mathbf{c}=0,$$
$$\left(\mathbf{S}^{-\frac{1}{2}}\mathbf{H}\mathbf{S}^{-\frac{1}{2}}\mathbf{S}^{\frac{1}{2}}-\varepsilon\mathbf{S}^{\frac{1}{2}}\right)\mathbf{c}=0,$$
$$\left(\mathbf{S}^{-\frac{1}{2}}\mathbf{H}\mathbf{S}^{-\frac{1}{2}}-\varepsilon\mathbf{1}\right)\mathbf{S}^{\frac{1}{2}}\mathbf{c}=0,$$
$$\left(\tilde{\mathbf{H}}-\varepsilon\mathbf{1}\right)\tilde{\mathbf{c}}=0,$$

with $\tilde{\mathbf{H}}=\mathbf{S}^{-\frac{1}{2}}\mathbf{H}\mathbf{S}^{-\frac{1}{2}}$ and $\tilde{\mathbf{c}}=\mathbf{S}^{\frac{1}{2}}\mathbf{c}$.

The new equation represents the eigenvalue problem, which we solve by diagonalization of $\tilde{\mathbf{H}}$ (Appendix L, p. 703). Thus,

---

[1] See Chapter 5.

the equation $(\mathbf{H}-\varepsilon\mathbf{S})\,\mathbf{c}=\mathbf{0}$ is equivalent to the eigenvalue problem $\left(\tilde{\mathbf{H}}-\varepsilon\mathbf{1}\right)\tilde{\mathbf{c}}=\mathbf{0}$. The nonorthogonality does not change the energy eigenvalues. In order to obtain $\tilde{\mathbf{H}}$, we have to diagonalize $\mathbf{S}$ to compute $\mathbf{S}^{\frac{1}{2}}$ and $\mathbf{S}^{-\frac{1}{2}}$.

Secular equation and normalization

What about normalization? If in the Ritz method we used nonnormalized basis functions, then this would not change the eigenvalues either. The only thing that would change are eigenvectors. Indeed, imagine we have solved the secular equation for the normalized basis set functions: $(\mathbf{H}-\varepsilon\mathbf{S})\,\mathbf{c}=\mathbf{0}$. The eigenvalues ε have been obtained from the secular determinant $\det(\mathbf{H}-\varepsilon\mathbf{S})=0$. Now, one wishes to destroy the normalization and takes new basis functions, which are the old basis set functions multiplied by some numbers, the i-th function by a_i. Then, a new overlap integral and the corresponding matrix element of the Hamiltonian $\hat{H}$ would be $S'_{ij}=a_i a_j S_{ij}$, $H'_{ij}=a_i a_j H_{ij}$. The new secular determinant $\det\left(\mathbf{H}'-\varepsilon\mathbf{S}'\right)$ may be expressed by the old secular determinant times a number.[2] This number is irrelevant, since what matters is that the determinant is equal to 0. Thus, whether in the secular equation we use the normalized basis set or not, the eigenvalues do not change. The eigenfunctions are also identical, although the eigenvectors $\mathbf{c}$ are different – they have to, because they multiply different functions (that are proportional to each other).

If we asked whether the eigenvalues of the matrices $\mathbf{H}$ are $\mathbf{H}'$ identical, the answer would be no.[3] However, in quantum chemistry we do not calculate the eigenvalues[4] of $\mathbf{H}$, but rather solve the secular equation $\left(\mathbf{H}'-\varepsilon\mathbf{S}'\right)\mathbf{c}=\mathbf{0}$. If $\mathbf{H}'$ changes with respect to $\mathbf{H}$, then there is a corresponding change of $\mathbf{S}'$ when compared to $\mathbf{S}$. This guarantees that ε's do not change.

---

[2] We divide the new determinant by a_1, which means dividing the elements of the first row by a_1 and in this way removing from them a_1, both in $\mathbf{H}'$ and in $\mathbf{S}'$. Doing the same with a_2 and the second row, etc., and then repeating the procedure for columns (instead of rows) we get finally the old determinant times a number.

[3] This is evident; just think of diagonal matrices.

[4] Although we often say it that way.

APPENDIX N

Slater–Condon Rules

The Slater determinants represent something like daily bread of quantum chemists. Our goal is to learn how to use the Slater determinants in computational practice and when they are involved in calculation of the mean values or the matrix elements of some important operators. We will need that in the Hartree–Fock method as well as in other important methods of quantum chemistry.

Most important are only the final results of the derivations presented in this appendix.

Antisymmetrization operator

The antisymmetrization operator is defined as

$$\hat{A} = \frac{1}{N!}\sum_P (-1)^p \hat{P}, \tag{N.1}$$

where $\hat{P}$ represents a permutation operator of N objects (in our case electrons), while $(-1)^p$ stands for the parity of the permutation P, which is "even" ("odd") if a given permutation P can be created from an even (odd) number p of the transpositions (i.e., exchanges) of two elements.

The operator $\hat{A}$ has some nice features. The most important one is that when applied to *any* function it produces either a function that is antisymmetric with respect to the permutations of N elements, or zero.[1] This means that $\hat{A}$ represents a sort of fairy's magic wand: whatever it touches it makes antisymmetric or it causes to disappear! The antisymmetrizer is also idempotent, i.e., does not change any function that is already antisymmetric, which means $\hat{A}^2 = \hat{A}$.

Let us check that $\hat{A}$ is indeed idempotent. First we obtain

$$\hat{A}^2 = (N!)^{-1}\sum_P (-1)^p \hat{P}(N!)^{-1}\sum_P (-1)^p \hat{P} = (N!)^{-2}\sum_{PP'} (-1)^{p+p'} \hat{P}\hat{P}'. \tag{N.2}$$

Of course $\hat{P}\hat{P}'$ represents a permutation operator,[2] which is then multiplied by its own parity $(-1)^{p+p'}$ and there is a sum over such permutations at a given fixed $\hat{P}'$. Independently of what

---

[1] In the near future these elements will be identified with the electronic coordinates (one element will be represented by the space and spin coordinates of a single electron: x, y, z, σ).

[2] The permutations form the permutation group.

$\hat{P}'$ is *we obtain the same result*[3] $N!$ times, and therefore $\hat{A}^2 = (N!)^{-2} N! \sum_P (-1)^p \hat{P} = \hat{A}$. This is what we wanted to show.

The operator $\hat{A}$ is Hermitian. Since $\hat{P}$ represents a (permutational) symmetry operator, it conserves the scalar product. This means that for the two vectors ψ_1 and ψ_2 of the Hilbert space we obtain[4]

$$\left\langle \psi_1(1,2,\ldots,N) | \hat{A}\psi_2(1,2,\ldots,N) \right\rangle =$$

$$(N!)^{-1} \sum_P (-1)^p \left\langle \hat{P}^{-1}\psi_1(1,2,\ldots,N) | \psi_2(1,2,\ldots,N) \right\rangle.$$

The summation over $\hat{P}$ can be replaced by the summation over $\hat{P}^{-1}$, i.e.,

$$(N!)^{-1} \sum_{P^{-1}} (-1)^p \left\langle \hat{P}^{-1}\psi_1(1,2,\ldots,N) | \psi_2(1,2,\ldots,N) \right\rangle.$$

Since the parity p of the permutation $\hat{P}^{-1}$ is the same as that of $\hat{P}$, we have $(N!)^{-1} \sum_{P^{-1}} (-1)^p \hat{P}^{-1} = \hat{A}$, which shows that $\hat{A}$ is Hermitian: $\left\langle \psi_1 | \hat{A}\psi_2 \right\rangle = \left\langle \hat{A}\psi_1 | \psi_2 \right\rangle$, or[5]

$$\hat{A}^\dagger = \hat{A}. \tag{N.3}$$

From "*Solid State and Molecular Theory*," Wiley, London, 1975 by John Slater on the permutation group: "*(...) It was at this point that Wigner, Hund, Heitler, and Weyl entered the picture, with their "Gruppenpest": the pest of the group theory, as certain disgruntled individuals who had never studied group theory in school described it. (...) The authors of the "Gruppenpest" wrote papers, which were incomprehensible to those like me who had not studied group theory (...). The practical consequences appeared to be negligible, but everyone felt that to be in the mainstream of quantum mechanics, we had to learn about it. (...) It was a frustrating experience, worthy of the name of a pest.*".

[3] Of course, $\hat{P}\hat{P}' = \hat{P}''$ has the parity $(-1)^{p+p'}$, because this is how such a permutation parity is to be calculated: first we make p transpositions in order to get $\hat{P}$, and next making p' transpositions we obtain the permutation $\hat{P}\hat{P}'$. Note that when keeping $\hat{P}'$ fixed and taking $\hat{P}$ from all possible permutations, we are running with $\hat{P}\hat{P}'$ over all possible permutations as well. This is because the complete set of permutations is obtained independently of what the starting permutation looks like, i.e., independently of $\hat{P}'$.

[4] The conservation of the scalar product $\langle \psi_1 | \psi_2 \rangle = \left\langle \hat{P}\psi_1 | \hat{P}\psi_2 \right\rangle$ means that the lengths of the vectors ψ_1 and $\hat{P}\psi_1$ are the same (similarly with ψ_2), and that the angle between the vectors is also conserved. If $\hat{P}$ is acting on ψ_2 alone and ψ_1 does not change the angle resulting from the scalar product, $\left\langle \psi_1 | \hat{P}\psi_2 \right\rangle$ is of course different, because only one of the vectors (ψ_2) has been transformed (which means a rotation of a unit vector in the Hilbert space). *The same angle* would be obtained if its partner ψ_1 were transformed in the *opposite* direction, i.e., when the operation $\hat{P}^{-1}\psi_1$ is performed. Hence from the equality of the angles we have $\left\langle \psi_1 | \hat{P}\psi_2 \right\rangle = \left\langle \hat{P}^{-1}\psi_1 | \psi_2 \right\rangle$.

[5] $\hat{A}^\dagger$ stands for the *adjoint* operator with respect to $\hat{A}$, i.e., for arbitrary functions belonging to its domain we have $\left\langle \psi_1 | \hat{A}\psi_2 \right\rangle = \left\langle \hat{A}^\dagger\psi_1 | \psi_2 \right\rangle$. There is a subtle difference (ignored in the present book) among the self-adjoint $\left(\hat{A}^\dagger = \hat{A}\right)$ and Hermitian operators in mathematical physics (they differ by definition of their domains).

Slater–Condon rules

The Slater–Condon rules serve for expressing the matrix elements involving the Slater determinants (they are many-electron wave functions the integer in parentheses stands for the space and spin coordinates of this electron):

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(N) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(N) \\ \multicolumn{4}{c}{\dots\dots\dots\dots\dots\dots\dots\dots} \\ \phi_N(1) & \phi_N(2) & \dots & \phi_N(N) \end{vmatrix}. \tag{N.4}$$

> The normalized Slater determinant has the form $\Psi = \sqrt{N!}\hat{A}\,(\phi_1\phi_2...\phi_N)$, where $\phi_1\phi_2...\phi_N$ represents a product $\phi_1\,(1)\,\phi_2\,(2)\,...\phi_N\,(N)$, and therefore, the normalization constant before the determinant itself $\det[\phi_1\,(1)\,\phi_2\,(2)\,...\phi_N\,(N)]$ is equal to $(N!)^{-1/2}$.

Quantum chemists love the Slater determinants, because they are built of the simple *one-electron* "bricks" ϕ_i called the spin orbitals (we assume them orthonormal) and because any Slater determinant is *automatically* antisymmetric with respect to exchange of the coordinates of any two electrons (shown as arguments of ϕ_i's), the factor $\frac{1}{\sqrt{N!}}$ ensures the normalization. At the same time any Slater determinant automatically satisfies the Pauli exclusion principle, because any attempt to use the same spin orbitals results in two rows being equal, and as a consequence, in having $\Psi = 0$ everywhere.[6]

The Slater determinants give quantum chemists a kind of comfort since all the integrals that appear when calculating the matrix elements of the Hamiltonian are relatively simple. The most complicated ones contain the coordinates of two electrons.

> WHAT KIND OF OPERATORS WILL WE BE DEALING WITH?
>
> 1. The sum of one-electron operators, $\hat{F} = \sum_i \hat{h}(i)$.
> 2. The sum of two-electron operators, $\hat{G} = \sum_{i<j} \hat{g}(i, j)$.
>
> In both cases the summation goes over all the electrons. Note that $\hat{h}$ has the identical form independent of the particular electron; the same pertains to $\hat{g}$.

The future meaning of the $\hat{F}$ and $\hat{G}$ operators is quite obvious; the first pertains to the noninteracting electrons (electronic kinetic energy with $\hat{h}(i) = -\frac{1}{2}\Delta_i$ or the interaction of the electrons with the nuclei), the second operator deals with the electronic repulsion, with $\hat{g}(i, j) = \frac{1}{r_{ij}}$.

---

[6] Which is a kind of catastrophe in theory, because our system *is somewhere* and can be found there with a certain nonzero probability.

WHAT ARE THE SLATER–CONDON RULES ALL ABOUT?

The Slater–Condon rules show how to express the matrix elements of many-electron operators $\hat{F}$ and $\hat{G}$ with the Slater determinants by *the matrix elements of the operators* $\hat{h}$ *and* $\hat{g}$ *computed with the orthonormal spin orbitals* ϕ_i.

The operators $\hat{F}$ and $\hat{G}$ are invariant with respect to any permutation of the electrons (Chapter 2). In other words, the formulae for $\hat{F}$ and $\hat{G}$ do not change before and after any relabeling of the electrons. This means that any permutation operator commutes with $\hat{F}$ and $\hat{G}$. Since $\hat{A}$ is a linear combination of such commuting operators, we have $\hat{A}\hat{F} = \hat{F}\hat{A}$ and $\hat{A}\hat{G} = \hat{G}\hat{A}$.

A simple trick used in the proofs below

All the proofs given below are based on the same simple trick. First, the integral under consideration is transformed into a sum of the following terms: $\langle \phi_1(1)\,\phi_2(2)\ldots\phi_N(N)\,|\hat{A}\hat{X}|\phi_1(1)\,\phi_2(2)\ldots\phi_N(N)\rangle$, where $\hat{X} = \hat{h}(i)$ or $\hat{g}(i,j)$. We recall that $\hat{A}$ is a linear combination of the permutation operators and that in the integral $\langle \phi_1(1)\,\phi_2(2)\ldots\phi_N(N)\,|\hat{X}|\phi_{n_1}(1)\,\phi_{n_2}(2)\ldots\phi_{n_N}(N)\rangle$ only a few terms will survive.

- In case $\hat{X} = \hat{h}(i)$ we obtain a product of one-electron integrals,

$$\langle \phi_1(1)\,\phi_2(2)\ldots\phi_N(N)\,|\hat{X}|\phi_{n_1}(1)\,\phi_{n_2}(2)\ldots\phi_{n_N}(N)\rangle =$$
$$= \left\langle \phi_1(1)\,|\phi_{n_1}(1)\right\rangle\left\langle \phi_2(2)\,|\phi_{n_2}(2)\right\rangle\ldots\left\langle \phi_i(i)\,|\hat{h}(i)\,|\phi_{n_i}(i)\right\rangle\ldots\left\langle \phi_N(N)\,|\phi_{n_N}(N)\right\rangle.$$

  Since the spin orbitals are orthonormal, only one term will survive, i.e., this one which has $(n_1, n_2, \ldots, n_{i-1}, n_{i+1}, \ldots n_N) = (1, 2, \ldots, i-1, i+1, \ldots N)$. All the overlap integrals that appear there are equal to 1. Only one of the one-electron integrals will give something else: $\left\langle \phi_i(i)\,|\hat{h}(i)\,|\phi_{n_i}(i)\right\rangle$, but also in this integral we have to have $n_i = i$, because of the abovementioned overlap integrals that force the matching of the indices.
- In case $\hat{X} = \hat{g}(i,j)$ we make the same transformations, but the rule for survival of the integrals pertains to the *two*-electron integral that involves the coordinates of the electrons i and j (not one electron as before). Note that this time we will have some *pairs* of the integrals that are going to survive, because the exchange of the indices $ij \to ji$ makes also an integral to survive.

Slater–Condon rule I

If ψ represents a normalized Slater determinant, then

$$\overline{F} = \langle\psi|\hat{F}|\psi\rangle = \sum_{i=1}^{N}\langle i|\hat{h}|i\rangle, \tag{N.5}$$

$$\overline{G} = \langle\psi|\hat{G}|\psi\rangle = \frac{1}{2}\sum_{i,j}(\langle ij|ij\rangle - \langle ij|ji\rangle), \tag{N.6}$$

where

$$\langle i|\hat{h}|r\rangle \equiv \sum_{\sigma_1}\int \phi_i^*(1)\hat{h}(1)\phi_r(1)\mathrm{d}V_1, \tag{N.7}$$

$$\langle ij|kl\rangle \equiv \sum_{\sigma_1}\sum_{\sigma_2}\int\int \phi_i^*(1)\phi_j^*(2)g(1,2)\phi_k(1)\phi_l(2)\mathrm{d}V_1\mathrm{d}V_2, \tag{N.8}$$

where the summation pertains to two spin coordinates (for electrons 1 and 2).

Proof.

Operator $\hat{F}$

We have

$$\overline{F} = \langle\psi|\hat{F}|\psi\rangle = N!\langle\hat{A}(\phi_1\phi_2\ldots\phi_N)|\hat{F}|\hat{A}(\phi_1\phi_2\ldots\phi_N)\rangle.$$

Using $\hat{A}\hat{F} = \hat{F}\hat{A}$, $\hat{A}^\dagger = \hat{A}$, and $\hat{A}^2 = \hat{A}$ one gets

$$\overline{F} = N!\langle\phi_1\phi_2\ldots\phi_N|\hat{A}\Big[(\hat{h}(1)\phi_1\phi_2\ldots\phi_N) + ...(\phi_1\phi_2\ldots\hat{h}(N)\phi_N)\Big]\rangle =$$
$$= \frac{N!}{N!}\langle\phi_1\phi_2\ldots\phi_N|\Big[(\hat{h}(1)\phi_1\phi_2\ldots\phi_N) + ...(\phi_1\phi_2\ldots\hat{h}(N)\phi_N)\Big]\rangle,$$

because what gives a nonzero contribution from the antisymmetrizer $\hat{A} = (N!)^{-1}(1 +$ other permutations) is only the first term with the operator of multiplication by 1. Other terms disappear after any attempt of integration. As a result we have

$$\overline{F} = \langle\phi_1|\hat{h}|\phi_1\rangle + \langle\phi_2|\hat{h}|\phi_2\rangle + \ldots\langle\phi_N|\hat{h}|\phi_N\rangle = \sum_i h_{ii}, \tag{N.9}$$

which we wanted to show.

Operator $\hat{G}$

Now let us consider the expression for $\overline{G}$. We have

$$\overline{G} = N!\langle \hat{A}(\phi_1\phi_2\ldots\phi_N)|\hat{G}\hat{A}|\phi_1\phi_2\ldots\phi_N)\rangle,$$

where once again $N!$ comes from the normalization of ψ. Taking (similarly as above) into account that $\hat{A}^\dagger = \hat{A}$, $\hat{A}^2 = \hat{A}$, $\hat{G}\hat{A} = \hat{A}\hat{G}$, we get

$$\begin{aligned}\overline{G} &= N!\langle(\phi_1\phi_2\ldots\phi_N)|\hat{A}|[\hat{g}(1,2)\phi_1\phi_2\ldots\phi_N + \hat{g}(1,3)\phi_1\phi_2\ldots\phi_N + \ldots]\rangle\\ &= \langle\phi_1(1)\phi_2(2)|\hat{g}(1,2)|\phi_1(1)\phi_2(2)\rangle - \langle\phi_1(1)\phi_2(2)|\hat{g}(1,2)|\phi_2(1)\phi_1(2)\rangle\\ &\quad + \langle\phi_1(1)\phi_3(3)|\hat{g}(1,3)|\phi_1(1)\phi_3(3)\rangle - \langle\phi_1(1)\phi_3(3)|\hat{g}(1,3)|\phi_3(1)\phi_1(3)\rangle + \ldots. \qquad \text{(N.10)}\end{aligned}$$

This transformation needs some explanation. The factor $N!$ before the integral is annihilated by $1/N!$ coming from the antisymmetrizer. The remainder of the antisymmetrizer permutes the electrons in the ket $|[\hat{g}(1,2)\phi_1\phi_2\ldots\phi_N + \hat{g}(1,3)\phi_1\phi_2\ldots\phi_N + \ldots]\rangle$. In the first term [with $\hat{g}(1,2)$] the integrals with only those permutations of the electrons $3, 4, \ldots, N$ that perfectly match the permutation $\phi_1(1)\phi_2(2)\ldots\phi_N(N)$ will survive, because otherwise the overlap integrals of the spin orbitals (over the coordinates of electrons 2, 3, ...N) will make them zero. This is why the first term will give rise to only *two* permutations that result in nonzero integrals: we will have on the first two positions $\phi_1(1)\phi_2(2)$, the other one $\phi_1(2)\phi_2(1)$. Of course, they will differ by sign; this is why we have the minus sign in the second surviving integral. A similar reasoning may be done for the term with $\hat{g}(1,3)$, as well as other terms.

Thus, we have shown that

$$\overline{G} = \sum_{i<j}(\langle ij|ij\rangle - \langle ij|ji\rangle) = \frac{1}{2}\sum_{i,j}(\langle ij|ij\rangle - \langle ij|ji\rangle), \qquad \text{(N.11)}$$

where the factor $\frac{1}{2}$ takes care of the fact that there are only $\frac{N(N-1)}{2}$ interelectronic interactions $g(i,j)$ (the upper triangle of the table $N \times N$). There is no restriction in the summation over $i, j = 1, 2, \ldots N$, because any attempt of taking the “illegal” self-interaction (corresponding to $i = j$) gives zero due to the identity of the Coulomb ($\langle ij|ij\rangle$) and exchange ($\langle ij|ji\rangle$) integrals. This is the formula we wanted to prove.

Special case: double occupation

The integrals in the expressions for $\overline{F}$ and G contain spin orbitals and the integration goes over the electronic space-and-spin coordinates. When the spin orbitals are expressed by the orbitals and the spin functions, we may perform the summation over spin coordinates. The most popular and the most important is the double occupation case, when every orbital is used to form two spin orbitals[7]:

$$\begin{aligned}\phi_1(1) &= \varphi_1(1)\alpha(1),\\ \phi_2(1) &= \varphi_1(1)\beta(1),\\ \phi_3(1) &= \varphi_2(1)\alpha(1),\\ \phi_4(1) &= \varphi_2(1)\beta(1), \qquad \text{(N.12)}\\ &\ldots \qquad \text{(N.13)}\end{aligned}$$

or

$$\begin{aligned}\phi_{2i-1}(1) &= \varphi_i(1)\alpha(1),\\ \phi_{2i}(1) &= \varphi_i(1)\beta(1), \qquad \text{(N.14)}\end{aligned}$$

for $i = 1, 2, \ldots, N/2$.

Thus, the one-electron spin orbitals that represent the building blocks of the Slater determinant are products of a spatial function (orbital φ) and one of the two simple functions of the spin coordinate σ (α or β functions, cf. p. 35).

The first Slater–Condon rule, Eq. (N.9), may be transformed as follows (for the definition of the integrals see p. 469):

$$\overline{F} = \sum_{i=1}^{N} \langle i|\hat{h}|i\rangle = \sum_{i=1}^{MO}\sum_{\sigma} \langle i\sigma|\hat{h}|i\sigma\rangle = 2\sum_{i}^{MO}(i|\hat{h}|i) \equiv 2\sum_{i}^{MO} h_{ii}, \qquad \text{(N.15)}$$

where the summations denoted by MO go over the occupied *orbitals* (their number being $N/2$), and the factor 2 results from the summation over σ, which gives the same result for the two values of σ (because of the double occupation of the orbitals).

Let us make a similar operation with $\overline{G}$. The formula for $\overline{G}$ is composed of the two parts

$$\overline{G} = \mathrm{I} - \mathrm{II}. \qquad \text{(N.16)}$$

The first part reads as

$$\mathrm{I} = \frac{1}{2}\sum_{i}^{\mathrm{MO}}\sum_{\sigma_i}\sum_{j}^{\mathrm{MO}}\sum_{\sigma_j} \langle i\ \sigma_i, j\ \sigma_j|i\ \sigma_i, j\ \sigma_j\rangle,$$

---

[7] The functions below are written as if they were dependent on the coordinates of electron 1. The reason is that we want to stress that they all are *one-electron functions*. Electron 1 serves here as an example (and when needed may be replaced by another electron). The symbol "1" means $(x_1, y_1, z_1, \sigma_1)$ if it is an argument of a spin orbital, (x_1, y_1, z_1) if it corresponds to an orbital, and σ_1 if it corresponds to a spin function.

where $i\ \sigma_i, \ldots$, etc., stand for the spin orbital composed of the orbital φ_i and a spin function that depends on σ_i. For any pair of the values of σ_i, σ_j the integral yields the same value (at a given pair of i, j) and therefore (cf. p. 469)

$$\mathrm{I} = \frac{1}{2}\sum_i^{\mathrm{MO}}\sum_j^{\mathrm{MO}} 4(ij|ij) = 2\sum_i^{\mathrm{MO}}\sum_j^{\mathrm{MO}}(ij|ij).$$

The fate of part II will be a little different:

$$\mathrm{II} = \frac{1}{2}\sum_i^{\mathrm{MO}}\sum_{\sigma_i}\sum_j^{\mathrm{MO}}\sum_{\sigma_j}\langle i\ \sigma_i, j\ \sigma_j | j\ \sigma_j, i\ \sigma_i\rangle = \frac{1}{2}\sum_i^{\mathrm{MO}}\sum_j^{\mathrm{MO}} 2(ij|ji) = \sum_i^{\mathrm{MO}}\sum_j^{\mathrm{MO}}(ij|ji),$$

because this time the summation over σ_i and σ_j gives the nonzero result in half of the cases when compared to the previous case. The pairs $(\sigma_i, \sigma_j) = (\frac{1}{2}, \frac{1}{2}), (-\frac{1}{2}, -\frac{1}{2})$ give a nonzero (and the same) result, while $(\frac{1}{2}, -\frac{1}{2}), (-\frac{1}{2}, \frac{1}{2})$ end up with zero (recall that by convention in the integral the electrons have the order 1 2 1 2). Finally the double occupation leads to

$$\overline{G} = \sum_{i,j}^{\mathrm{MO}}[2(ij|ij) - (ij|ji)]. \tag{N.17}$$

Slater–Condon rule II

Suppose we are interested in two matrix elements, $F_{12} \equiv \langle\psi_1|\hat{F}|\psi_2\rangle$ and $G_{12} \equiv \langle\psi_1|\hat{G}|\psi_2\rangle$, and the two Slater determinants ψ_1 and ψ_2 differ only by that spin orbital ϕ_i in ψ_1 has been replaced by ϕ_i' (*orthogonal to all other spin orbitals*) in ψ_2. Then the Slater–Condon rule states that

$$F_{12} = \langle i|\hat{h}|i'\rangle, \tag{N.18}$$

$$G_{12} = \sum_{j=1}\big(\langle ij|i'j\rangle - \langle ij|ji'\rangle\big). \tag{N.19}$$

Proof.

Operator $\hat{F}$

Using $\hat{F}\hat{A} = \hat{A}\hat{F}$, $\hat{A}^\dagger = \hat{A}$, and $\hat{A}^2 = \hat{A}$, we obtain $\hat{A}^\dagger\hat{F}\hat{A} = \hat{A}\hat{F}\hat{A} = \hat{A}\hat{A}\hat{F} = \hat{A}\hat{F}$ and therefore

$$F_{12} = N!\langle\phi_1 ... \phi_i ... |\hat{A}\hat{F}|\phi_1, ... \phi_i' \ldots \phi_N\rangle,$$

so we have

$$
\begin{aligned}
F_{12} &= N!\langle \phi_1\phi_2\ldots\phi_i\ldots\phi_N|\hat{A} \\
&\qquad |\left[\hat{h}(1)\phi_1,\ldots\phi_i'\ldots\phi_N+\phi_1\hat{h}(2)\phi_2\ldots\phi_i'\ldots\phi_N+\ldots\phi_1,\ldots\phi_i'\ldots\hat{h}(N)\phi_N\right]\rangle \\
&= \sum_P(-1)^p\langle \phi_1\phi_2\ldots\phi_i\ldots\phi_N| \\
&\qquad \hat{P}\left[\hat{h}(1)\phi_1,\ldots\phi_i'\ldots\phi_N+\phi_1\hat{h}(2)\phi_2\ldots\phi_i'\ldots\phi_N+\ldots\phi_1,\ldots\phi_i'\ldots\hat{h}(N)\phi_N\right]\rangle.
\end{aligned}
$$

Note first that the only integral to survive should involve ϕ_i and ϕ_i' in such a way that it leads to the one-electron integral $\langle\phi_i|\hat{h}|\phi_i'\rangle$. This however happens only if the i-th term in the square bracket intervenes (that with $\hat{h}(i)$). Indeed, let us take an integral which is *not* like that ($i \neq 1$): $\langle\phi_1\phi_2\ldots\phi_i\ldots\phi_N|\hat{P}\hat{h}(1)\phi_1\phi_2\ldots\phi_i'\ldots\phi_N\rangle$. Whatever permutation $\hat{P}$ is $\hat{h}$ will always go with ϕ_1, while ϕ_i' will be therefore *without* $\hat{h}$. When integrating over the electronic coordinates we obtain a product of one-electron integrals (for subsequent electrons), and in this product one always pinpoints the *overlap* integral of ϕ_i' multiplied by one of the spin orbitals $\phi_1,\phi_2,\ldots,\phi_N$. This integral (and therefore the whole product) is equal to 0, because ϕ_i' is orthogonal to *all* the spin orbitals. An identical reasoning can be given for $\hat{h}(2),\hat{h}(3),\ldots$, but not for $\hat{h}(i)$, and we obtain $F_{12}=\sum_P(-1)^p\langle\phi_1\phi_2\ldots\phi_i\ldots\phi_N|\hat{P}[\phi_1\phi_2\ldots\hat{h}(i)\phi_i'\ldots\phi_N]\rangle$.

The only integral to survive is that which corresponds to $\hat{P}=1$, because in other cases the orthogonality of the spin orbitals will make the product of the one-electron integrals equal to zero. Thus, finally we prove that

$$F_{12}=\langle i|h|i'\rangle. \tag{N.20}$$

Operator $\hat{G}$

From $\hat{A}^\dagger=\hat{A}$, $\hat{A}\hat{G}\hat{A}=\hat{A}\hat{A}\hat{G}=\hat{A}\hat{G}$ we obtain the following transformation

$$
\begin{aligned}
G_{12} &= N!\langle\hat{A}(\phi_1\phi_2\ldots\phi_N)|\hat{A}\hat{G}|\phi_1,\ldots\phi_i'\ldots\phi_N\rangle\rangle \\
&= N!\langle\hat{A}(\phi_1\phi_2\ldots\phi_N)|\{[\hat{g}(1,2)|\phi_1,\ldots\phi_i'\ldots\phi_N\rangle]+[\hat{g}(1,3)|\phi_1,\ldots\phi_i'\ldots\phi_N\rangle]+\ldots\}\rangle \\
&= \frac{1}{2}\sum_{k,l}{}^{'}\sum_P(-1)^p\langle\hat{P}\,(\phi_1\ldots\phi_i\ldots\phi_N)\,|\hat{g}(k,l)|\phi_1,\ldots\phi_i'\ldots\phi_N\rangle.
\end{aligned}
$$

The number of g terms is equal to the number of the interelectronic interactions. The prime in the summation $k,l=1,2,\ldots,N$ over interactions $\hat{g}(k,l)$ means that $k\neq l$ (we count the interactions twice, but the factor $\frac{1}{2}$ takes care of that). Note that, due to the orthogonality of the spin orbitals, for a given $\hat{g}(k,l)$ the integrals are all zero if $k \neq i$ and $l \neq i$. Thus, the integrals to survive have to have $k = i$ or $l = i$. Therefore (prime in the summation means the summation index i to be excluded),

$$G_{12} = \frac{1}{2}\sum_{l}{}^{'}\sum_{P}(-1)^{p}\langle\hat{P}\,(\phi_1...\phi_i...\phi_N)\,|\hat{g}(i,l)|\phi_1,...\phi_i'...\phi_N\rangle +$$
$$\frac{1}{2}\sum_{k}{}^{'}\sum_{P}(-1)^{p}\langle\hat{P}\,(\phi_1...\phi_i...\phi_N)\,|\hat{g}(k,i)|\phi_1,...\phi_i'...\phi_N\rangle =$$
$$= \frac{1}{2}\sum_{l}{}^{'}\big[\langle\phi_i\phi_l|\phi_i'\phi_l\rangle - \langle\phi_i\phi_l|\phi_l\phi_i'\rangle\big] + \frac{1}{2}\sum_{k}{}^{'}\big[\langle\phi_i\phi_k|\phi_i'\phi_k\rangle - \langle\phi_i\phi_k|\phi_k\phi_i'\rangle\big] =$$
$$\sum_{j}{}^{'}\big[\langle\phi_i\phi_j|\phi_i'\phi_j\rangle - \langle\phi_i\phi_j|\phi_j\phi_i'\rangle\big],$$

because only those two-electron integrals will survive that involve both ϕ_i *and* ϕ_i', and two other spin orbitals involved are bound to be identical (and have either the index k or l, depending on whether $l = i$ or $k = i$). The difference in the square brackets results from two successful permutations $\hat{P}$ in which we have the order i, j or j, i (in the last term). Finally, leaving for the sake of simplicity only the indices for the spin orbitals we obtain

$$G_{12} = \sum_{j(\neq i)}[\langle ij|i'j\rangle - \langle ij|ji'\rangle] \tag{N.21}$$

and after adding $0 = \langle ii|i'i\rangle - \langle ii|ii'\rangle$ we have[8]

$$G_{12} = \sum_{j}\{\langle ij|i'j\rangle - \langle ij|ji'\rangle\}. \tag{N.22}$$

This is *quod erat demonstrandum.*

Slater–Condon rule III

If ψ_1 and ψ_2 differ by two spin orbitals, say, in ψ_1 are ϕ_i and ϕ_s, and in ψ_2 we have ϕ_i' and ϕ_s' (normalized and orthogonal to themselves and to all other spin orbitals), i.e., ϕ_i' replaces ϕ_i while ϕ_s' replaces ϕ_s (all other spin orbitals are in the same order), then

---

[8] Having this formula we may forget right away that the integration has been carried out over the coordinates of the electrons i and j. It does not matter what is the symbol of the coordinate over which an integration is performed in a definite integral. When in the future we will have to declare which coordinates we are going to integrate over in $\langle ij|i'j\rangle$, then safely we can put any electrons, in the present book it will be electron 1 and electron 2.

$$F_{12} = 0, \tag{N.23}$$

$$G_{12} = \langle is|i's'\rangle - \langle is|s'i'\rangle. \tag{N.24}$$

Proof.

Operator $\hat{F}$

We have

$$\begin{aligned} F_{12} &= N!\langle(\phi_1\phi_2\ldots\phi_N)|\hat{A}\hat{F}(\phi_1'\phi_2'\ldots\phi_N')\rangle \\ &= N!\langle(\phi_1\phi_2\ldots\phi_N)|\hat{A}\{(\hat{h}(1)\phi_1'\phi_2'\ldots\phi_N') + (\phi_1'\hat{h}(2)\phi_2'\ldots\phi_N') + \ldots(\phi_1'\phi_2'\ldots\hat{h}(N)\phi_N')\}\rangle \\ &= 0, \end{aligned}$$

where the spin orbitals in ψ_2 have been labeled additionally by primes (to stress they *may* differ from those of ψ_1). In each term there will be $(N-1)$ overlap integrals between spin orbitals and one integral involving the $\hat{h}$. Therefore, there will be *always* at least one overlap integral involving *different* spin orbitals. This will produce zero.

Operator $\hat{G}$

There will be something to survive in G_{12}. Using the previous arguments we have

$$\begin{aligned} G_{12} &= N!\langle(\phi_1\phi_2\ldots\phi_N)|\hat{A}(g(1,2)\phi_1'\phi_2'\ldots\phi_N') + (g(1,3)\phi_1'\phi_2'\ldots\phi_N') + \ldots\rangle \\ &= \langle\phi_1\phi_2|g(1,2)|\phi_1'\phi_2'\rangle - \langle\phi_1\phi_2|g(1,2)|\phi_2'\phi_1'\rangle \\ &\quad + \langle\phi_1\phi_3|g(1,3)|\phi_1'\phi_3'\rangle - \langle\phi_1\phi_3|g(1,3)|\phi_3'\phi_1'\rangle + \ldots \\ &= \langle\phi_1\phi_2|\phi_1'\phi_2'\rangle - \langle\phi_1\phi_2|\phi_2'\phi_1'\rangle \\ &\quad + \langle\phi_1\phi_3|\phi_1'\phi_3'\rangle - \langle\phi_1\phi_3|\phi_3'\phi_1'\rangle \\ &\quad + \ldots \end{aligned}$$

Note that $N!$ cancels out $1/N!$ from the antisymmetrizer, and in the ket we have all possible permutations. The only term to survive has to engage all four spin orbitals: i, i', s, s', otherwise the overlap integrals will kill it. Therefore, no way, only two terms will survive and give

$$G_{12} = \langle is|i's'\rangle - \langle is|s'i'\rangle. \tag{N.25}$$

Slater–Condon rule IV

Using the above technique it is easy to show that if the Slater determinants ψ_1 and ψ_2 differ by *more than two* (orthogonal) spin orbitals, then the matrix elements $F_{12} = 0$ and $G_{12} = 0$.

This happens because the operators $\hat{F}$ and $\hat{G}$ represent a sum of at most two-electron operators, which will involve at most four spin orbitals and there will be always an extra overlap integral over orthogonal spin orbitals.[9]

Summary of the Slater–Condon rules

The Slater–Condon rules are schematically depicted in Fig. N.1.

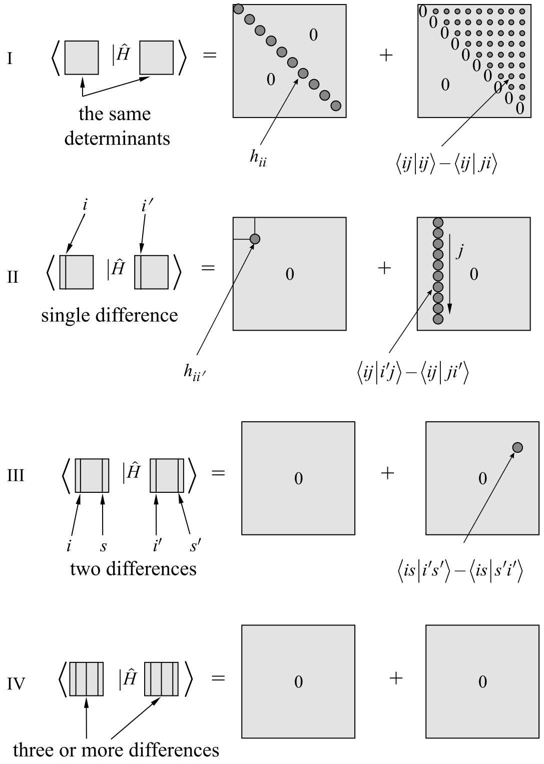


Fig. N.1. Four Slater–Condon rules (I–IV) in pictorial representation for easy remembering. On the left-hand side we see pictorial representations of matrix elements of the total Hamiltonian $\hat{H}$: the squares inside the brackets represent the Slater determinants. Vertical lines in bra stand for those spin orbitals which are different in bra (solid lines) and in ket functions (dashed lines). On the right-hand side we have two square matrices collecting the h_{ij}'s and $\langle ij|ij\rangle - \langle ij|ji\rangle$ for $i, j = 1, \ldots N$, which all have to be summed up. The dots in the matrices symbolize nonzero elements.

---

[9] If the operators were more than two-particle ones, then the result would be different.

APPENDIX O

Lagrange Multipliers Method

Imagine a Cartesian coordinate system of dimension $(n + m)$ with the axes labeled $x_1, x_2, ..., x_{n+m}$ and a function[1] $E(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, ..., x_{n+m})$. Suppose that we are interested in finding the lowest value of E, but only among such $\mathbf{x}$ that satisfy m conditions (*conditional extremum*):

$$W_i(\mathbf{x}) = 0 \tag{O.1}$$

for $i = 1, 2, ..., m$. The constraints cause the number of the *independent* variables to be n.

If we calculated the differential $\mathrm{d}E$ in point $\mathbf{x}_0$ that corresponds to an extremum of E, then we obtain 0, i.e.,

$$0 = \sum_{j=1}^{n+m} \left(\frac{\partial E}{\partial x_j}\right)_0 \mathrm{d}x_j, \tag{O.2}$$

where the derivatives are computed at the point of the extremum. The quantities $\mathrm{d}x_j$ stand for the infinitesimally small increments. From (O.2) we cannot draw the conclusion that $\left(\frac{\partial E}{\partial x_j}\right)_0$ are equal 0. This would be true if the increments $\mathrm{d}x_j$ were independent, but they are not. Indeed, one finds the relations between them by taking the differentials of the conditions W_i:

$$\sum_{j=1}^{n+m} \left(\frac{\partial W_i}{\partial x_j}\right)_0 \mathrm{d}x_j = 0 \tag{O.3}$$

for $i = 1, 2, ..., m$ (the derivatives are computed for the extremum).

---

[1] Symbol E is chosen to suggest that in our applications the quantity will have the meaning of energy.

Joseph Louis de Lagrange (1736–1813), French mathematician of Italian origin, self-taught; professor at the Artillery School of Torino, then at École Normale Supérieure in Paris. His main achievements are in variational calculus and mechanics, as well as in number theory, algebra, and mathematical analysis.

This means that the number of the truly independent increments is only n. Let us try to exploit that. To this end let us multiply each of the equations (O.3) by a number ϵ_i (*Lagrange multiplier*), which will be fixed in a moment. Then let us add together all the conditions (O.3) and subtract the result from Eq. (O.2). One gets

$$\sum_{j=1}^{n+m}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0,$$

where the summation extends over $(n+m)$ terms. The summation may be carried out in two steps. First, let us sum up the first n terms, and afterwards the sum of the other terms

$$\sum_{j=1}^{n}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j+\sum_{j=n+1}^{n+m}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0.$$

The multipliers ϵ_i until now had been treated as “undetermined.” However, we may force them to make each of the terms in the second summation to be equal to zero,[2] i.e.,

$$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0 \tag{O.4}$$

for $j=n+1,...,n+m$.

Hence, for such values of the multipliers the first summation alone is 0

$$\sum_{j=1}^{n}\left[\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0\right]\mathrm{d}x_j=0,$$

which means that now we have only n increments $\mathrm{d}x_j$ and therefore they are *independent*. Since for any (small) $\mathrm{d}x_j$ the sum is always 0, the only reason for that could be that each term in brackets equals *individually zero*, i.e.,

$$\left(\frac{\partial E}{\partial x_j}\right)_0-\sum_i \epsilon_i\left(\frac{\partial W_i}{\partial x_j}\right)_0=0 \quad \text{for } j=1,...,n.$$

---

[2] This is possible if the determinant built of coefficients $\left(\frac{\partial W_i}{\partial x_j}\right)_0$ is nonzero (this is what we have to assume). For example, if several conditions were identical, then the determinant would be zero.

This set of n equations (the so-called *Euler equation*) together with the m conditions (O.1) and m equations (O.4) gives a set of $(n+2m)$ equations with $(n+2m)$ unknowns (m epsilons and $(n+m)$ components x_i of the vector $\mathbf{x}_0$).

For a conditional extremum the constraint $W_i(\mathbf{x}) = 0$ has to be satisfied for $i = 1, 2, \ldots, m$ and $\left(\frac{\partial E}{\partial x_j}\right)_0 - \sum_i \epsilon_i \left(\frac{\partial W_i}{\partial x_j}\right)_0 = 0$ for $j = 1, \ldots, n+m$. The x_i found from these equations determine the position $\mathbf{x}_0$ of the conditional extremum of function E.

Whether it is a minimum, a maximum, or a saddle point decides the analysis of the matrix of the second derivatives (Hessian). If its eigenvalues computed at $\mathbf{x}_0$ are all positive (negative), then it is a minimum[3] (maximum); in other cases it is a saddle point.

Example 1 (Minimizing a paraboloid going along a straight line off-center). Let us take a paraboloid

$$E(x, y) = x^2 + y^2.$$

This function has, of course, a minimum at (0, 0), but the minimum is of no interest to us. What we want to find is a minimum of E, but *when* x and y satisfy some conditions. In our case there will be only one of them, i.e.,

$$W = \frac{1}{2}x - \frac{3}{2} - y = 0. \tag{O.5}$$

This means that we are interested in a minimum of E *when going along a straight line* $y = \frac{1}{2}x - \frac{3}{2}$.

The Lagrange multipliers method works as follows:

- We differentiate W and multiply by an unknown (Lagrange) multiplier ϵ, thus getting $\epsilon(\frac{1}{2}dx - dy) = 0$.
- This result (i.e., 0) is subtracted[4] from $dE = 2xdx + 2ydy = 0$ and we obtain $dE = 2xdx + 2ydy - \frac{1}{2}\varepsilon dx + \varepsilon dy = 0$.
- In the last expression the coefficients at dx and at dy have to be equal to zero.[5] In this way we obtain two equations, i.e., $2x - \frac{1}{2}\epsilon = 0$ and $2y + \epsilon = 0$.
- The third equation needed is the constraint $y = \frac{1}{2}x - \frac{3}{2}$.

---

[3] In this way we find a minimum; no information is available whether is it a global or a local one.

[4] Or added – does not matter (in such a case we get another value of ε).

[5] Only now this is possible.

- The solution to these three equations gives a set of x, y, ϵ that corresponds to an extremum. We obtain $x = \frac{3}{5}, y = -\frac{6}{5}, \epsilon = \frac{12}{5}$. Thus, we have obtained not only the position of the minimum, $x = \frac{3}{5}, y = -\frac{6}{5}$, but also the Lagrange multiplier ϵ. The minimum value of E that has been encountered when going along the straight line $y = \frac{1}{2}x - \frac{3}{2}$ is equal to $E(\frac{3}{5}, -\frac{6}{5}) = \left(\frac{3}{5}\right)^2 + \left(-\frac{6}{5}\right)^2 = \frac{9+36}{25} = \frac{9}{5}$.

Example 2 (Minimizing a paraboloid going along a circle (off-center)). Let us take the same paraboloid (O.5), but put another constraint, i.e.,

$$W = (x-1)^2 + y^2 - 1 = 0. \tag{O.6}$$

This condition means that we want to go around a circle of radius 1 centered at $(1, 0)$ and watch for which point (x, y) we will have the lowest value[6] of E. The example is chosen in such a way, as to answer the question first *without any calculations*. Indeed, the circle goes through $(0, 0)$; therefore, this point has to be found as minimum. Besides that, we should find a maximum at $(2, 0)$, because this is the point of the circle that is most distant from $(0, 0)$.

Well, let us see whether the Lagrange multipliers method will give the same.

After differentiation of W, multiplying it by the multiplier ϵ, subtracting the result from dE, and rearranging the terms, we obtain

$$dE = [2x - \epsilon\,(2x-2)]\,dx + 2y(1-\epsilon)dy = 0,$$

which (after forcing the coefficients at dx and dy to be zero) gives the following set of equations:

$$\begin{gathered} 2x - \epsilon(2x-2) = 0, \\ 2y(1-\epsilon) = 0, \\ (x-1)^2 + y^2 = 1. \end{gathered}$$

Please check that this set has the following solutions: $(x, y, \epsilon) = (0, 0, 0)$ and $(x, y, \epsilon) = (2, 0, 2)$. The solution $(x, y) = (0, 0)$ corresponds to the minimum, while the solution $(x, y) = (2, 0)$ corresponds to the maximum.[7] This is what we expected to get.

---

[6] Or, in other words, we intersect the paraboloid by the cylinder surface of radius 1 and the cylinder axis (parallel to the symmetry axis of the paraboloid) shifted to $(1, 0)$.

[7] The method does not give us information about the kind of extremum found.

Example 3 (Minimizing the mean value of the harmonic oscillator Hamiltonian). This example is different: it will pertain to the extremum of a *functional*.[8] *This is what we are often going to encounter in the methods of quantum chemistry.* Let us take the energy functional

$$E[\phi] = \int_{-\infty}^{\infty} dx \phi^* \hat{H}\phi \equiv \left\langle \phi | \hat{H}\phi \right\rangle,$$

where $\hat{H}$ stands for the harmonic oscillator Hamiltonian, $\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2$. If we were asked about what function ϕ ensures the minimum value of $E[\phi]$, then such a function could be found right away; it is $\phi = 0$. Yes, indeed, because the kinetic energy integral and the potential energy integral are positive numbers, except the situation when $\phi = 0$; then the result is zero. Wait a minute! This is not what we thought. We want that ϕ has a probabilistic interpretation, as any wave function, and therefore $\langle \phi | \phi \rangle = 1$, and not zero. Well, in such a case we want to minimize $E[\phi]$, *but the normalization condition is all the time satisfied.* Therefore, we search for the extremum of $E[\phi]$ *with the condition* $W = \langle \phi | \phi \rangle - 1 = 0$. It is easy to foresee that what the method has to produce (if it is of any value) is the normalized ground-state wave function for the harmonic oscillator. How will the Lagrange multipliers method get such a result?

The answer can be found on p. 267.

---

[8] The argument of a functional is a function that produces the value of the functional (a number).

Penalty Function Method

Very often we are interested in minimization of a ("target") function,[1] i.e., in finding such values of the variables that ensure a minimum of the function when some constraints are satisfied. Just imagine a strange Smoky Mountains hike: we want to find the point of the lowest ground elevation provided we hike along a straight line from town A to B.

Suppose the target function for minimization (that corresponds to the elevation of the ground in the Smoky Mountains region) is the function $f(x_1, x_2, ..., x_{n+m})$, but the variables x_i have to fulfill m equations ("constraints"), so

$$\phi_i(x_1, x_2, ..., x_{n+m}) = 0, \quad \text{for } i = 1, 2, ..., m.$$

At such tasks we have at least three possibilities. The first is to eliminate m variables (by using the conditions) and expressing them by the other ones. In this way the target function f takes into account all the constraints and depends only on n independent variables. Then the target function is to be minimized. The second possibility is to use the Lagrange multipliers method (see Appendix O). In both cases there is, however, a complication that the conditions to be satisfied might be quite complex and therefore the solution of the corresponding equations may be difficult to achieve. An easier solution may be to choose a penalty method. The idea behind the penalty method is quite simple. Why to take pains and try to satisfy the conditions $\phi_i = 0$, while one could propose the following: instead of the function f let us minimize its modification

$$F = f + \sum_{i=1}^{m} K_i \phi_i^2,$$

where the penalty coefficients $K_i > 0$ are chosen large.[2] When minimizing F we *admit that the conditions* $\phi_i = 0$ *could be nonsatisfied*, but any attempt to violate them introduces to F a positive contribution $\sum_{i=1}^{m} K_i \phi_i^2$. This means that for minimization of F it would be always better to explore such points of space (Fig. P.1) for which $\sum_{i=1}^{m} K_i \phi_i^2 = 0$. If K's are large enough the procedure will force to choose $\phi_i^2 = 0$, or $\phi_i = 0$, for $i = 1, 2, ..., m$, and this is what has to be satisfied during minimization.

---

[1] If we change the sign of the target function, then the task is equivalent to maximization.

[2] This means a large penalty.

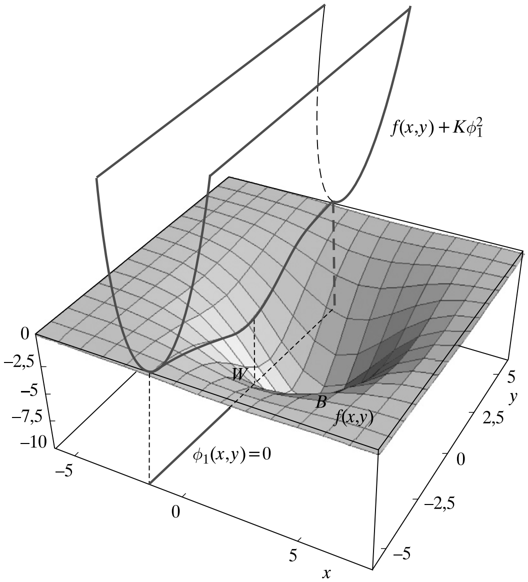


Fig. P.1. How does the penalty method work? We have to minimize $f(x, y)$, but under the condition that x and y satisfy the equation $\phi_1(x, y) = 0$ (black line). Function $f(x, y)$ exhibits a single minimum at point B, but this minimum is of no interest to us, because we are looking for a conditional minimum. To find it we minimize the sum $f(x, y) + K\phi_1^2$ with the penalty function $K\phi_1^2 \geq 0$ *allowing* any deviation from the black line $\phi_1(x, y) = 0$. However, going off this line does not pay, because this is precisely what switches the penalty on. As a result at K sufficiently large we obtain the conditional minimum W. This is what the game was all about.

Note that the task would be much more difficult if ϕ_i^2 had more than one minimum that corresponds to $\phi_i = 0$. This penalty method is worth to be in our tool box, because it is general and easily applicable. The method to work has to have a sufficiently large K. However, if K were too large, then the numerical results might be of poor quality, since the procedure would take care first of all of the penalty, paying little attention to f. It is recommended to take a few values of K and check whether the results depend on that.

As an example of the penalty function method let us take docking of two molecules. Our goal is to give such values of the atomic coordinates of both molecules as to ensure the contacts of some particular atoms of both molecules within some precise distance limits for the contacting atoms. The task sounds trivial, until we try to accomplish it in practice (especially for large molecules). The goal can be rather easily achieved when the penalty function method is used. We do the following. To the existing force field (i.e., an approximate electronic energy, Chapter 7) we add simply a penalty for not satisfying the desired contacts. For a single pair of the atoms (a contact) the penalty could be set as

$$K\left(r - r_0\right)^2,$$

where r stands for the distance of the atoms, r_0 is the optimum (desired) contact distance. At a chosen starting geometry the atoms are far from achieving the optimum distance, and therefore the force field energy is supplemented by a large distance-dependent penalty. The energy is so high that the minimization procedure tries to remove the penalty and relax the system. Often this can be done in only one way: to dock the molecules in a way to achieve the proper contact distance.

APPENDIX Q

Molecular Integrals with Gaussian-Type Orbitals 1*s*

The normalized 1*s* spherically symmetric Gaussian-type orbital (GTO) centered in the point shown by the vector $\mathbf{R}_p$ reads as

$$\chi_p \equiv \left(\frac{2\alpha_p}{\pi}\right)^{\frac{3}{4}} \exp(-\alpha_p|\mathbf{r}-\mathbf{R}_p|^2).$$

The molecular integrals usually involve at most four such orbitals: $\chi_p, \chi_q, \chi_r, \chi_s$, with the corresponding centers $\mathbf{R}_p, \mathbf{R}_q, \mathbf{R}_r, \mathbf{R}_s$, and the exponents $\alpha_p, \alpha_q, \alpha_r, \alpha_s$, respectively. Since any product of the 1*s* GTOs represents a (nonnormalized) 1*s* GTO centered between the centers of the individual GTOs (see p. 497), let us denote the center of $\chi_p\chi_q$ by $\mathbf{R}_k = \frac{\alpha_p\mathbf{R}_p+\alpha_q\mathbf{R}_q}{\alpha_p+\alpha_q}$ and the center of $\chi_r\chi_s$ by $\mathbf{R}_l = \frac{\alpha_r\mathbf{R}_r+\alpha_s\mathbf{R}_s}{\alpha_r+\alpha_s}$. Then all the integrals needed are as follows[1]:

overlap integral:

$$S_{pq} = \langle\chi_p|\chi_q\rangle = \left(\frac{4\alpha_p\alpha_q}{(\alpha_p+\alpha_q)^2}\right)^{\frac{3}{4}} \exp\left(\frac{-\alpha_p\alpha_q}{\alpha_p+\alpha_q}|\mathbf{R}_p-\mathbf{R}_q|^2\right), \tag{Q.1}$$

kinetic energy integral:

$$T_{pq} = \left\langle\chi_p\left|-\frac{1}{2}\Delta\right|\chi_q\right\rangle = \frac{\alpha_p\alpha_q}{\alpha_p+\alpha_q}\left(3-\frac{2\alpha_p\alpha_q}{\alpha_p+\alpha_q}|\mathbf{R}_p-\mathbf{R}_q|^2\right)S_{pq}, \tag{Q.2}$$

nuclear attraction integral[2]:

$$V_{pq}^{\alpha} = \left\langle\chi_p\left|\frac{1}{|\mathbf{r}-\mathbf{R}_\alpha|}\right|\chi_q\right\rangle = 2\sqrt{\frac{\alpha_p+\alpha_q}{\pi}}F_0\left((\alpha_p+\alpha_q)|\mathbf{R}_\alpha-\mathbf{R}_k|^2\right)S_{pq}, \tag{Q.3}$$

electron repulsion integral:

$$(pr|qs) = (\chi_p\chi_r|\chi_q\chi_s) = \int \chi_p(1)^*\chi_q(1)\frac{1}{r_{12}}\chi_r^*(2)\chi_s(2)dv_1dv_2 \tag{Q.4}$$

---

[1] S.F. Boys, *Proc. Roy.Soc. (London)A*, 200(1950)542.

[2] In order to interpret this integral (in a.u.) as the Coulombic attraction of the electronic charge $\chi_p^*(1)\chi_q(1)$ by the nucleus (of charge Z, located at $\mathbf{R}_\alpha$) we have to multiply the integral by $-Z$.

$$= \frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha_p + \alpha_q}\sqrt{\alpha_r + \alpha_s}}{\sqrt{\alpha_p + \alpha_q + \alpha_r + \alpha_s}} F_0 \left(\frac{(\alpha_p + \alpha_q)(\alpha_r + \alpha_s)}{\alpha_p + \alpha_q + \alpha_r + \alpha_s} |\mathbf{R}_k - \mathbf{R}_l|^2 \right) S_{pq} S_{rs},$$

with F_0 defined as[3]

$$F_0(t) = \frac{1}{\sqrt{t}} \int_0^{\sqrt{t}} \exp(-u^2) \mathrm{d}u. \qquad \text{(Q.5)}$$

Note that for an atom (all the centers coincide) we have $t = 0$ and $F_0(0) = 1$.

Do these formulae work?

The formulae look quite complex. If they are correct, they have to work in several simple situations. For example, if the electronic distribution $\chi_p^*(1)\,\chi_q(1)$ centered at $\mathbf{R}_k$ were far away from the nucleus, then as a matter of fact we had to obtain the Coulombic interaction of the charge of $\chi_p^*(1)\,\chi_q(1)$ and the nucleus. The total charge of the electron cloud $\chi_p^*(1)\,\chi_q(1)$ is obviously equal to S_{pq}, and therefore $\frac{S_{pq}}{|\mathbf{R}_\alpha - \mathbf{R}_k|}$ should be a very good estimation of the nuclear attraction integral, right?

What we need is the asymptotic form of $F_0(t)$, if $t \to \infty$. This can be deduced from our formula for $F_0(t)$. The integrand is concentrated close to $t = 0$. For $t \to \infty$, the contributions to the integral become negligible and the integral itself can be replaced by $\int_0^\infty \exp(-u^2)\mathrm{d}u = \sqrt{\pi}/2$. This gives $[F_0(t)]_{asympt.} = \frac{\sqrt{\pi}}{2\sqrt{t}}$ and $\left(V_{pq}^{\alpha}\right)_{asympt.} = 2\sqrt{\frac{\alpha_p+\alpha_q}{\pi}} F_0\left((\alpha_p + \alpha_q)|\mathbf{R}_\alpha - \mathbf{R}_k|^2\right) S_{pq} = 2\sqrt{\frac{\alpha_p+\alpha_q}{\pi}} \frac{\sqrt{\pi}}{2\sqrt{(\alpha_p+\alpha_q)|\mathbf{R}_\alpha - \mathbf{R}_k|^2}} S_{pq} = \frac{S_{pq}}{|\mathbf{R}_\alpha - \mathbf{R}_k|}$, exactly what we expected. If $\chi_p = \chi_q$, then $S_{pq} = 1$ and we get simply the Coulombic law for the unit charges. It works!

Similarly, if in the electronic repulsion integral $\chi_p = \chi_q$, $\chi_r = \chi_s$, and the distance $|\mathbf{R}_k - \mathbf{R}_l| = R$ is large, then what we should get is the Coulombic law for the two point-like unit charges at distance R. Let us see. Asymptotically

$$(pr|qs)_{asympt.} = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha_p + \alpha_q}\sqrt{\alpha_r + \alpha_s}}{\sqrt{\alpha_p + \alpha_q + \alpha_r + \alpha_s}} F_0 \left(\frac{(\alpha_p + \alpha_q)(\alpha_r + \alpha_s)}{\alpha_p + \alpha_q + \alpha_r + \alpha_s} |\mathbf{R}_k - \mathbf{R}_l|^2 \right) =$$
$$\frac{2}{\sqrt{\pi}} \frac{\sqrt{\alpha_p + \alpha_q}\sqrt{\alpha_r + \alpha_s}}{\sqrt{\alpha_p + \alpha_q + \alpha_r + \alpha_s}} \frac{\sqrt{\pi}}{2\sqrt{\frac{(\alpha_p+\alpha_q)(\alpha_r+\alpha_s)}{\alpha_p+\alpha_q+\alpha_r+\alpha_s}|\mathbf{R}_k - \mathbf{R}_l|^2}} = \frac{1}{R},$$

which is exactly we should obtain.

---

[3] The values of $F_0(t)$ are reported in L.J. Schaad, G.O. Morrell, *J. Chem. Phys.*, 54(1971)1965.

APPENDIX R

Singlet and Triplet States for Two Electrons

An angular momentum is a vector; this pertains also to spin angular momenta (see Chapter 1). The spin angular momentum of a certain number of elementary particles is a sum of their spin vectors. To obtain the total spin vector we have to add therefore the x components of the spins of the particles, and similarly the y and z components, and to construct from them the total vector. Then, we might be interested in the corresponding spin operators. These operators will be created using the Pauli matrices.[1]

Using them we obtain immediately that for a single particle the following identity holds:

$$\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2 = \hat{S}_z^2 + \hat{S}_+\hat{S}_- - \hbar\hat{S}_z, \tag{R.1}$$

where $\hat{S}_+$ and $\hat{S}_-$ are the *lowering* and *raising* operators, respectively, i.e.,

$$\hat{S}_+ = \hat{S}_x + i\hat{S}_y, \tag{R.2}$$

$$\hat{S}_- = \hat{S}_x - i\hat{S}_y, \tag{R.3}$$

that satisfy the useful relations justifying their names:

$$\begin{aligned} \hat{S}_+\alpha &= 0, \\ \hat{S}_+\beta &= \hbar\alpha, \\ \hat{S}_-\alpha &= \hbar\beta, \\ \hat{S}_-\beta &= 0. \end{aligned}$$

For any stationary state the wave function is an eigenfunction of the square of the total spin operator and of the z component of the total spin operator. The one- and two-electron cases are the only ones for which the total wave function is a *product* of space and spin parts.

The maximum projection of the electron spin on the z axis is equal to $\frac{1}{2}$ (in a.u.). Hence, the maximum projection for the total spin of two electrons is equal to 1. This means that in this case only two spin states are possible: the *singlet state* corresponding to $S = 0$ and the *triplet state* with $S = 1$ (see Postulate V). In the singlet state the two electronic spins are opposite ("pairing of electrons"), while in the triplet state the spin vectors are "parallel" (cf. Fig. 1.11 in

---

[1] See Postulate VI in Chapter 1.

Chapter 1). As always the possible projection of the total spin takes one of the following values: $M_S = -S, -S+1, \ldots, +S$, i.e., $M_S = 0$ for the singlet state and $M_S = -1, 0, +1$ for the triplet state.

Now, it will be shown that the two-electron spin function $\alpha(1)\beta(2) - \alpha(2)\beta(1)$ ensures the singlet state. First, let us construct the square of the total spin of the two electrons:

$$\mathbf{S}^2 = (\mathbf{s}_1 + \mathbf{s}_2)^2 = s_1^2 + s_2^2 + 2\mathbf{s}_1\mathbf{s}_2.$$

Thus, to create the operator $\hat{\mathbf{S}}^2$ we need the operators $\hat{s}_1^2$ and $\hat{s}_2^2$, which will be expressed by the lowering and raising operators according to Eq. (R.1), and the scalar product $\hat{\mathbf{s}}_1\hat{\mathbf{s}}_2$ expressed as a sum of products of the corresponding components x, y, and z (we know how they act, see Postulate V in Chapter 1). If $\hat{S}^2$ acts on $\alpha(1)\beta(2)$, then after five lines of derivation we obtain

$$\hat{S}^2\left[\alpha(1)\beta(2)\right] = \hbar^2[\alpha(1)\beta(2) + \alpha(2)\beta(1)],$$

and similarly

$$\hat{S}^2\left[\alpha(2)\beta(1)\right] = \hbar^2[\alpha(1)\beta(2) + \alpha(2)\beta(1)].$$

Now we will use this result to calculate $\hat{S}^2[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$ and $\hat{S}^2[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$. We have

$$\hat{S}^2[\alpha(1)\beta(2) - \alpha(2)\beta(1)] = 0 \equiv S(S+1)\hbar^2[\alpha(1)\beta(2) - \alpha(2)\beta(1)],$$

where $S = 0$ (singlet) and

$$\hat{S}^2[\alpha(1)\beta(2) + \alpha(2)\beta(1)] = 2\hbar^2[\alpha(1)\beta(2) + \alpha(2)\beta(1)] \equiv S(S+1)\hbar^2[\alpha(1)\beta(2) + \alpha(2)\beta(1)],$$

where $S = 1$ (triplet).

If the operator $\hat{S}_z = \hat{s}_{1z} + \hat{s}_{2z}$ acts on $[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$, then we obtain $0 \times [\alpha(1)\beta(2) - \alpha(2)\beta(1)]$. This means that in the singlet state the projection of the spin on the z axis is equal to 0. This is what we expect from a singlet-state function!

On the other hand, if $\hat{S}_z = \hat{s}_{1z} + \hat{s}_{2z}$ acts on $[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$, then we have $0 \times [\alpha(1)\beta(2) + \alpha(2)\beta(1)]$, i.e., the function $[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$ is such a triplet function, which corresponds to the zero projection of the total spin. A similarly simple calculation for the spin functions $\alpha(1)\alpha(2)$ and $\beta(1)\beta(2)$ gives the eigenvalue $S_z = \hbar$ and $S_z = -\hbar$, respectively. Therefore, after normalization[2] we finally have

$\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$ is a singlet function, while $\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) + \alpha(2)\beta(1)]$, $\alpha(1)\alpha(2)$, and $\beta(1)\beta(2)$ represent three triplet functions.

---

[2] For example, let us check the normalization of the singlet function $\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]$:

$$\sum_{\sigma_1}\sum_{\sigma_2}\{\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \alpha(2)\beta(1)]\}^2 =$$

$$\sum_{\sigma_1}\sum_{\sigma_2}\frac{1}{2}\{[\alpha(1)]^2[\beta(2)]^2 + [\alpha(2)]^2[\beta(1)]^2 - 2[\alpha(2)\beta(2)][\alpha(1)\beta(1)]\} =$$

$$\frac{1}{2}\{\sum_{\sigma_1}[\alpha(1)]^2\sum_{\sigma_2}[\beta(2)]^2 + \sum_{\sigma_2}[\alpha(2)]^2\sum_{\sigma_1}[\beta(1)]^2 - 2\sum_{\sigma_2}[\alpha(2)\beta(2)]\sum_{\sigma_1}[\alpha(1)\beta(1)]\} =$$

$$\frac{1}{2}\{1\cdot 1 + 1\cdot 1 - 2\cdot 0\cdot 0\} = 1.$$

The Hydrogen Molecular Ion in the Simplest Atomic Basis Set

Consider a quantum mechanical description of the hydrogen molecular ion in its simplest version. Let us use the molecular orbital theory with the atomic basis set composed of only two Slater-type orbitals (STOs): $1s_a$ and $1s_b$ centered on the nuclei a and b. The mean value of the Hamiltonian computed with the bonding (+) and antibonding (−) orbital (see Chapter 8 and Appendix D) reads as

$$E_{\pm} = \frac{H_{aa} \pm H_{ab}}{1 \pm S},$$

where the Hamiltonian (in a.u.)[1] $\hat{H} = -\frac{1}{2}\Delta - \frac{1}{r_a} - \frac{1}{r_b} + \frac{1}{R}$ and S stands for the overlap integral of the two atomic orbitals. Thus, we have

$$E_{\pm} = \frac{1}{R} + \frac{H_{aa} \pm H_{ab}}{1 \pm S} = \frac{1}{R} + \frac{\left(-\frac{1}{2}\Delta - \frac{1}{r_a} - \frac{1}{r_b}\right)_{aa} \pm \left(-\frac{1}{2}\Delta - \frac{1}{r_a} - \frac{1}{r_b}\right)_{ab}}{1 \pm S} =$$
$$\frac{1}{R} + \frac{E_H + V_{aa,b} \pm E_H S \pm V_{ab,b}}{1 \pm S} = E_H + \frac{1}{R} + \frac{V_{aa,b} \pm V_{ab,b}}{1 \pm S},$$

where E_H means the energy of the hydrogen atom, while the nuclear attraction integrals

$$V_{aa,b} = -(a|\frac{1}{r_b}|a),$$
$$V_{ab,b} = -(a|\frac{1}{r_b}|b).$$

The energy $E_{\pm}$ is a function of the internuclear distance R, which is hidden in the dependence of the integrals on R. We want to have this function explicitly. To this end we have to compute the integrals S, $V_{aa,b}$, and $V_{ab,b}$. We use the elliptic coordinates (Fig. S.1)

$$\mu = \frac{r_a + r_b}{R},$$
$$\nu = \frac{r_a - r_b}{R},$$
$$\phi = \arctan(\frac{y}{x}).$$

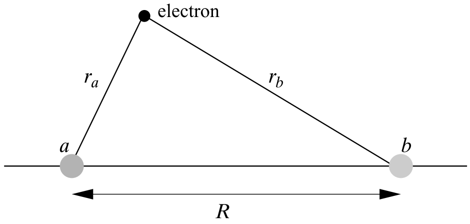


Fig. S.1. The elliptic coordinates $\mu = \frac{r_a + r_b}{R}$, $\nu = \frac{r_a - r_b}{R}$, built using the distances r_a and r_b from the two foci (where the nuclei are, their distance is R) of the ellipse. The angle ϕ measures the rotation of the plane defined by ab and the electron about the ab axis.

The volume element in the elliptic coordinates is $dV = R^3/8(\mu^2 - \nu^2)d\mu d\nu d\phi$, where $1 \leq \mu < \infty$, $-1 \leq \nu \leq 1, 0 \leq \phi \leq 2\pi$.

We will need two auxiliary integrals:

$$A_n(\sigma, \alpha) = \int_{\sigma}^{\infty} x^n \exp(-\alpha x)\, dx = \exp(-\alpha\sigma) \sum_{k=0}^{n} \frac{n!}{(n-k)!} \frac{\sigma^{n-k}}{\alpha^{k+1}},$$

$$B_n(\alpha) = \int_{-1}^{+1} x^n \exp(-\alpha x)\, dx = A_n(-1, \alpha) - A_n(1, \alpha).$$

The integrals $A_n(\sigma, \alpha)$ satisfy the following recurrence relation:

$$A_n(\sigma, \alpha) = \sigma^n A_0(\sigma, \alpha) + \frac{n}{\alpha} A_{n-1}(\sigma, \alpha),$$

$$A_0(\sigma, \alpha) = \frac{1}{\alpha} \exp(-\sigma\alpha).$$

These are some simplest auxiliary integrals (we will need them in a moment):

$$A_1(\sigma, \alpha) = \sigma \frac{1}{\alpha} \exp(-\sigma\alpha) + \frac{1}{\alpha}\frac{1}{\alpha} \exp(-\sigma\alpha) = \frac{1}{\alpha}\left(\sigma + \frac{1}{\alpha}\right) \exp(-\sigma\alpha),$$

$$A_2(\sigma, \alpha) = \sigma^2 \frac{1}{\alpha} \exp(-\sigma\alpha) + \frac{2}{\alpha}\left(\frac{1}{\alpha}\left(\sigma + \frac{1}{\alpha}\right)\exp(-\sigma\alpha)\right) =$$

$$\frac{1}{\alpha}\exp(-\sigma\alpha)\left[\sigma^2 + \frac{2}{\alpha}\left(\sigma + \frac{1}{\alpha}\right)\right],$$

$$B_0(\alpha) = \frac{1}{\alpha}\exp(\alpha) - \frac{1}{\alpha}\exp(-\alpha) = \frac{1}{\alpha}\left[\exp(\alpha) - \exp(-\alpha)\right],$$

---

[1] See Fig. S.1 for explaining symbols.

$$B_1(\alpha) = \frac{1}{\alpha}\left(-1+\frac{1}{\alpha}\right)\exp(\alpha) - \frac{1}{\alpha}\left(1+\frac{1}{\alpha}\right)\exp(-\alpha) =$$
$$\frac{1}{\alpha}\left[\left(\frac{1}{\alpha}-1\right)\exp(\alpha) - \left(\frac{1}{\alpha}+1\right)\exp(-\alpha)\right].$$

The overlap integral S is calculated in the following way:

$$S = \frac{R^3}{8\pi}\int_1^\infty d\mu \exp(-R\mu)\int_{-1}^{+1} d\nu(\mu^2-\nu^2)\int_0^{2\pi} d\phi =$$
$$\frac{R^3}{2}\left[\int_1^\infty d\mu\mu^2\exp(-R\mu) - \frac{1}{3}\int_1^\infty d\mu\exp(-R\mu)\right] =$$
$$\frac{R^3}{2}\left[A_2(1,\alpha) - \frac{1}{3}A_0(1,\alpha)\right] =$$
$$\frac{R^3}{2}\left[\frac{1}{R}\exp(-R)\left(1+\frac{2}{R}+\frac{2}{R^2}\right) - \frac{1}{3}\frac{1}{R}\exp(-R)\right] = \exp(-R)\left(\frac{R^2}{3}+R+1\right).$$

Thus, we have the explicit dependence on R. The formula for S satisfies correctly the limiting cases, $\lim_{R\to\infty} S(R) = 0$ and $\lim_{R\to 0} S(R) = 1$ (normalization of the 1s orbital). Besides $\frac{dS}{dR} = -\exp(-R)\left(\frac{R^2}{3}+R+1\right) + \exp(-R)\left(\frac{2}{3}R+1\right) = -\exp(-R)\left(\frac{R^2+R}{3}\right) < 0$, i.e., the overlap integral of the 1s functions decreases monotonically with increasing R (see Fig. S.2a).

It is seen that for small R the function S decreases gently, while for larger R its decreasing becomes fast.[2]

Using the elliptic coordinates and the formulae for the integrals $A_n(\sigma,\alpha)$ and $B_n(\alpha)$ we obtain

$$-V_{aa,b} = (a|\frac{1}{r_b}|a) = \frac{1}{\pi}\int \exp(-2r_a)\frac{1}{r_b}d\tau =$$
$$\frac{R^3}{8\pi}\frac{2}{R}\int_1^\infty d\mu \exp[-R(\mu+\nu)]\int_{-1}^{+1} d\nu\frac{(\mu^2-\nu^2)}{\mu-\nu}\int_0^{2\pi} d\phi =$$
$$\frac{R^2}{4\pi}2\pi\int_1^\infty d\mu\int_{-1}^{+1} d\nu\exp(-R\mu)\exp(-R\nu)(\mu+\nu) =$$
$$\frac{R^2}{2}\left[\int_1^\infty d\mu\mu\exp(-R\mu)\int_{-1}^{+1} d\nu\exp(-R\nu) + \int_1^\infty d\mu\exp(-R\mu)\int_{-1}^{+1} d\nu\nu\exp(-R\nu)\right] =$$
$$\frac{R^2}{2}[A_1(1,R)B_0(R) + A_0(1,R)B_1(R)] = \frac{1}{R} - \exp(-2R)\left(1+\frac{1}{R}\right).$$

---

[2] Just to get an idea: at $R = 5$ a.u. (quite typical for van der Waals complexes) the value of the overlap integral is of the order of 0.1.

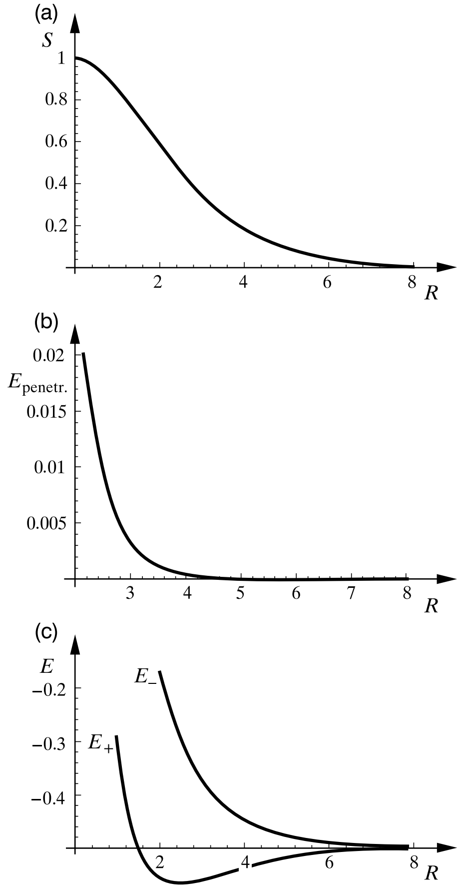


Fig. S.2. The hydrogen molecular ion in the simplest basis set of two $1s$ Slater-type orbitals (STOs). (a) The overlap integral S as a function of the internuclear distance R. (b) The penetration energy represents the difference between the electron–proton interaction calculated assuming the electronic charge distribution and the same energy calculated with the point charges (the electron is located on the nucleus a). (c) The energies E_+ and E_- of the bonding (lower curve) and of the antibonding (upper curve) orbitals, respectively.

This is an interesting result. The integral $-V_{aa,b}$ means $(a|\frac{1}{r_b}|a)$, which at large R should give the Coulombic interaction of the two unit point charges, i.e., $\frac{1}{R}$. This is what we have as the first term. The second term, $E_{\text{penetr}} = -\exp(-2R)\left(1+\frac{1}{R}\right)$, represents what is known as *penetration energy* resulting from the nonpoint-like character of one of the interacting charges.[3]

---

[3] The electron cloud of the electronic density a^2.

From Fig. S.2b it is seen that the penetration energy vanishes much faster than the overlap integral. No wonder; it vanishes as $\exp(-2R)$, while the overlap integral vanishes only as $\exp(-R)$.

It is seen that

> the diffuse charges interact weaker.

Diffuse charges offer a chance to be close in space (this increases the interaction), on the other hand in most configurations charges become more distant. The second effect prevails and therefore the penetration energy makes the Coulombic interaction weaker.

What will happen if $R \to 0$?

Let us expand the exponential function in the Taylor series. We obtain

$$\lim_{R\to 0}\left[V_{aa,b}(R)\right] = -\lim_{R\to 0}\left[\frac{1}{R} - \left[1-2R+\frac{1}{2}R^2+...\right]\left(1+\frac{1}{R}\right)\right] =$$

$$-\lim_{R\to 0}(\frac{1}{R} - 1 + 2R - \frac{1}{2}R^2 - \frac{1}{R} + 2 + \frac{1}{2}R + ...) = -1.$$

This is exactly what we get for the hydrogen atom when computing:

$$V_{aa,a} = -\int dv\frac{1}{r}(1s)^2 =$$

$$-\frac{1}{\pi}\int \exp(-2r)\,\frac{1}{r}r^2\sin\theta\, dr\, d\theta\, d\phi = -4\int_0^\infty r\exp(-2r)\,dr = -4\times 2^{-2} = -1.$$

Thus, everything is all right.

Similarly we compute

$$-V_{ab,b} = \left(a|\frac{1}{r_b}|b\right) = \frac{1}{\pi}\int \exp\left(-\left(r_a+r_b\right)\right)\frac{1}{r_b}dv =$$

$$\frac{1}{\pi}\frac{2}{R}\int \exp(-R\mu)\,\frac{1}{(\mu-\nu)}\,\frac{R^3}{8}\left(\mu^2-\nu^2\right)d\mu\, d\nu\, d\phi =$$

$$\frac{R^2}{2}\int_1^\infty\int_{-1}^{+1} d\mu\, d\nu\left[\mu\exp(-R\mu)+\nu\exp(-R\mu)\right] = \frac{R^2}{2}2A_1(1,R)+0 = (1+R)\exp(-R).$$

If $R \to \infty$, then $-V_{ab,b} \to 0$, which is the correct behavior. Do we get $V_{aa,a} = -1$ if $R \to 0$? Again, let us expand the exponential function:

$$V_{aa,a} = -\lim_{R\to 0}(1+R)\exp(-R) = -\lim_{R\to 0}(1+R)\left(1-R+\frac{R^2}{2}+...\right) =$$

$$-\lim_{R\to 0}\left[1+R-R-R^2+\frac{R^2}{2}+\ldots\right]=-1.$$

This is what we expected.

Bonding and antibonding orbital energy

If we insert the results obtained into the formula for the energy of the bonding and antibonding orbitals, then we obtain the most important formulae for the problem under consideration:

$$E_{\pm}=E_H+\frac{1}{R}+\frac{V_{aa,b}\pm V_{ab,b}}{1\pm S}=E_H+\frac{1}{R}+\frac{-\frac{1}{R}+\exp\left(-2R\right)\left(1+\frac{1}{R}\right)\pm\left(-1-R\right)\exp\left(-R\right)}{1\pm\left[\exp\left(-R\right)\left(\frac{R^2}{3}+R+1\right)\right]}.$$

The plots of $E_{\pm}$ are shown in Fig. S.2c. It is seen that in the quite primitive LCAO MO approximation the bonding energy is lower than the energy of the hydrogen atom E_H for all sufficiently large R (a single minimum), while the energy of the antibonding orbital is higher than E_H for all R (no minimum). This simple theory predicts the position of the energy minimum for the ground state as $R_e=2.5$ a.u., while the experimental value is equal[4] to ca. 2.0 a.u.

---

[4] These two quantities are not directly comparable, because the experimental value does not correspond exactly to the position of the minimum (because of anharmonicity).

APPENDIX T

Dipole Moment of a Lone Pair

The electronic lone pairs play an important role in the intermolecular interactions. In particular, a lone pair protruding in space towards its partner has a large dipole moment,[1] which may interact electrostatically with partners' multipole moments (see Appendix V2-G, p. V2-613). Let us see how the dipole moment depends on the atom to which it belongs and on the type of hybridization.

Suppose the electronic lone pair is described by the normalized hybrid

$$h = \frac{1}{\sqrt{1+\lambda^2}}[(2s) + \lambda(2p_x)],$$

with the normalized $2s$ and $2p_x$ atomic orbitals. The coefficient λ may change from $-\infty$ to $+\infty$, giving a different degree of hybridization. Fig. T.1 shows for comparison two series of the hybrids: for the carbon and for the fluorine atoms. If $\lambda = 0$, then we have the pure $2s$ orbital; if $\lambda = \pm\infty$ then we obtain the pure $\pm 2p_x$ orbital.

The dipole moment of a single electron described by h is calculated[2] as ($N = \frac{1}{\sqrt{1+\lambda^2}}$)

$$\mu_x = \langle h | -x | h \rangle = -N^2 \left[\langle 2s|x|2s \rangle + \lambda^2 \langle 2p_x|x|2p_x \rangle + 2\lambda \langle 2s|x|2p_x \rangle \right],$$

$$\mu_y = \mu_z = 0,$$

where x stands for the x coordinate of the electron.

The first two integrals are equal to zero, because the integrand represents an odd function[3] with respect to the reflection in the plane $x = 0$. As a result

$$\mu_x = -N^2 2\lambda \langle 2s|x|2p_x \rangle .$$

We will limit ourselves to $\lambda \geq 0$, which means we are considering the hybrids protruding to the right-hand side,[4] as in Fig. T.1, and since $\langle 2s|x|2p_x \rangle > 0$, we have $\mu_x \leq 0$. The negative

---

[1] Calculated with respect to the nucleus; large dipole moment means here a large length of the dipole moment vector.

[2] The atomic units have been used throughout, and therefore, μ is expressed in a.u.

[3] Please recall that the orbital $2p_x$ represents a spherically symmetric factor multiplied by x.

[4] The hybrids with $\lambda < 0$ differ only by protruding to the left-hand side.

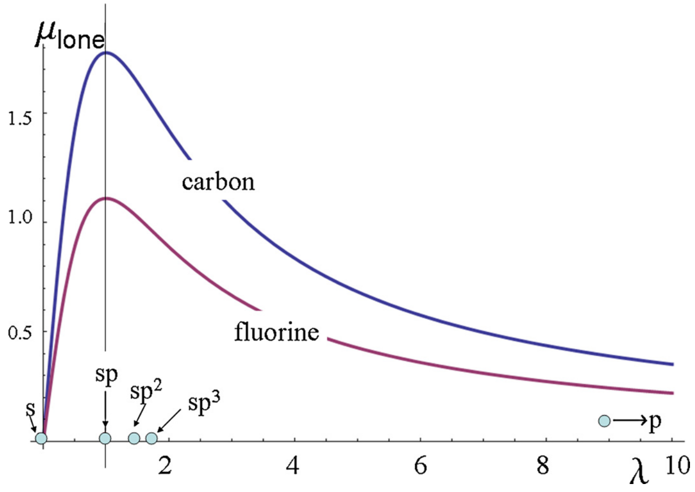


Fig. T.1. The length of the dipole moment vector (in a.u.) as a function of the mixing parameter λ for the carbon (upper curve) and fluorine (lower curve) for hybrids $h = \frac{1}{\sqrt{1+\lambda^2}}[(2s) + \lambda(2p_x)]$ that correspond to various mixing of the $2s$ and $2p_x$ carbon Slater orbitals (with the exponential factor $\zeta = 1.625$) and the fluorine orbitals ($\zeta = 2.60$); from the left (shown as small circles): $\lambda = 0$, $\lambda = 1$ (sp), $\lambda = 1.41$ (sp^2), $\lambda = 1.73$ (sp^3), λ tending to infinity. A hybrid orbital that corresponds to $\lambda < 0$ looks exactly as that with $\lambda' = -\lambda$, except it is reflected with respect to the yz plane. The maximum dipole moment corresponds to the sp hybridization.

sign stresses the fact that a negative electron is displaced to the right-hand side (positive x coordinates).

To calculate $\langle 2s|x|2p_x\rangle$ *we need to specify the atomic orbitals* $2s$ *and* $2p$. As the $2s$ and $2p$ atomic orbitals let us take the Slater-type ones, i.e.,

$$2s = N' r \exp(-\zeta r),$$
$$2p_x = N'' x \exp(-\zeta r),$$

where the exponential factor ζ (the same for both orbitals) is calculated using simple rules for building the Slater orbitals (see p. 494).

Using the integral $\int_0^\infty x^n \exp(-\alpha x) dx = n! \alpha^{-(n+1)}$, in a five-minute-long calculation we obtain the normalization constants $N' = \zeta^2 \sqrt{\frac{\zeta}{3\pi}}$ and $N'' = \zeta^2 \sqrt{\frac{\zeta}{\pi}}$. The contribution of two electrons ("electron pair") to the dipole moment is, therefore, equal to

$$\mu_{lone} = 2\mu_x = -N^2 |\lambda| (2s|xp_x) = -2N^2 N' N'' (2\lambda) \int r x^2 \exp(-2\zeta r) dv =$$
$$-2N^2 N' N'' 2\lambda \int r^3 x^2 \exp(-2\zeta r) \sin\theta dr d\theta d\phi =$$

$$-2N^2N'N''2\lambda\int_0^\infty drr^5\exp(-2\zeta r)\int_0^\pi \sin^3\theta d\theta\int_0^{2\pi}\cos^2\phi d\phi = -2N^2N'N''2\lambda\frac{5!}{(2\zeta)^6}\frac{4}{3}\pi =$$

$$-\frac{4\lambda}{(1+\lambda^2)}\zeta^2\sqrt{\frac{\zeta}{3\pi}}\zeta^2\sqrt{\frac{\zeta}{\pi}}\frac{5!}{(2\zeta)^6}\frac{4}{3}\pi = -\frac{\lambda}{(1+\lambda^2)}\frac{10}{\zeta\sqrt{3}}.$$

DIPOLE MOMENT OF A LONE PAIR

We have $\mu_{lone} = -\frac{\lambda}{(1+\lambda^2)}\frac{10}{\zeta\sqrt{3}}$.

The dipole moment at $\lambda = 0$, i.e., for the pure $2s$ orbital, is equal to 0; for $\lambda = \infty$, i.e., for the pure $2p_x$, it is also equal to 0. It is interesting for which hybridization the dipole moment has a maximum. We easily find $\frac{\partial|\mu_{lone}|}{\partial\lambda} = \frac{10}{\zeta\sqrt{3}}\frac{(1+\lambda^2)-2\lambda^2}{(1+\lambda^2)} = 0$, which gives $\lambda = \pm 1$, independently of ζ.

Thus,

the maximum of the dipole moment is at the $1:1$ mixing of $2s$ and $2p$, i.e., for the digonal hybridization (for any element) (Fig. T.1).

Table T.1. The length of dipole moments (a.u.) corresponding to doubly occupied hybrid atomic orbitals. The orbital exponents of $2s$ and $2p$ STOs are identical and computed using the rules given by Slater: $\zeta_C = 1.625$, $\zeta_N = 1.95$, $\zeta_O = 2.275$, $\zeta_F = 2.60$.

Atom	Digonal $\lambda = 1$	Trigonal $\lambda = \sqrt{2}$	Tetrahedral $\lambda = \sqrt{3}$
C	1.776	1.675	1.538
N	1.480	1.396	1.282
O	1.269	1.196	1.099
F	1.110	1.047	0.962

From Table T.1 it is seen that the dipole moment of a lone pair strongly depends on the chemical element,[5] and to a smaller extent on hybridization.

---

[5] From a practical point of view probably the most important is to compare the nitrogen and the oxygen lone pairs. Thus, a coordination of a cation by amines should correspond to a stronger interaction than that by hydroxyl groups.

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

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	$\frac{e\hbar}{2m_0c}$	$0.92731 \cdot 10^{-20}$ erg/gauss
polarizability		$\frac{e^2a_0^2}{E_h}$	$1.648778 \cdot 10^{-41}$ C^{2}m^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 1: From Quantum Physics to Chemistry

Volume 1 in a 2-volume expansion of this accessible guide to both foundational concepts and complex problems in the field of quantum chemistry

- Includes new sections on Time-Energy Uncertainty and Virial Theorem
- Provides an extensive, in-depth coverage ranging from fundamental topics to the most up-to-date concepts in quantum chemistry
- Uses an innovative structure to denote relationships among topics and allows material to be tailored to the specific needs of the readers

Ideas of Quantum Chemistry, Volume 1: From Quantum Physics to Chemistry shows how quantum mechanics is applied to molecular sciences. Beginning with an introduction to the quantum world, this book dives deep in reviewing key topics such as the Schrödinger equation, the molecular shape, the nuclear motion, and the orbital model of electronic structure.

As the 1st volume in this 2-volume expansion, *Ideas of Quantum Chemistry: From Quantum Physics to Chemistry* has been updated and revised to cover the latest developments in the field. It can be used either on its own as a detailed introduction to quantum chemistry or in combination with volume 2 as a complete overview. 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 978-0-444-64246-2

9 780444 642462

ELSEVIER

elsevier.com/books-and-journals