Simple Theorems, Proofs, and Derivations in Quantum Chemistry

MATHEMATICAL AND COMPUTATIONAL CHEMISTRY

Series Editor: PAUL G. MEZEY
University of Saskatchewan
Saskatoon, Saskatchewan

FUNDAMENTALS OF MOLECULAR SIMILARITY
Edited by Ramon Carbó-Dorca, Xavier Gironés, and Paul G. Mezey

MANY-ELECTRON DENSITIES AND REDUCED DENSITY MATRICES
Edited by Jerzy Cioslowski

SIMPLE THEOREMS, PROOFS, AND DERIVATIONS IN QUANTUM CHEMISTRY
István Mayer

A Continuation Order Plan is available for this series. A continuation order will bring delivery of each new volume immediately upon publication. Volumes are billed only upon actual shipment. For further information please contact the publisher.

Simple Theorems, Proofs, and Derivations in Quantum Chemistry

István Mayer
Chemical Research Center
Hungarian Academy of Sciences
Budapest, Hungary

Springer Science+Business Media, LLC

Library of Congress Cataloging-in-Publication Data

Mayer, István, 1943–
Simple theorems, proofs, and derivations in quantum chemistry/István Mayer.
p. cm. — (Mathematical and computational chemistry)
ISBN 978-1-4419-3389-8 ISBN 978-1-4757-6519-9 (eBook)
DOI 10.1007/978-1-4757-6519-9
1. Quantum chemistry. I. Title. II. Series.

QD462 .M39 2002
541.2'8—dc21

2002027573

ISBN 978-1-4419-3389-8

©2003 Springer Science+Business Media New York
Originally published by Kluwer Academic/Plenum Publishers , New York in 2003
Softcover reprint of the hardcover 1st edition 2003

http://www.wkap.nl/

10 9 8 7 6 5 4 3 2 1

A C.I.P. record for this book is available from the Library of Congress

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

To my wife Márti;
and to the memory of my mother and aunts,
who saved me and raised me in difficult times.

Preface

Since 1983 I have been delivering lectures at Budapest University that are mainly attended by chemistry students who have already studied quantum chemistry in the amount required by the (undergraduate) chemistry curriculum of the University, and wish to acquire deeper insight in the field, possibly in preparation of a master's or Ph.D. thesis in theoretical chemistry. In such a situation, I have the freedom to discuss, in detail, a limited number of topics which I feel are important for one reason or another. The exact coverage may vary from year to year, but I usually concentrate on the general principles and theorems and other basic theoretical results which I foresee will retain their importance despite the rapid development of quantum chemistry.

I commonly organize my lectures by treating the subject from the beginning, without referring explicitly to any actual previous knowledge in quantum chemistry—only some familiarity with its goals, approaches and, to a lesser extent, techniques is supposed. I concentrate on the formulae and their derivation, assuming the audience essentially understands the reasons for deriving these results.

This book is basically derived from the material of my lectures. The special feature, distinguishing it from most other textbooks, is that all results are explicitly proved or derived, and the derivations are presented completely, step by step. True understanding of a theoretical result can be achieved only if one has gone through its derivation.

Accordingly, this book may be considered a textbook "from the beginning but not for beginners." This means that to fully understand the book, only a familiarity with the simplest notions of quantum mechanics are required. These notions are the concepts of wave functions, operators and their matrix elements, the time-dependent and time-independent Schrödinger equation, as well as some basic knowledge of calculus, linear and matrix algebra. Any material which I suppose that may be absent from a chemist's curriculum is included either in the text or in an appendix. Therefore, anybody to whom the

book appears to be of interest based upon its title should have the necessary knowledge to read it. The book, however, is not intended as a first introduction to quantum chemistry (except, possibly, for a professional physicist); much emphasis is put on the formulae and their derivation and little on their practical utilization. (For these practical aspects I can recommend studying Veszprémi and Fehér's *Quantum Chemistry*, Kluwer Academic/Plenum Publishers 1999.) Thus, despite the limited amount of knowledge assumed, the book does not represent very easy reading: it requires the ability to follow the (sometimes-not-so-short) derivations and proofs. I think the title of the book reflects these aspects—relative simplicity and the emphasis put on formalism.

The book contains many results supposed to be familiar to professionals working in the field, but not included in introductory textbooks. At the same time, it may be rather difficult to find sources containing detailed derivations or proofs of these "commonplaces," especially such derivations which do not use a more refined formalism, e.g., second quantization. No second quantization, no density matrices, no diagrams and practically no group theory are used in this book, so it is also suited well to the needs of those who have their background in chemistry, rather than in theoretical physics.

Whereas the book is mainly intended for future professionals (students and young specialists) in theoretical chemistry and related fields (spectroscopy, material research, molecular physics etc.) I hope my fellow quantum chemists and molecular physicists will also find the book a useful source of simple proofs and derivations when they prepare their lectures.

The material in the book is presented as simply as possible. The treatment remains at the so-called "physical level of rigor". I did not follow the conventional presentation if I could develop a better one. In particular, I treat all the variational problems based on the same equivalent reformulation of the variation principle Eq.(2.13), which leads, for instance, to a very simple derivation of the Hartree-Fock equations via the Brillouin theorem. (However, in some cases the alternative derivations by using the Lagrangian multiplier technique are also shown for completeness.)

The book is self-contained and for the most part avoids questions of priority. Therefore it does not contain extensive literature surveys, only short bibliographical notes quoting some basic references and giving credit to the sources where I have learned some non-standard material.

Last, but not least, I should express my gratitude to my colleague Dr. Andrea Hamza who has read the entire manuscript and made substantial contributions to its improvement, I also thank Professor Péter R. Surján, Professor Dalia Satkovskiene, Dr. Imre Pápai and Dr. Gábor Schubert for their useful comments on some parts of the manuscript. The preparation of the manuscript has been done, for the most part, by Zsuzsa Kertész. I am indebted to her for her meticulous work in placing the bulky formulae into LaTeX.

CONTENTS

1. THE BORN–OPPENHEIMER HAMILTONIAN 1
1. Separating the center of mass motion in quantum mechanics . 1
1.1. Reducing the two-body problem to two one-body ones 1
1.2. The center of mass in quantum mechanics 6
1.3. Free atoms and atomcules 9
2. The Born–Oppenheimer approximation 12
2.1. Introductory remarks 12
2.2. The Born–Oppenheimer separation 13
2.3. Why the Born–Oppenheimer separation is not exact . 14
2.4. Approximate decoupling 15
2.5. A note on the Born–Oppenheimer separation 16
Notes 17

2. GENERAL THEOREMS AND PRINCIPLES 19
1. The variation principle 19
1.1. The Rayleigh quotient 19
1.2. The variation principle for the ground state 19
1.3. The variation principle as an equivalent of the Schrödinger equation: A useful formulation of the variation principle 21
1.4. Eckart's inequality 23
1.5. Excited states 24
2. The Hellmann–Feynman theorem 26
2.1. The differential Hellmann–Feynman theorem 26
2.2. The integral Hellmann–Feynman theorem 31
3. The virial theorem in quantum mechanics 32
3.1. Time dependence of a physical quantity 32
3.2. The virial theorem 33
3.3. Scaling—a connection with the variation principle . . 35
3.4. The virial theorem in the Born–Oppenheimer approximation 36
3.5. The virial theorem and chemical bonding 41
Notes 42

3. THE LINEAR VARIATIONAL METHOD AND LÖWDIN'S ORTHOGONALIZATION SCHEMES 45
1. The linear variational method (Ritz method) 45
2. Löwdin's symmetric orthogonalization 52
2.1. Matrix $\mathbf{S}^{-1/2}$ 52
2.2. The $\mathbf{S}^{-1/2}$ transformation 53
2.3. The Löwdin basis 54
2.4. The stationary property of Löwdin's symmetric orthogonalization scheme 55
2.5. Löwdin orthogonalization: A two-dimensional example 59
3. Linear independence of the basis and Löwdin's canonic orthogonalization 62
3.1. Eigenvalues of the overlap matrix: A measure for the linear independence of the basis 62
3.2. Löwdin's canonic orthogonalization 64
Notes 67

4. PERTURBATIONAL METHODS 69
1. Nondegenerate Rayleigh–Schrödinger perturbation theory . . 69
1.1. The problem 69
1.2. "Algebraic" expansion 72
1.3. The use of the reduced resolvent in the Rayleigh–Schrödinger perturbation theory 76
1.4. Wigner's $2n + 1$ theorem 81
1.4.1 The accuracy of the energy expectation value calculated by the nth order perturbational wave function 81
1.4.2 Computing the exact energy corrections up to order $2n + 1$ by using the first n wave function corrections . 82
2. Variational–perturbational method: The Hylleraas functional . 88
3. Degenerate Rayleigh–Schrödinger perturbation theory 93
4. Brillouin–Wigner perturbation theory 95
4.1. The size–consistency problem 98
5. Size consistency of the Rayleigh–Schrödiger perturbation theory 100
5.1. Formal considerations based on the properties of power series 100
5.2. Size consistency of the perturbational expansions . . . 102
6. Löwdin's partitioning method 115
Notes 118

5. DETERMINANT WAVE FUNCTIONS 121
1. Spin orbitals 121
2. Many-electron spin states 124
3. Slater determinants 126

3.1. Two-electron examples 127
4. The antisymmetrizing operator 129
4.1. The projection character of the antisymmetrizing operator 130
4.2. Commutation properties of the antisymmetrizing operator 134
4.3. Antisymmetrization between distant subsystems . . . 135
5. Invariance of the determinant wave function with respect to "mixing" the occupied orbitals 137
6. Matrix elements between determinant wave functions: Löwdin's general formulae for nonorthogonal orbitals 140
6.1. Overlap 140
6.1.1. General case 141
6.1.2. Factorization 142
6.1.3. Special cases for orthonormalized orbitals (Slater rules) 144
6.2. One-electron operators 145
6.2.1. General case 145
6.2.2. Special cases for orthonormalized orbitals (Slater rules) 147
6.3. Two-electron operators 149
6.3.1. General case 149
6.3.2. Special cases for orthonormalized orbitals (Slater rules) 151
7. Löwdin's pairing theorem and its generalization 153
7.1. Löwdin's pairing theorem 154
7.2. Karadakov's extended pairing theorem 157
8. Theorem about the existence of orbitals of special structure . 160
8.1. Existence theorem 161
8.2. *A posteriori* determination 162
Notes 163

6. THE HARTREE–FOCK METHOD 165
1. The variation principle for single determinant wave functions: The Brillouin theorem 166
1.1. The Brillouin theorem for the determinant with the absolute minimum of the energy 166
1.2. The Brillouin theorem for a determinant with a stationary energy 167
1.3. Algorithm of calculations based on the Brillouin theorem 169
2. The Hartree–Fock equations 170
2.1. Unrestricted Hartree–Fock equations 171
2.2. Alternate derivation by using Lagrangian multipliers . 177
2.3. Alternate derivation by using specified variations . . . 180
3. Koopmans theorem 182
3.1. The orbital energies and the total energy 184

4. The RHF method 187
4.1. The RHF and UHF schemes 187
4.2. Symmetry and the RHF method 189
4.3. The "dissociation catastrophe"; different Hartree–Fock methods 190
4.4. Singlet and triplet excitations 195
5. Finite basis Hartree–Fock theory 198
5.1. Hartree–Fock–Roothaan equations 198
5.2. The P-matrix 202
5.3. Example of using projectors in the LCAO framework: The UHF equations for successive optimization of the orbitals 204
5.4. The Fock matrix and the energy 206
6. Restricted open-shell Hartree–Fock method 209
6.1. The ROHF equations 211
6.2. The coupling coefficients for some open-shell systems 215
7. The gradient of the energy 216
Notes 223

7. POPULATION ANALYSIS, BOND ORDERS, AND VALENCES 227
1. Mulliken's population analysis 228
1.1. Electron density 228
1.2. Population analysis 230
2. Bond orders and valences 235
2.1. Bond order index 235
2.2. Valence indices 238
2.3. Exchange density and bond order 241
2.4. Bond orders in three-center bonds 245
Notes 247

8. THE ELECTRON CORRELATION 251
1. The CI expansion 251
2. The correlation energy and Nesbet's theorem 253
3. Methods of electron correlation calculations 260
3.1. The CI method 260
3.2. MC–SCF methods 262
3.3. The MBPT or Møller–Plesset method 264
3.4. The Coupled Cluster Approach 269
Notes 278

9. MISCELLANEOUS 281
1. The 1/Z expansion 281
2. The cusp condition 283
2.1. The nuclear cusp 283
2.2. The electronic cusp 286

3. The asymptotical behavior of the wave function at very large distances 286
4. Basis functions: The Gaussian product theorem 287
5. The integral transformation problem 290
Notes 291

Appendices 293
I. Separating the motion of the center of mass in classical mechanics 295
II. Reducing the two-body problem to two one-body ones in classical mechanics 299
III. Analogy between differentials and variations 303
IV. Euler's theorem for homogenous functions 305
V. The virial theorem in classical mechanics 307
VI. The electronic Schrödinger equation in atomic units 309
VII. The "bra-ket" formalism 311
1. Dirac's "bra" and "ket" vectors 311
2. Analogy with the matrix formalism 313
3. The use of an overlapping basis 314
4. Example of using the bra-ket formalism: The hypervirial theorem 315
5. Projection operators 316
6. Resolution of identity 317
7. Spectral resolution of Hermitian operators 318
8. The case of non-Hermitian operators—biorthogonal sets of functions 319
9. The trace of the projector 321
VIII. Collection of formulas for Rayleigh–Schrödinger perturbation theory (nondegenerate case) 323
IX. Direct products of matrices 325
X. Permutations 329
XI. An orthogonalization algorithm 331

Index 333

Chapter 1

THE BORN–OPPENHEIMER HAMILTONIAN

1. Separating the center of mass motion in quantum mechanics

1.1. Reducing the two-body problem to two one-body ones

The Hamiltonian for two interacting particles is, as known

$$\hat{H} = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 + V(\vec{r}_1,\, \vec{r}_2)\ , \tag{1.1}$$

where

$$\Delta_1 \equiv \nabla_1^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2} \tag{1.2}$$

and Δ_2 is analogously defined. In a closed system the potential energy V depends only on the relative position of the particles, so we may write

$$\hat{H} = -\frac{\hbar^2}{2m_1}\Delta_1 - \frac{\hbar^2}{2m_2}\Delta_2 + V(\vec{r}_1 - \vec{r}_2)\ . \tag{1.3}$$

Exactly in the manner as one does it in classical mechanics (Appendix II), we introduce the total mass of the system

$$M = m_1 + m_2 \tag{1.4}$$

and define the new coordinates $\vec{R}$ and $\vec{r}$, describing the center of the mass of the system and the relative motion, respectively, and their Cartesian components:

$$\vec{r} \;=\; \vec{r}_1 - \vec{r}_2 = \vec{i}x + \vec{j}y + \vec{k}z\ ; \tag{1.5}$$

$$\vec{R} \;=\; \frac{m_1}{M}\vec{r}_1 + \frac{m_2}{M}\vec{r}_2 = \vec{i}X + \vec{j}Y + \vec{k}Z\ ; \tag{1.6}$$

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

i. e.,

$$x = x_1 - x_2 \tag{1.7}$$

and y and z are analogously defined;

$$X = \frac{m_1}{M} x_1 + \frac{m_2}{M} x_2 \tag{1.8}$$

and again analogously for Y and Z.

Let us investigate the sum

$$\frac{1}{m_1} \frac{\partial^2 \Psi}{\partial x_1^2} + \frac{1}{m_2} \frac{\partial^2 \Psi}{\partial x_2^2} \tag{1.9}$$

by considering Ψ as a function of the coordinates $\vec{r}$ and $\vec{R}$; then it is an indirect function of x_1 and x_2 (and of the other coordinates of particles). Using the "chain rule" of differentiating indirect functions and utilizing that $\frac{\partial y}{\partial x_1} = \frac{\partial Y}{\partial x_1} = 0$, etc., we have

$$\frac{\partial \Psi}{\partial x_1} = \frac{\partial \Psi}{\partial x} \frac{\partial x}{\partial x_1} + \frac{\partial \Psi}{\partial X} \frac{\partial X}{\partial x_1} = \frac{\partial \Psi}{\partial x} \cdot 1 + \frac{\partial \Psi}{\partial X} \frac{m_1}{M} = \left(\frac{\partial}{\partial x} + \frac{m_1}{M} \frac{\partial}{\partial X} \right) \Psi \,. \tag{1.10}$$

Then similarly

$$\frac{\partial^2 \Psi}{\partial x_1^2} = \left(\frac{\partial}{\partial x} + \frac{m_1}{M} \frac{\partial}{\partial X} \right)^2 \Psi = \left[\frac{\partial^2}{\partial x^2} + 2 \frac{m_1}{M} \frac{\partial^2}{\partial x \partial X} + \left(\frac{m_1}{M} \right)^2 \frac{\partial^2}{\partial X^2} \right] \Psi \,. \tag{1.11}$$

One obtains analogously also the term $\frac{\partial^2 \Psi}{\partial x_2^2}$, but in this case $\frac{\partial x}{\partial x_2} = -1$ (and not $+1$, which we had in the previous case):

$$\begin{aligned} \frac{\partial \Psi}{\partial x_2} &= \left(-\frac{\partial}{\partial x} + \frac{m_2}{M} \frac{\partial}{\partial X} \right) \Psi \\ \frac{\partial^2 \Psi}{\partial x_2^2} &= \left[\frac{\partial^2}{\partial x^2} - 2 \frac{m_2}{M} \frac{\partial^2}{\partial x \partial X} + \left(\frac{m_2}{M} \right)^2 \frac{\partial^2}{\partial X^2} \right] \Psi \,. \end{aligned} \tag{1.12}$$

When summing, the terms containing $\frac{\partial^2}{\partial x \partial X}$ cancel:

$$\frac{1}{m_1} \frac{\partial^2 \Psi}{\partial x_1^2} + \frac{1}{m_2} \frac{\partial^2 \Psi}{\partial x_2^2} = \left[\left(\frac{1}{m_1} + \frac{1}{m_2} \right) \frac{\partial^2}{\partial x^2} + \frac{m_1 + m_2}{M^2} \frac{\partial^2}{\partial X^2} \right] \Psi \,. \tag{1.13}$$

Introducing the reduced mass μ defined by

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \,, \tag{1.14}$$

we get, by taking into account (1.4)

$$\frac{1}{m_1}\frac{\partial^2\Psi}{\partial x_1^2}+\frac{1}{m_2}\frac{\partial^2\Psi}{\partial x_2^2}=\left(\frac{1}{\mu}\frac{\partial^2}{\partial x^2}+\frac{1}{M}\frac{\partial^2}{\partial X^2}\right)\Psi\ . \tag{1.15}$$

Analogous transformations can be performed for the directions y and z, therefore

$$\hat{H}=-\frac{\hbar^2}{2\mu}\Delta_{rel}-\frac{\hbar^2}{2M}\Delta_{C.O.M.}+V(\vec{r}_{rel}) \qquad (\vec{r}_{rel}\equiv\vec{r})\ . \tag{1.16}$$

The first and third terms of this expression form the Hamiltonian

$$\hat{H}_{rel}=-\frac{\hbar^2}{2\mu}\Delta_{rel}+V(\vec{r}_{rel}) \tag{1.17}$$

for the relative motion of the two particles; this represents an effective one-particle problem corresponding to the reduced mass μ and the original potential $V(\vec{r})$ of the interparticle interaction. (We have assumed the system to be closed, i.e., there are no external fields.) The second term is the Hamiltonian corresponding to the free motion of the center of mass (C.O.M.)—also an effective one-particle Hamiltonian (with the mass M):

$$\hat{H}_{C.O.M.}=-\frac{\hbar^2}{2M}\Delta_{C.O.M.}\ . \tag{1.18}$$

The eigenfunctions of such a Hamiltonian, as it is well known, are the plane waves.

When the Hamiltonian represents a sum of two independent Hamiltonians that do not contain common variables, then the wave function represents the simple product of the respective eigenfunctions of the two Hamiltonians and the energy is the sum of the two energies. This means, that in our case the wave function Ψ depending on all the variables characterizing the system can be written as

$$\Psi(\vec{r}_1,\vec{r}_2)=\Psi(\vec{R},\vec{r})=\Phi(\vec{R})\varphi(\vec{r})\ , \tag{1.19}$$

where $\Phi(\vec{R})$ and $\varphi(\vec{r})$ represent the solutions of the time-independent Schrödinger equations written down for the C.O.M. motion and the relative motion, respectively:

$$\hat{H}_{C.O.M.}\Phi(\vec{R})=E_{C.O.M.}\Phi(\vec{R}) \tag{1.20}$$

$$\hat{H}_{rel}\varphi(\vec{r})=E_{rel}\varphi(\vec{r}) \tag{1.21}$$

and the energy E is the sum $E_{C.O.M.}+E_{rel}$. In fact, owing to the absence of the common variables, one may interchange the order of the Hamiltonian $\hat{H}_{C.O.M.}$

and of the function $\varphi(\vec{r})$, and one may similarly interchange $\hat{H}_{rel}$ with $\Phi(\vec{R})$; we have accordingly

$$\begin{aligned}\left(\hat{H}_{C.O.M.}+\hat{H}_{rel}\right)\Phi(\vec{R})\varphi(\vec{r}) &= \hat{H}_{C.O.M.}\Phi(\vec{R})\varphi(\vec{r})+\hat{H}_{rel}\Phi(\vec{R})\varphi(\vec{r})\\ &= \left(\hat{H}_{C.O.M.}\Phi(\vec{R})\right)\varphi(\vec{r})+\Phi(\vec{R})\left(\hat{H}_{rel}\varphi(\vec{r})\right) \qquad (1.22)\\ &= \left(E_{C.O.M.}+E_{rel}\right)\Phi(\vec{R})\varphi(\vec{r})\end{aligned}$$

Q.E.D.[1]

There is another way of deriving the previous Hamiltonian for the relative motion. We may assume that we are using a coordinate system moving with the same speed as the center of the mass of the two particles. That means that the overall momentum of the system is zero with respect to that system of reference. (This is equivalent to the requirement that the center of mass be at rest, but we do not make any assumption on its position.)

It is known that the overall momentum of a system is described by the operator

$$\hat{\vec{P}} = \frac{\hbar}{i}\sum_i \nabla_i \equiv \frac{\hbar}{i}\sum_i \frac{\partial}{\partial \vec{r}_i}\ , \qquad (1.23)$$

where we use the notation for the differentiation according to a vector

$$\frac{\partial}{\partial \vec{a}} = \vec{i}\frac{\partial}{\partial a_x}+\vec{j}\frac{\partial}{\partial a_y}+\vec{k}\frac{\partial}{\partial a_z}\ . \qquad (1.24)$$

It is also known that the operator $\hat{\vec{P}}$ commutes with the Hamiltonian, therefore the momentum can be measured simultaneously with the energy and represents an integral of motion (c.f., Section 3.1 of Chapter 2).

Now we shall express the sum $\frac{1}{m_1}\Delta_1+\frac{1}{m_2}\Delta_2$ by turning to new coordinates

$$\begin{aligned}\vec{r} &= \vec{r}_1-\vec{r}_2\\ \vec{\rho} &= \vec{r}_1+\vec{r}_2\ .\end{aligned} \qquad (1.25)$$

The first coordinate $\vec{r}$ in (1.25) is the same relative coordinate as used previously, whereas the second, $\vec{\rho}$ is a simple sum of $\vec{r}_1$ and $\vec{r}_2$, i.e., it is not weighted by the masses as was $\vec{R}$, the radius vector of the center of mass. Using the chain rule we have

$$\begin{aligned}\frac{\partial}{\partial \vec{r}_1} &= \frac{\partial}{\partial \vec{\rho}}\frac{\partial \vec{\rho}}{\partial \vec{r}_1}+\frac{\partial}{\partial \vec{r}}\frac{\partial \vec{r}}{\partial \vec{r}_1} = \frac{\partial}{\partial \vec{\rho}}+\frac{\partial}{\partial \vec{r}}\\ \frac{\partial}{\partial \vec{r}_2} &= \frac{\partial}{\partial \vec{\rho}}\frac{\partial \vec{\rho}}{\partial \vec{r}_2}+\frac{\partial}{\partial \vec{r}}\frac{\partial \vec{r}}{\partial \vec{r}_2} = \frac{\partial}{\partial \vec{\rho}}-\frac{\partial}{\partial \vec{r}}\ .\end{aligned} \qquad (1.26)$$

[1] Quod erat demonstrandum = which was to be demonstrated.

Now, we request the wave function Ψ to be an eigenfunction of $\hat{\vec{P}} = \frac{\hbar}{i}\left(\frac{\partial}{\partial \vec{r}_1} + \frac{\partial}{\partial \vec{r}_2}\right)$ with an eigenvalue of 0. From this requirement and the equalities in (1.26) it follows that

$$\left(\frac{\partial}{\partial \vec{r}_1} + \frac{\partial}{\partial \vec{r}_2}\right)\Psi = 2\frac{\partial}{\partial \vec{\rho}}\Psi = 0\,, \tag{1.27}$$

i.e., Ψ is independent of $\vec{\rho}$. Therefore, we may neglect all the differentiations according to $\vec{\rho}$ when computing $\Delta_1\Psi$ and $\Delta_2\Psi$ *in this particular case* and only the differentiations according to $\vec{r}$ will survive. Therefore, we are left with

$$\left(\frac{1}{m_1}\Delta_1 + \frac{1}{m_2}\Delta_2\right)\Psi\Big|_{P=0} = \left(\frac{1}{m_1} + \frac{1}{m_2}\right)\Delta_{rel}\Psi\,, \tag{1.28}$$

which means that if the overall momentum of the system is zero then its kinetic energy coincides with that of the relative motion corresponding to a reduced mass μ.

According to quantum mechanics, the energy of an electron moving in the Coulombic field of a fixed elementary positive charge is $\frac{m_e e_0^4}{2\hbar^2}$; this gives the Rydberg constant in spectroscopic units

$$R = 109737.31\mathrm{cm}^{-1}\,.$$

From spectroscopy we have for the light H isotope

$$R_H = 109677.581\mathrm{cm}^{-1}\,.$$

As the mass of the proton is ca. 1836.15 times the mass of the electron, the reduced mass for the H-atom is

$$\mu = \frac{m_e}{1 + \frac{m_e}{M_p}} \cong \frac{m_e}{1 + \frac{1}{1836.15}}\,.$$

From this we can calculate the Rydberg constant corresponding to an infinite nuclear mass (nucleus at rest):

$$R_\infty = R_H\left(1 + \frac{1}{1836.15}\right) \cong 109737.31$$

in perfect agreement with the theoretical value of R.

As proton is the lightest nucleus, the small difference between R_H and R_∞ may be considered as a hint for introducing the Born–Oppenheimer approximation.

1.2. The center of mass in quantum mechanics

If there are more than two particles, one may proceed in the following manner. At first one transforms the coordinates of the first two particles to the coordinates of relative motion and the coordinates of their center of mass, as described previously. Then one should repeat this procedure for the center of mass of the first two particles (as being a "particle" with mass $m_1 + m_2$) and the third particle, and so on: in each case the first set of three new coordinates is for the relative motion of the new particle with respect to the center of mass of the previous ones, and the second set is for the center of mass calculated for all the particles considered until that point. Then the very last three of these "Jacobi coordinates" represent the coordinates of the center of mass of the overall system. This coordinate transformation scheme can be used in both classical and quantum mechanics. However, it has the serious disadvantage that it is not symmetric with respect to the individual particles, which means that in the quantum mechanical case the indistinguishability of the electrons is not "built in." (An obvious and transparent symmetry with respect of the identical particles is a highly desirable property, of course.)

As the procedure outlined here represents a generalization of the scheme described in the previous section, one obtains again that the motion of the center of mass is described by a Hamiltonian characteristic for a free particle carrying the overall mass of the system. It is a very important fact that the center of mass of a closed system behaves as a free particle, so it should be described by a plane wave in quantum mechanics. One of the conclusions is that the wave-particle dualism is not connected with the internal structure of the individual particles but is a general law of Nature on the atomic and subatomic scale.

Another conclusion is that one cannot simply assume that we are using the system of coordinates in which the center of mass is at rest and simultaneously is at the origin of the coordinate system—that would contradict the principles of quantum mechanics. (This is possible in classical mechanics, one should only select the initial conditions appropriately—c.f. Appendix I). In quantum mechanics we can either make a statement about the position of the center of the mass or about its velocity, but not both, according to the uncertainty principle.

From the formal point of view, the essence of the problem is that in quantum mechanics the coordinates we consider are not the actual coordinates of the particles moving along a well-defined trajectory, as was in classical mechanics, but the *independent variables* for the wave function describing the probabilistic behavior of the system. Also, the wave function is defined for all possible

combinations of the coordinates of the individual particles, not just for those for which the center of mass is at the origin, i.e., for which

$$\vec{R} = \sum_{i=1}^{N} \frac{m_i}{M} \vec{r}_i = 0 \,. \tag{1.29}$$

Here and further on we denote by M the overall mass of the system:

$$M = \sum_{i=1}^{N} m_i \,. \tag{1.30}$$

We shall note in this connection that one should distinguish the concept of the center of mass in the sense of the function of the independent coordinates as shown by the first half of (1.29) from its quantum mechanical expectation value, which can be obtained by integrating over the wave function of the system.

Obviously, once the center of mass motion is separated out, we may use any set of *relative* (or internal) coordinates, not only the Jacobi ones discussed previously. The probably simplest possibility is to consider the relative coordinates of the first $N-1$ particles with respect to the center of mass of the whole system:

$$\vec{r}_i{}' = \vec{r}_i - \vec{R} = \vec{r}_i - \sum_{j=1}^{N} \frac{m_j}{M} \vec{r}_j \,. \tag{1.31}$$

Then the relative coordinates of the Nth particle are not independent parameters but can be determined on the basis of the requirement that in the C.O.M. system the center of the mass is at the origin, from which one has

$$\vec{r}_N{}' = -\sum_{j=1}^{N-1} \frac{m_j}{m_N} \vec{r}_j{}' \,. \tag{1.32}$$

This means that instead of the original set of coordinates $\vec{r}_i$, $i = 1, 2, \ldots, N$, we refer the system to the coordinates $\vec{R}$ and $\vec{r}_i{}'$, $i = 1, 2, \ldots, N-1$.

When calculating partial derivatives one has to take into account that the coordinates of the C.O.M. depend on all the original coordinates, so $\frac{\partial x_j'}{\partial x_i} = \delta_{ij} - \frac{m_i}{M}$. Then using the chain rule, we have

$$\frac{\partial}{\partial x_i} = \frac{\partial}{\partial X}\frac{\partial X}{\partial x_i} + \sum_{j=1}^{N-1} \frac{\partial}{\partial x_j'}\frac{\partial x_j'}{\partial x_i} = \frac{m_i}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial x_i'} - \frac{m_i}{M}\sum_{j=1}^{N-1}\frac{\partial}{\partial x_j'} \tag{1.33}$$

for $i = 1, 2, \ldots, N-1$, and

$$\frac{\partial}{\partial x_N} = \frac{\partial}{\partial X}\frac{\partial X}{\partial x_N} + \sum_{j=1}^{N-1} \frac{\partial}{\partial x_j'}\frac{\partial x_j'}{\partial x_N} = \frac{m_N}{M}\left(\frac{\partial}{\partial X} - \sum_{j=1}^{N-1}\frac{\partial}{\partial x_j'}\right) \tag{1.34}$$

for the Nth particle.

The second derivatives therefore are

$$\begin{aligned}\frac{\partial^2}{\partial x_i^2} &= \frac{m_i^2}{M^2}\left(\frac{\partial^2}{\partial X^2}+\sum_{j,k=1}^{N-1}\frac{\partial^2}{\partial x_j'\partial x_k'}-2\sum_{j=1}^{N-1}\frac{\partial^2}{\partial X\partial x_j'}\right)\\ &+ \frac{\partial^2}{\partial x_i'^2}+2\frac{m_i}{M}\left(\frac{\partial^2}{\partial X\partial x_i'}-\sum_{j=1}^{N-1}\frac{\partial^2}{\partial x_j'\partial x_i'}\right)\end{aligned} \tag{1.35}$$

and

$$\frac{\partial^2}{\partial x_N^2}=\frac{m_N^2}{M^2}\left(\frac{\partial^2}{\partial X^2}+\sum_{j,k=1}^{N-1}\frac{\partial^2}{\partial x_j'\partial x_k'}-2\sum_{j=1}^{N-1}\frac{\partial^2}{\partial X\partial x_j'}\right), \tag{1.36}$$

respectively.

By dividing each of these equations with the respective mass m_i (m_N) and summing, one obtains, after some trivial algebra in which equality (1.30) is utilized:

$$\sum_{i=1}^{N}\frac{1}{m_i}\frac{\partial^2}{\partial x_i^2}=\frac{1}{M}\frac{\partial^2}{\partial X^2}+\sum_{i=1}^{N-1}\frac{1}{m_i}\frac{\partial^2}{\partial x_i'^2}-\frac{1}{M}\sum_{i,j=1}^{N-1}\frac{\partial^2}{\partial x_i'x_j'}\,. \tag{1.37}$$

This result indicates that the terms with the mixed partial derivatives connecting the C.O.M. motion and the internal motion cancel, similarly to the case discussed in the previous section, but there remain some mixed partial derivatives containing pairs of different internal coordinates.

Performing the analogous calculations for the other two directions and utilizing the operator $\nabla_{i'}=\vec{i}\frac{\partial}{\partial x_{i'}}+\vec{j}\frac{\partial}{\partial y_{i'}}+\vec{k}\frac{\partial}{\partial z_i'}$ behaving as a vector, one gets the Hamiltonian in the form

$$\hat{H}=-\frac{\hbar^2}{2}\left(\frac{1}{M}\Delta_R+\sum_{i=1}^{N-1}\frac{1}{m_i}\Delta_{i'}-\frac{1}{M}\sum_{i,j=1}^{N-1}\nabla_{i'}\nabla_{j'}\right)+V\,. \tag{1.38}$$

The kinetic energy part of this Hamiltonian is a sum of those for the C.O.M. motion and for the internal one; however, the latter is not a simple sum of the kinetic energy operators for the $N-1$ individual particles the relative coordinates of which are considered as independent variables, but contains also the cross terms of the type $\nabla_{i'}\nabla_{j'}$. The appearance of similar terms is inevitable if one uses any set of internal coordinates other than the orthogonal Jacobi ones. The sum containing these terms may be expected to be relatively small if the overall mass M of the system is much larger than each of the $N-1$ masses m_i, $i=1,2,\ldots,N-1$. This is practically the case only if we consider the case

of a free atom and the Nth particle (the coordinates of which are not considered as independent variables) is its nucleus. In this case, however, the approach discussed in the next section is much more appropriate. We shall also note that in the case of a free atom with the Nth particle being the nucleus, *neglecting* these relatively small cross terms reduces the internal part of the Hamiltonian (1.38) to the Born–Oppenheimer one; c.f., Section 2.

Another disadvantage of the Hamiltonian (1.38) is that it contains the potential energy that depends on the relative coordinates of all particles, not only the $N - 1$ selected ones. This means that in the previous framework the relative coordinates of the Nth particle should always be determined by using (1.32), which makes the potential energy a more complex function of the independent coordinates. For the case of free atoms, this complication can be avoided by the scheme discussed in the next section.

1.3. Free atoms and atomcules

As discussed in the previous section, for an N-body system we have to separate out the motion of the center of mass and to select an appropriate set of $3N$-3 relative (internal) coordinates. The adequate relative coordinates in the two-body case were the three Cartesian components of the radius vector of one particle with respect to the other one. This approach can be generalized to the case of a larger number of particles, as well—we may select one particle as a reference, and consider the Cartesian components of the radius vectors of the others with respect to the reference particle as the set of internal coordinates. These internal coordinates will play the role of the independent variables of the wave function describing the internal motion. (The wave function of the motion of the center of mass is again a plane wave. It is of no interest for us.)

For a many-electron atom it is quite natural to select the *nucleus* as the reference particle and to consider the coordinates of the electrons with respect to the nucleus as the relative coordinates. These coordinates— contrary to the Jacobi coordinates, but similarly to the system of coordinates considered in the previous section–have the disadvantage that the kinetic energy part of the Hamiltonian cannot be expressed strictly as a sum of terms, each of which contains the Laplacian ($\Delta \equiv \nabla^2$ operator) corresponding to one individual coordinate. This set of coordinates, however, has the big advantage that the electrons enter it on equal footing. Also, this system can include the case of different "exotic atoms"—the "muonic atom" in which one of the electrons is replaced by a muon (μ-meson) and the antiprotonic "atomcule" (a name combined from words "atom" and "molecule") representing a system—usually atom He—in which one electron is replaced by an antiproton. (At the level of the non-relativistic Hamiltonian these changes manifests only in the larger value of the mass of one negatively charged particle.)

This system of coordinates is similar to that discussed in the previous section, in which the C.O.M. has been used as reference, but does not coincide with it: the mass of the nucleus is large with respect to that of the electrons, but the ratio of masses is not infinite; therefore the positions of the nucleus and of the C.O.M. do not strictly coincide. As the potential energy of the Coulombic system is a simple function of the distance between the particles, the present choice is more advantageous for treating free atoms.

According to the previous discussion, we are going to describe the atom (atomcule) by using the following set of coordinates:
a) The coordinates of the center of mass (including the nucleus)

$$\vec{R} = \frac{m_N}{M}\vec{R}_N + \sum_{i=1}^{N_e} \frac{m_i}{M}\vec{r}_i \ , \tag{1.39}$$

where m_N and $\vec{R}_N$ are the mass and the radius vector of the nucleus, m_i and $\vec{r}_i$ are those for the i-th electron (negatively changed particle), N_e is the number of the latter and M is the overall mass of the system:

$$M = m_N + \sum_{i=1}^{N_e} m_i \ . \tag{1.40}$$

b) The coordinates of the N_e negatively charged particles (electrons, possibly muon or antiproton) with respect to the position of the nucleus, considered as the reference:

$$\vec{r}_i{}' = \vec{r}_i - \vec{R}_N \qquad (i = 1, 2, \ldots, N_e) \ . \tag{1.41}$$

(Note that $\vec{r}_i{}'$ has a different meaning here than in the previous section.)

We have in this case $\frac{\partial X}{\partial x_i} = \frac{m_i}{M}$, $\frac{\partial X}{\partial X_N} = \frac{m_N}{M}$, $\frac{\partial x_j'}{\partial x_i} = \delta_{ij}$, and $\frac{\partial x_j'}{\partial X_N} = -1$, so the use of the chain rule gives for the first derivatives:

$$\frac{\partial}{\partial x_i} = \frac{\partial}{\partial X}\frac{\partial X}{\partial x_i} + \sum_{j=1}^{N_e} \frac{\partial}{\partial x_j'}\frac{\partial x_j'}{\partial x_i} = \frac{m_i}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial x_i'} \tag{1.42}$$

for $i = 1, 2, \ldots, N_e$; and

$$\frac{\partial}{\partial X_N} = \frac{\partial}{\partial X}\frac{\partial X}{\partial X_N} + \sum_{j=1}^{N_e} \frac{\partial}{\partial x_j'}\frac{\partial x_j'}{\partial X_N} = \frac{m_N}{M}\frac{\partial}{\partial X} - \sum_{j=1}^{N_e} \frac{\partial}{\partial x_j'} \ , \tag{1.43}$$

and for the second derivatives:

$$\frac{\partial^2}{\partial x_i^2} = \left(\frac{m_i}{M}\frac{\partial}{\partial X} + \frac{\partial}{\partial x_i'}\right)^2 = \frac{m_i^2}{M^2}\frac{\partial^2}{\partial X^2} + 2\frac{m_i}{M}\frac{\partial^2}{\partial X \partial x_i'} + \frac{\partial^2}{\partial x_i'^2}$$

$$\begin{aligned}\frac{\partial^2}{\partial X_N^2} &= \left(\frac{m_N}{M}\frac{\partial}{\partial X} - \sum_{i=1}^{N_e}\frac{\partial}{\partial x_i'}\right)\left(\frac{m_N}{M}\frac{\partial}{\partial X} - \sum_{j=1}^{N_e}\frac{\partial}{\partial x_j'}\right) \\ &= \frac{m_N^2}{M^2}\frac{\partial^2}{\partial X^2} - 2\frac{m_N}{M}\sum_{i=1}^{N_e}\frac{\partial^2}{\partial X \partial x_i'} + \sum_{i,j=1}^{N_e}\frac{\partial^2}{\partial x_i' \partial x_j'} .\end{aligned} \tag{1.44}$$

By dividing each second derivative with the respective mass and summing for all particles, we easily get

$$\frac{1}{m_N}\frac{\partial^2}{\partial X_N^2} + \sum_{i=1}^{N_e}\frac{1}{m_i}\frac{\partial^2}{\partial x_i^2} = \frac{1}{M}\frac{\partial^2}{\partial X^2} + \sum_{i=1}^{N_e}\frac{1}{m_i}\frac{\partial^2}{\partial x_i'^2} + \frac{1}{m_N}\sum_{i,j=1}^{N_e}\frac{\partial^2}{\partial x_i' \partial x_j'} . \tag{1.45}$$

This result indicates that the cross terms between the C.O.M. motion and the internal one again cancel, but there again appear some terms related to the internal motion that contain mixed partial derivatives.

By performing analogous derivations for the directions y and z, we arrive at the transformed Hamiltonian

$$\hat{H} = -\frac{\hbar^2}{2}\left(\frac{1}{M}\Delta_R + \sum_{i=1}^{N_e}\frac{1}{m_i}\Delta_{i'} + \frac{1}{m_N}\sum_{i,j=1}^{N_e}\nabla_{i'}\nabla_{j'}\right) + V . \tag{1.46}$$

Omitting the terms containing $\nabla_{i'}\nabla_{j'}$, we again arrive at the sum of the Hamiltonian of the free C.O.M. motion and of the Born–Oppenheimer electronic Hamiltonian for the internal motion (see Section 2); this is equivalent with assuming $m_N \to \infty$, which would result in a "clamped" nucleus.

However, using Hamiltonian (1.46) it is easy to account for the nuclear motion partially. This may be particularly important in the case of an "atomcule," for which one negatively charged particle has a mass m_i that is comparable with the nuclear mass m_N. One should observe that the summation indices i, j in the last sum of (1.46) run independently, so the sum contains the cases of $\nabla_{i'}\nabla_{j'}$, with $i = j$. As $\nabla_{i'}^2 \equiv \Delta_{i'}$, one can trivially transform (1.46) to

$$\hat{H} = -\frac{\hbar^2}{2}\left(\frac{1}{M}\Delta_R + \sum_{i=1}^{N_e}\left(\frac{1}{m_i} + \frac{1}{m_N}\right)\Delta_{i'} + \frac{2}{m_N}\sum_{i<j}^{N_e}\nabla_{i'}\nabla_{j'}\right) + V . \tag{1.47}$$

This result means that we have to replace the inverse of the mass of each negative particle $\frac{1}{m_i}$ by the inverse of the reduced mass corresponding to its interaction with the nucleus, i.e. by $\frac{1}{\mu_i} = \frac{1}{m_i} + \frac{1}{m_N}$. The last sum of equation (1.47) contains only true mixed partial derivatives. These remaining cross terms can be taken into account by using e.g., perturbational techniques.

Although for electrons the reduced mass differs from the mass of the free electron only slightly, the change is quite significant for an antiproton in an atomcule: assuming the nucleus to be ^{4}He, the reduced mass is only ca. 80% of the mass of the free antiproton. Also, in this case it is imperative to explicitly treat the term with the product $\nabla_{i'}\nabla_{j'}$ as the coefficient $\frac{1}{m_N}$ is only four times smaller than the coefficient $\frac{1}{m_i}$ corresponding to the antiproton. Alternately, one may treat the atomcule in the framework of the Born–Oppenheimer approximation, considering the antiproton as the second (although negatively charged) nucleus.

2. The Born–Oppenheimer approximation

2.1. Introductory remarks

Most studies in quantum chemistry and theoretical molecular physics are dealing with the (approximate) solutions of the time-independent non-relativistic Schrödinger equation written in the Born–Oppenheimer approximation.

Pictorially one justifies the Born–Oppenheimer scheme by saying "The motion of the light electrons follows the motion of the heavy nuclei," i.e., the nuclei usually move slowly enough so the "electronic cloud" could be rearranged instantaneously according to the change in the position of the nuclei.

Here is another formulation of the same idea: as the electrons move faster, they "see" the instantaneous position of the nuclei, whereas the nuclei "feel" only the averaged potential of the electronic cloud (electrostatic field caused by the spatially spread electronic charge). Therefore, the electronic distribution can be studied by considering the nuclei as being fixed (so-called "clamped nuclei" approximation). Then the resulting total energy of the electrons will depend on the nuclear configuration; this electronic energy, together with the energy of the mutual repulsion of the nuclei, determines the potential energy (hyper-)surface (PES) governing the motion of the nuclei. This is called the "Born–Oppenheimer separation of the electronic and nuclear motion," or simply the "Born–Oppenheimer approximation."

In reality there is a weak ("vibronic"=vibrational–electronic or vibro–electronic) coupling between the equations describing the electronic and nuclear motions; these terms, however, may be neglected in most of the actual calculations.

The time-independent Schrödinger equation is

$$\hat{H}(R,r)\Psi(R,r) = \varepsilon\Psi(R,r) \ . \tag{1.48}$$

Here $\{R\}$ and $\{r\}$ are the sets of all nuclear and electronic coordinates, respectively, and $\hat{H}(R,r)$ is the total (electrons + nuclei) molecular Hamiltonian:

$$\hat{H}(R,r) = \underbrace{\underbrace{\sum_{\alpha=1}^{N_N} -\frac{\hbar^2}{2M_\alpha}\Delta_\alpha}_{\hat{T}_N \text{ nuclear}} + \underbrace{\sum_{i=1}^{N_e} -\frac{\hbar^2}{2m_e}\Delta_i}_{\hat{T}_e \text{ electronic}}}_{\text{kinetic energy}} + \underbrace{V(R,r)}_{V \text{ potential energy}} \tag{1.49}$$

with

$$V \equiv V(R,r) = \underbrace{-\sum_{\alpha=1}^{N_N}\sum_{i=1}^{N_e} e_0^2\frac{Z_\alpha}{r_{\alpha i}}}_{V_{eN}} + \underbrace{\sum_{\alpha<\beta} e_0^2\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}}_{V_{NN}} + \underbrace{\sum_{i<j} e_0^2\frac{1}{r_{ij}}}_{V_{ee}} \tag{1.50}$$

As we are using a nonrelativistic approximation, the Hamiltonian $\hat{H}$ contains only kinetic energy and electrostatic interactions; the weaker, e.g., magnetic (spin–spin, spin–orbit) interactions are neglected. These neglected terms should be taken into account in *very* accurate calculations or in describing some magnetic phenomena; also, the importance of the relativistic effects is increasing with the atomic number, so relativistic effects are often accounted for in the case of heavy atoms.

2.2. The Born–Oppenheimer separation

One considers $\Psi(R,r)$ as a simple product

$$\Psi(R,r) \approx \Psi_e(r;R)\Psi_N(R)\ , \tag{1.51}$$

where $\Psi_e(r;R)$ is the electronic wave function in the field of the *fixed* nuclei ($\{R\} = const.$), and $\Psi_N(R)$ is the nuclear wave function. The problem is divided into two separate steps.

1. Electronic Schrödinger equation:

$$\hat{H}_e\Psi_e(r;R) = E(R)\Psi_e(r;R) \tag{1.52}$$

where

$$\hat{H}_e = \hat{T}_e + V(R,r) = -\frac{\hbar^2}{2m_e}\sum_{i=1}^{N_e}\Delta_i + V(R,r) \tag{1.53}$$

is the electronic Hamiltonian (including the energy of the nuclear–nuclear repulsion).

$E(R)$ and $\Psi_e(r;R)$ are computed at *fixed* values of the coordinates $\{R\}$; accordingly, they depend on the nuclear coordinates as parameters. Equation

(1.52) usually has many different solutions $\Psi_e(r;R)$; each of them determines a PES $E(R)$ for the given (ground or excited) electronic state. $E(R)$ determined in this manner is then used as the potential energy for investigating the nuclear motion:

2. Nuclear Schrödinger equation:

$$\hat{H}_N \Psi_N(R) = \varepsilon \Psi_N(R) \tag{1.54}$$

where

$$\hat{H}_N = -\sum_{\alpha=1}^{N_N} \frac{\hbar^2}{2M_\alpha} \Delta_\alpha + E(R) \tag{1.55}$$

is the Hamiltonian of the nuclear motion. Note that (contrary to Ψ_e) Ψ_N is a function of the nuclear coordinates *only*.

2.3. Why the Born–Oppenheimer separation is not exact

We substitute the product form of the wave function in the left-hand-side (l.h.s.) of the overall (electrons + nuclei) Schrödinger equation:

$$\hat{H}(R,r)\Psi(R,r) = \hat{H}(R,r)\Psi_e(r;R)\Psi_N(R) \ . \tag{1.56}$$

When expanding this, the critical term is

$$\hat{T}_N = \sum_\alpha -\frac{\hbar^2}{2M_\alpha} \Delta_\alpha \ . \tag{1.57}$$

(In the followings we shall omit the summation limits when they are obvious.)

$$\Delta_\alpha = \frac{\partial^2}{\partial X_\alpha^2} + \frac{\partial^2}{\partial Y_\alpha^2} + \frac{\partial^2}{\partial Z_\alpha^2} \tag{1.58}$$

We compute the second derivative of a product:

$$\frac{\partial^2}{\partial X_\alpha^2} [\Psi_e(r;R)\Psi_N(R)] = \Psi_N \frac{\partial^2 \Psi_e}{\partial X_\alpha^2} + 2\frac{\partial \Psi_e}{\partial X_\alpha}\frac{\partial \Psi_N}{\partial X_\alpha} + \Psi_e \frac{\partial^2 \Psi_N}{\partial X_\alpha^2} \ . \tag{1.59}$$

Performing analogously the differentiation according to Y_α Z_α, we have

$$\begin{aligned} &\sum_\alpha -\frac{\hbar^2}{2M_\alpha} \Delta_\alpha [\Psi_e(r;R)\Psi_N(R)] \\ &= \sum_\alpha -\frac{\hbar^2}{2M_\alpha} [\Psi_N \Delta_\alpha \Psi_e + 2(\nabla_\alpha \Psi_e \nabla_\alpha \Psi_N) + \Psi_e \Delta_\alpha \Psi_N] \ . \end{aligned} \tag{1.60}$$

The first two terms, $\Psi_N \Delta_\alpha \Psi_e$ and $2(\nabla_\alpha \Psi_e \nabla_\alpha \Psi_N)$, contain derivatives of the electronic wave function according to the nuclear coordinates—no such derivatives enter the electronic and nuclear Schrödinger equations (1.52) and (1.54),

indicating that the consecutive solution of (1.52) and (1.54) cannot provide an exact solution of the overall molecular Schrödinger equation. The terms in question produce some (usually weak, but not always negligible) coupling between different Born–Oppenheimer electronic states through the nuclear motion.

Substituting into the overall (electrons + nuclei) Schrödinger equation and denoting by B the sum of all coupling terms

$$B = \sum_{\alpha} -\frac{\hbar^2}{2M_\alpha}\left[\Psi_N \Delta_\alpha \Psi_e + 2(\nabla_\alpha \Psi_e \nabla_\alpha \Psi_N)\right] \tag{1.61}$$

we can write

$$\begin{aligned}
\hat{H}\Psi_e\Psi_N &= B + \Psi_e \underbrace{\sum_{\alpha} -\frac{\hbar^2}{2M_\alpha}\Delta_\alpha \Psi_N}_{\hat{T}_N\Psi_N} + \underbrace{\sum_{i} -\frac{\hbar^2}{2m_e}\Delta_i \Psi_e \Psi_N}_{(\hat{T}_e\Psi_e)\Psi_N = \Psi_N(\hat{T}_e\Psi_e)} + V\Psi_e\Psi_N \\
&= B + \Psi_e\hat{T}_N\Psi_N + \Psi_N \underbrace{(\hat{T}_e\Psi_e + V\Psi_e)}_{\substack{\text{the l.h.s. of the electronic} \\ \text{Schrödinger equation} = E(R)\Psi_e}} \qquad (1.62) \\
&= B + \Psi_e\hat{T}_N\Psi_N + \Psi_N E(R)\Psi_e = B + \Psi_e \underbrace{\left[\hat{T}_N + E(R)\right]\Psi_N}_{\substack{\text{the l.h.s. of the nuclear} \\ \text{Schrödinger equation} = \varepsilon\Psi_N}} \\
&= B + \varepsilon\Psi_e\Psi_N
\end{aligned}$$

We see that $\Psi_e(r;R)\Psi_N(R)$ is not an exact eigenfunction because of the presence of the coupling terms B.

2.4. Approximate decoupling

a) Born–Oppenheimer approximation (separation):

Neglecting the coupling terms completely—the most common approach.

The coupling terms are relatively small because they represent derivatives of the electronic wave function, but the denominators contain the nuclear masses which are much larger than the mass of the electron. (The electronic wave function strongly depends on the electron–nuclear distances, so the derivatives according to $\{r\}$ and $\{R\}$ are expected to be of the same order of magnitude.) The coupling terms can usually be neglected around the equilibrium configurations of molecules, and usually are (much) less than the errors we admit when solving the equations in the Born–Oppenheimer approximation.

b) Adiabatic approximation:

(This term has no relationship with the concept of adiabaticity in thermodynamics. Some authors call the Born–Oppenheimer approximation also "adiabatic"—not quite correct.)

One computes the average value of the coupling terms for a given electronic state Ψ_e (the diagonal matrix elements in terms of the different electronic states) and adds

$$\sum_{\alpha} -\frac{\hbar^2}{2M_\alpha}(< \Psi_e|\Delta_\alpha|\Psi_e > +2 < \Psi_e|\nabla_\alpha|\Psi_e > \nabla_\alpha) \tag{1.63}$$

to $\hat{H}_N$ used in the nuclear Schrödinger equation. Note: the second term is still an operator in terms of R, which acts on $\Psi_N(R)$.

It may be interesting to compare the results obtained at different levels of theory for the simplest system, for which very accurate calculations are available: the H_2 molecule. The table shows the dissociation energy of the different isotopic variants in cm^{-1} (the zero-point energy of vibrations is included).

	H_2	HD	D_2
Born–Oppenheimer	36112.2	36401.5	36745.6
Adiabatic	36118.0	36405.7	36748.3
Nonadiabatic	36114.7	36402.9	36746.2
Experiment	36113.6	36400.5	36744.2

The results show that the Born–Oppenheimer approximation is quite acceptable and there is (at least in this case) a fortuitous error compensation.

It may be noted that most of the (thermal) chemical reactions are also adiabatic in the sense that the system remains on the PES corresponding to a given (usually lowest) electronic state. However, there exist nonadiabatic reactions too, with a transition from one PES to another, usually at the configurations where the PESs corresponding to different electronic states go close to each other (e.g., "avoided crossings").

2.5. A note on the Born–Oppenheimer separation

It is to be emphasized that writing the wave function $\Psi(R,r)$ in the form $\Psi_e(r;R)\Psi_N(R)$ does not represent an approximation *by itself*: the approximation is in assuming that these functions are solutions of the separate electronic and nuclear Schrödinger equations (1.52) and (1.54). In fact, one may divide and multiply $\Psi(R,r)$ by any $\Psi_N(R)$ and write formally $\Psi_e(r;R) = \Psi(R,r)/\Psi_N(R)$. Moreover, if $\Psi(R,r)$ is normalized, then it can be presented as a product of the form $\Psi_e(r;R)\Psi_N(R)$, in which both the electronic wave

function Ψ_e and the nuclear wave function $\Psi_N(R)$ are separately normalized. (When calculating the norm of Ψ_e one has to integrate over the electronic coordinates only.)

To get a presentation of that type, we integrate $|\Psi(R,r)|^2$ over the *electronic* coordinates (including summations over the spins, $d\tau_i = dv_i d\sigma_i$), and define the function $q(R)$ as the result of this integration:

$$q(R) = \int |\Psi(R,r)|^2 d\tau_1 d\tau_2 \ldots d\tau_{N_e} \ . \tag{1.64}$$

Then the normalized electronic wave function for every nuclear configuration $\{R\}$ can be written as

$$\Psi_e(r;R) = \frac{\Psi(R,r)}{\sqrt{q(R)}} \tag{1.65}$$

and the nuclear wave function is simply

$$\Psi_N(R) = \sqrt{q(R)} \ . \tag{1.66}$$

Its normalization is a consequence of the assumption that the overall wave function $\Psi(R,r)$ is normalized. Unfortunately, there seems to be no direct connection between these functions and the functions obtained by solving the separate electronic and nuclear Schrödinger equations (1.52) and (1.54). (If there are identical fermions among the nuclei, then the wave function vanishes for the geometric configurations in which two of them are in the same point of the space and have the same spin. For such systems the previous definition should be extended by defining $\Psi_N(R) = 0$ in these points; then $\Psi_e(r;R)$ may be considered arbitrary in these points provided it remains finite.)

Notes

Section 1.1.

Most authors perform this transformation without discussing any details of derivation (e.g. [1–3]) or simply use the separation of the C.O.M motion taking it for granted (from classical mechanics —see Appendix II). The derivation based on the requirement of zero overall momentum may be (but is not claimed to be) original.

Section 1.2.

The derivation presented differs from that in [4] by the numbering of the particles only.

Section 1.3.

A derivation of this type is described concisely in e.g., [5, 6] in connection with the separation of the C.O.M. motion. This approach is discussed in detail in [7] for the special case of the He atom.

Sections 2.1.–2.4.

Essentially a standard presentation (e.g. [8]). The concept of the Born–Oppenheimer approximation first appeared [9] on qualitative arguments: if the motion of electrons and nuclei are considered as oscillators, then the frequencies assigned to them relate as the inverse square roots

of the masses, thus the electrons move with a frequency $\sqrt{\frac{m_N}{m_e}} \sim 100$ times larger than that of the nuclei. The Born–Oppenheimer separation was first introduced in [10]. There recurs in the literature the mistake that in this paper the accuracy of the separation of the electronic and nuclear motion was established only up to the order of the parameter $\kappa = \sqrt[4]{\frac{m_e}{m_N}} \sim \frac{1}{10}$. This is however, not the case; the parameter κ had been used in some involved series expansions devoted to establish the relative magnitude of electronic, vibrational and rotational energies; the accuracy of the separation of electronic and nuclear motions was not investigated. For the contemporary form of the Born–Oppenheimer approximation one usually refers to [11].

Section 2.5.
Present author's considerations.

Suggested readings

In [4] one can find a detailed discussion of using the chain rule for coordinate transformation as well as the "Podolsky trick" of replacing the classical Hamiltonian function with the quantum mechanical Hamiltonian if curvilinear coordinates are used.

Concerning the problems "beyond the Born–Oppenheimer approximation" we quote two survey papers [12,13] and mention the interesting attempt [14] to put forward a completely independent theory in which a "cluster-expansion" of the wave function is used to express the strong inter-dependence between the electronic and nuclear coordinates, without fixing the positions of the latter.

References

1 E. Fermi, *Notes on Quantum Mechanics*, University of Chicago Press, 1960.
2 N. Mott and I. Sneddon, *Wave Mechanics and its Applications*. Clarendon Press, Oxford 1948.
3 L.D. Landau and I. M. Lifshitz, *Kvantovaya Mekhanika*, Nauka, Moscow 1974.
4 P.R. Bunker and P. Jensen, *Molecular Symmetry and Spectroscopy*, NRC Research Press, Ottawa 1998.
5 L. Zülicke, *Quantenchemie*, Deutcher Verlag der Wissenschaften, Berlin 1973.
6 B.T. Suttcliffe *Fundamentals in Computational Quantum Chemistry*, in *Computational Techniques in Quantum Chemistry and Molecular Physics*, (ed. G.H.F. Dierksen, B.T. Suttcliffe and A. Veillard) Reidel, Dordrecht 1974.
7 H.A. Bethe, *Intermediate Quantum Mechanics*, Benjamin, New York 1964.
8 J.C. Slater, *Quantum Theory of Molecules and Solids vol. 1. Electronic Structure of Molecules*, McGraw-Hill, New York 1963.
9 M. Born and W. Heisenberg, Ann. Phys. **84**, 1 (1924)
10 M. Born and J.R. Oppenheimer, Ann. Phys. **84**, 457 (1927)
11 M. Born and K. Huang, *Dynamical Theory of Crystal Lattices*, Oxford University Press, London 1954, Appendix VIII.
12 T. Azumi and K. Matsuzaki, Photochem. Photobiol., **25**, 315 (1977)
13 M. Baer, Physics Reports, **358**, 75 (2002)
14 H.J. Monkhorst, Phys. Rev. **A36**, 1544 (1987)

Chapter 2

GENERAL THEOREMS AND PRINCIPLES

1. The variation principle

1.1. The Rayleigh quotient

For the solutions of the Schrödinger equations the expectation value (quantum mechanical average value) of the energy is equal to the eigenvalue of the Hamiltonian $\hat{H}$. In fact, the exact solutions are "energy eigenstates":

$$\hat{H}\Psi = E\Psi \ . \tag{2.1}$$

We multiply (2.1) with Ψ^* and integrate:

$$\langle\Psi|\hat{H}|\Psi\rangle \ = E\langle\Psi|\Psi\rangle \ . \tag{2.2}$$

Dividing both sides with $\langle\Psi|\Psi\rangle$ we obtain the energy as an *expectation value* ("Rayleigh quotient"):

$$E = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \ . \tag{2.3}$$

For approximate wave functions only the expectation value can be used because $\frac{\hat{H}\Psi}{\Psi} \neq const.$

1.2. The variation principle for the ground state

In the class of the wave functions satisfying the boundary conditions of the problem considered, the lowest possible expectation value E_0 of the energy E belongs to the exact ground state-wave function. Therefore, for any Ψ

$$E = \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \geq E_0 \ . \tag{2.4}$$

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

(The equality holds only for the exact ground-state solution.) Therefore an energy computed by using an approximate wave function gives an "upper limit" (upper bound) for the exact ground-state energy.

Proof.
The Hamiltonian $\hat{H}$ is Hermitian; consequently its eigenvectors Ψ_i form a complete orthonormalized set of functions and any wave function Ψ can be expanded as

$$\Psi = \sum_i c_i \Psi_i \ . \tag{2.5}$$

(If $\hat{H}$ has a continuous spectrum, then the sum should be replaced —or augmented—by an integration, and other complications arise; we shall not consider this possibility as it does not influence the essence.)

We calculate

$$\langle \Psi | \Psi \rangle = \langle \sum_i c_i \Psi_i | \sum_j c_j \Psi_j \rangle = \sum_{i,j} c_i^* c_j \langle \Psi_i | \Psi_j \rangle = \sum_{i,j} c_i^* c_j \delta_{ij} = \sum_i |c_i|^2 \tag{2.6}$$

and

$$\begin{aligned} \langle \Psi | \hat{H} | \Psi \rangle &= \langle \sum_i c_i \Psi_i | \hat{H} | \sum_j c_j \Psi_j \rangle = \sum_{i,j} c_i^* c_j \langle \Psi_i | \hat{H} | \Psi_j \rangle \\ &= \sum_{i,j} c_i^* c_j E_j \langle \Psi_i | \Psi_j \rangle = \sum_{i,j} c_i^* c_j E_j \delta_{ij} = \sum_i |c_i|^2 E_i \ . \end{aligned} \tag{2.7}$$

Therefore:

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\sum_i |c_i|^2 E_i}{\sum_i |c_i|^2} \ . \tag{2.8}$$

Because $E_i \geq E_0$, we get

$$E \geq \frac{\sum_i |c_i|^2 E_0}{\sum_i |c_i|^2} = E_0 \ . \tag{2.9}$$

Q.E.D.

The fact that the expectation value of the energy represents a minimum indicates that the energy calculated in that manner is usually much more accurate than the wave function from which it is determined. In fact, in the minimum the energy as the functional of the wave function must be stationary, i.e., it must have a vanishing first variation. Therefore, the error in the energy is quadratic with respect to the error in the wave function: if the deviation of the wave function from the exact one is proportional to some (small) parameter ε, then the expectation value of the energy will differ from the exact one by a quantity

proportional to ε^2. In practice this means that if we have determined the wave function expansion parameters with the accuracy of k decimal digits, than the energy is expected to be accurate up to about $2k$ digits, or even more.

1.3. The variation principle as an equivalent of the Schrödinger equation: A useful formulation of the variation principle

The Schrödinger equation can also be formulated as a variational problem: one has to find the wave function for which the energy (energy functional, expectation value of the Hamiltonian) is stationary:

$$\delta E = \delta \left[\frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \right] = 0 \tag{2.10}$$

The vanishing first variation of the energy and the Schrödinger equation are equivalent to each other—not only for the ground state but for excited states, as well. (Not sufficiently recognized.) If one uses a minimization procedure then the orthogonality to the states with lower energy should also be required. Exceptions are the lowest states of each symmetry species for which the orthogonality is satisfied automatically —they are essentially the ground state of their own species.

In (2.10) one may perform the variation of the fraction in a manner quite analogous to differentiation (for the analogy between variations and differentials we refer to Appendix III):

$$\delta E = \frac{\langle \delta\Psi | \hat{H} | \Psi \rangle \langle \Psi | \Psi \rangle - \langle \delta\Psi | \Psi \rangle \langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle^2} + \ c.c. \ = 0 \tag{2.11}$$

($c.c.$ = complex conjugate).

As $\delta\Psi$ is arbitrary, its phase is arbitrary as well (i.e., $\delta\Psi$ contains an arbitrary phase factor), therefore the expression written out explicitly and its complex conjugate should separately be equal to 0. After multiplying by $\langle \Psi | \Psi \rangle \neq 0$ (we consider only normalizable wave functions):

$$\langle \delta\Psi | \hat{H} | \Psi \rangle \underbrace{\frac{\langle \Psi | \Psi \rangle}{\langle \Psi | \Psi \rangle}}_{1} - \langle \delta\Psi | \Psi \rangle \underbrace{\frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}}_{E} = 0 \, . \tag{2.12}$$

That means

$$\langle \delta\Psi | \hat{H} - E | \Psi \rangle = 0 \, , \tag{2.13}$$

which is a very useful form of the variation principle that we shall use throughout. As $\delta\Psi$ is arbitrary if no restrictions are imposed on Ψ, condition (2.13) is

also fully equivalent to the Schrödinger equation

$$(\hat{H} - E)\Psi = 0 \,. \tag{2.14}$$

The form in (2.13) we have derived for the variation principle is applicable also in the cases when we do not use an arbitrary trial wave function Ψ in the variation procedure but approximate the exact wave function by a given type ("Ansatz") of variational wave function. In this case the variation $\delta\Psi$ is not fully arbitrary any more, but it must correspond to a variation of Ψ performed *within* the given class of wave functions. That means that the wave function $\Psi + \delta\Psi$ obtained after the variation must also belong to the given class, which puts restrictions on the possible variations $\delta\Psi$. In such cases (2.13) is not equivalent to the Schrödinger equation but permits to derive the respective variational equations corresponding to the given class of trial wave functions.

The use of Lagrangian multipliers

The equivalence of the variation principle with the Schrödinger equation can also be established by using the method of Lagrangian multipliers. The problem is then formulated in the following manner: let us request the functional (integral) $F = \langle\Psi|\hat{H}|\Psi\rangle$ to be stationary under the condition $\langle\Psi|\Psi\rangle = 1$. We form the auxiliary functional F' by adding to F the left-hand side of the auxiliary condition rewritten as

$$\langle\Psi|\Psi\rangle - 1 = 0 \tag{2.15}$$

multiplied by the undetermined Lagrangian multiplier λ:

$$F' = \langle\Psi|\hat{H}|\Psi\rangle + \lambda(\langle\Psi|\Psi\rangle - 1) \tag{2.16}$$

and request $\delta F'$ to vanish. One should join to the equation resulting from this variation also condition (2.15), or—which is the same—request F' to be stationary as a function of the scalar λ, as well. [From the requirement $\frac{\partial F'}{\partial\lambda} = 0$ one immediately recovers condition (2.15).] We get

$$\langle\delta\Psi|\hat{H}|\Psi\rangle + \lambda\langle\delta\Psi|\Psi\rangle + c.\,c. = 0 \,. \tag{2.17}$$

As $\delta\Psi$ is arbitrary, and also contains an arbitrary phase factor, it follows from (2.17) that

$$\hat{H}\Psi + \lambda\Psi = 0 \,. \tag{2.18}$$

We now multiply (2.18) by Ψ^* and integrate by taking into account (2.15), and obtain

$$\langle\Psi|\hat{H}|\Psi\rangle + \lambda = 0 \,, \tag{2.19}$$

i.e., $\lambda = -E$. Substituting this into (2.18) we have

$$\hat{H}\Psi - E\Psi = 0 \,. \tag{2.20}$$

Q.E.D.

1.4. Eckart's inequality

The variation principle tells us that the exact ground state has the absolute minimum of the energy, and Eckart's inequality shows that —once we are close enough to the solution—decreasing the energy of the trial function may be expected to improve the wave function, as well; in the limiting case, when the energy is approaching the exact value, then the wave function should also approach the exact solution.

For sake of simplicity, we shall assume all wave functions considered in this section to be normalized to 1.

Let us consider a trial wave function Ψ; it is a linear combination of the exact eigenfunctions of the Hamiltonian with the coefficients c_i:

$$\Psi = \sum_{i=0} c_i \Psi_i \tag{2.21}$$

and we have requested

$$\langle \Psi | \Psi \rangle = \sum_{i=0} |c_i|^2 = 1 \, . \tag{2.22}$$

We consider the overlap (scalar product, inner product) of the wave function Ψ and the exact ground-state solution Ψ_0

$$S = \langle \Psi_0 | \Psi \rangle = c_0 \tag{2.23}$$

where the last equality follows from the orthonormalization of the exact solutions Ψ_i. One may write

$$\begin{aligned} E &= \sum_{i=0} |c_i|^2 E_i = |S|^2 E_0 + \sum_{i=1} |c_i|^2 E_i \qquad (\text{as } E_i \geq E_1 \text{ if } i \geq 1) \\ &\geq |S|^2 E_0 + \sum_{i=1} |c_i|^2 E_1 = |S|^2 E_0 + E_1 \sum_{i=1} |c_i|^2 \, . \end{aligned} \tag{2.24}$$

As $\sum_{i=0} |c_i|^2 = 1$ and $|c_0|^2 = |S|^2$, we have obviously $\sum_{i=1} |c_i|^2 = 1 - |S|^2$, so

$$E \geq |S|^2 E_0 + (1 - |S|^2) E_1 \tag{2.25}$$

from which

$$|S|^2 \geq \frac{E_1 - E}{E_1 - E_0} \, . \tag{2.26}$$

Therefore, if $E \to E_0$, then $|S| \to 1$. ($|S| \leq 1$ according to its definition.) Q.E.D.

1.5. Excited states

The Hamiltonian is a Hermitian operator; its eigenvectors belonging to different eigenvalues are automatically orthogonal; those corresponding to the same (degenerate) eigenvalue can also be chosen orthogonal. However, as already noted in Section 1.3, when one searches for the eigenvector corresponding to an excited state, then one may meet the problem of the "variational collapse": the calculation is "trapped" by the ground-state solution, i.e. it converges back to the latter instead of yielding the required excited state. In such situations it may be necessary to introduce the explicit requirement of the orthogonality to the lower lying state(s) already determined. (We shall consider explicitly only one such state; the generalization to the case of several such ones is trivial.) The following considerations are applicable also in the case when the ground-state wave function considered is not the exact one but only some approximation to the latter.

Accordingly, we should solve the following problem. One has to find the wave functions $|\Psi\rangle$ having a stationary energy within the subspace of functions which are orthogonal to some (exact or approximate) ground-state wave function $|\Psi_0\rangle$. (It is convenient to use the "bra-ket" notations —see Appendix VII—in this section.) That means that we should request

$$(1-\hat{P})|\Psi\rangle = |\Psi\rangle \tag{2.27}$$

where

$$\hat{P} = |\Psi_0\rangle\langle\Psi_0| \tag{2.28}$$

is the projector on the normalized ground state wave function $|\Psi_0\rangle$.

As we are searching a wave function within the subspace orthogonal to $|\Psi_0\rangle$, the variation should not move $|\Psi\rangle$ out of this subspace, either. Consequently, one must consider only variations $|\delta\Psi\rangle$ that are also within the same subspace; this can be achieved by writing the most general permitted variation in the form $|\delta\Psi\rangle = (1-\hat{P})|\delta\Phi\rangle$ where $|\delta\Phi\rangle$ is fully arbitrary. Substituting this variation in the general form (2.13) of the variation principle, we get

$$\langle\delta\Phi|(1-\hat{P})(\hat{H}-E)|\Psi\rangle = 0\ . \tag{2.29}$$

Utilizing that $|\delta\Phi\rangle$ is fully arbitrary, this gives

$$(1-\hat{P})\hat{H}|\Psi\rangle = E(1-\hat{P})|\Psi\rangle\ . \tag{2.30}$$

This equation remains valid if one substitutes the restriction $(1-\hat{P})|\Psi\rangle = |\Psi\rangle$ on the right-hand side:

$$(1-\hat{P})\hat{H}|\Psi\rangle = E|\Psi\rangle\ . \tag{2.31}$$

In fact, by multiplying (2.31) with $(1-\hat{P})$ and utilizing the idempotency of this operator, we see that all solutions of (2.31) belonging to nonzero eigenvalues E are automatically orthogonal to $|\Psi_0\rangle$: $|\Psi\rangle = (1-\hat{P})|\Psi\rangle$.

Furthermore, one can utilize this equality once again, and insert a projector $(1-\hat{P})$ on the left-hand side of (2.31) and get the eigenvalue equation of a Hermitian operator:

$$(1-\hat{P})\hat{H}(1-\hat{P})|\Psi\rangle = E|\Psi\rangle \ , \tag{2.32}$$

which is essentially the eigenvalue equation of the Hamiltonian projected on the subspace of functions orthogonal to the ground state $|\Psi_0\rangle$. It has the property that the ground-state wave function $|\Psi_0\rangle$ is also its eigenvector with a zero eigenvalue. (Therefore, the required orthogonality is also guaranteed by the orthogonality of the eigenvectors of the Hermitian operators.)

The use of Lagrangian multipliers

The previous task can also be solved by using the method of Lagrangian multipliers. The problem is then formulated in the following manner: let us request the functional $F = \langle\Psi|\hat{H}|\Psi\rangle$ to be stationary under the conditions $\langle\Psi|\Psi\rangle = 1$ and $\langle\Psi|\Psi_0\rangle = 0$.

$\langle\Psi|\Psi\rangle$ and $\langle\Psi|\hat{H}|\Psi\rangle$ are necessarily real, but the condition $\langle\Psi|\Psi_0\rangle = 0$ in the general complex case represents two real conditions $Re\langle\Psi|\Psi_0\rangle = 0$ and $Im\langle\Psi|\Psi_0\rangle = 0$. Accordingly, we form the auxiliary functional F' by introducing three Lagrangian multipliers, λ, μ, and ν:

$$F' = \langle\Psi|\hat{H}|\Psi\rangle + \lambda(\langle\Psi|\Psi\rangle - 1) + \mu Re\langle\Psi|\Psi_0\rangle + \nu Im\langle\Psi|\Psi_0\rangle \ . \tag{2.33}$$

This can appropriately rewritten as follows. One has

$$Re\langle\Psi|\Psi_0\rangle = \frac{1}{2}(\langle\Psi|\Psi_0\rangle + \langle\Psi_0|\Psi\rangle) \tag{2.34}$$

and

$$Im\langle\Psi|\Psi_0\rangle = \frac{1}{2i}(\langle\Psi|\Psi_0\rangle - \langle\Psi_0|\Psi\rangle) \ , \tag{2.35}$$

We substitute these expressions into (2.33), collect the terms containing $\langle\Psi|\Psi_0\rangle$ and $\langle\Psi_0|\Psi\rangle$, introduce the notations $\tau = \frac{1}{2}(\mu - i\nu)$ and $\tau^* = \frac{1}{2}(\mu + i\nu)$ and get

$$F' = \langle\Psi|\hat{H}|\Psi\rangle + \lambda(\langle\Psi|\Psi\rangle - 1) + \tau\langle\Psi|\Psi_0\rangle + \tau^*\langle\Psi_0|\Psi\rangle \ . \tag{2.36}$$

We request $\delta F' = 0$, i.e.,

$$\delta F' = \langle\delta\Psi|\hat{H}|\Psi\rangle + \lambda\langle\delta\Psi|\Psi\rangle + \tau\langle\delta\Psi|\Psi_0\rangle + c.c. = 0 \ . \tag{2.37}$$

From this ($\delta\Psi$ is arbitrary, contains also an arbitrary phase factor) we get the system of equations

$$\begin{aligned} \hat{H}|\Psi\rangle + \lambda|\Psi\rangle + \tau|\Psi_0\rangle &= 0 \\ \langle\Psi|\Psi\rangle &= 1 \\ \langle\Psi|\Psi_0\rangle &= 0 \quad . \end{aligned} \tag{2.38}$$

By multiplying the first equation of (2.38) with $\langle\Psi|$ and $\langle\Psi_0|$, respectively, and utilizing the auxiliary conditions—the second and third equations of the system—we get

$$\lambda = -E \tag{2.39}$$

and

$$\tau = -\langle\Psi_0|\hat{H}|\Psi\rangle \ . \tag{2.40}$$

($|\Psi_0\rangle$ has been assumed normalized.) By substituting these values into the first equation of (2.38) we arrive at

$$\hat{H}|\Psi\rangle - |\Psi_0\rangle\langle\Psi_0|\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{2.41}$$

that is

$$(1 - \hat{P})\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{2.42}$$

We have obtained (2.31) that has already been discussed; it may be Hermitized in the same manner as described previously.

2. The Hellmann–Feynman theorem

When deriving the variation principle, we have considered arbitrary "mathematical" variations of the wave function for a given system (given Hamiltonian) in order to investigate the behavior of the energy functional under the slight changes of the wave function in the vicinity of the exact solution. Contrary to this, the Hellmann–Feynman theorem deals with the changes induced by the variations in the physical parameters of the system (parameters defining its Hamiltonian). Nevertheless, they have many aspects in common.

2.1. The differential Hellmann–Feynman theorem

Let us consider the case when some parameter α describing the system undergoes a change

$$\alpha \to \alpha + d\alpha \ . \tag{2.43}$$

Therefore, there will be a change in the Hamiltonian $\hat{H}$ and, as a consequence, in Ψ and E, as well:

$$\hat{H} \ \to \ \hat{H} + \frac{\partial\hat{H}}{\partial\alpha}d\alpha$$

$$\begin{aligned} \Psi &\rightarrow \Psi + \frac{\partial \Psi}{\partial \alpha} d\alpha \\ E &\rightarrow E + \frac{\partial E}{\partial \alpha} d\alpha \, . \end{aligned} \tag{2.44}$$

(We drop everywhere terms that are of higher order with respect to $d\alpha$.) The derivative of the energy is

$$\begin{aligned} \frac{\partial E}{\partial \alpha} &= \frac{\partial}{\partial \alpha} \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \\ &= \frac{\left(\langle \frac{\partial \Psi}{\partial \alpha} | \hat{H} | \Psi \rangle + \langle \Psi | \frac{\partial \hat{H}}{\partial \alpha} | \Psi \rangle + \langle \Psi | \hat{H} | \frac{\partial \Psi}{\partial \alpha} \rangle \right) \langle \Psi | \Psi \rangle}{\langle \Psi | \Psi \rangle^2} \\ &\quad - \frac{\left(\langle \frac{\partial \Psi}{\partial \alpha} | \Psi \rangle + \langle \Psi | \frac{\partial \Psi}{\partial \alpha} \rangle \right) \langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle^2} \\ &= \frac{1}{\langle \Psi | \Psi \rangle} \left\{ \langle \frac{\partial \Psi}{\partial \alpha} | \hat{H} | \Psi \rangle - E \langle \frac{\partial \Psi}{\partial \alpha} | \Psi \rangle + \langle \Psi | \hat{H} | \frac{\partial \Psi}{\partial \alpha} \rangle \right. \\ &\qquad \left. - E \langle \Psi | \frac{\partial \Psi}{\partial \alpha} \rangle + \langle \Psi | \frac{\partial \hat{H}}{\partial \alpha} | \Psi \rangle \right\} . \end{aligned} \tag{2.45}$$

The first four terms of this expression vanish if Ψ is a solution of the Schrödinger equation because in that case they are equal to $\langle \frac{\partial \Psi}{\partial \alpha} | \hat{H} - E | \Psi \rangle + c.c. = 0$. (This may also be formulated by stating that the variation $\delta \Psi = \frac{\partial \Psi}{\partial \alpha} d\alpha$ is one of the permitted "mathematical" variations for which the variation principle does hold.) Therefore we get:

$$\frac{\partial E}{\partial \alpha} = \frac{\langle \Psi | \frac{\partial \hat{H}}{\partial \alpha} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \, . \tag{2.46}$$

This is the "differential Hellmann–Feynman theorem" (often called simply *the* Hellmann–Feynman theorem). It indicates that for the calculation of the derivative of the energy one does not need the derivative of the wave function; the energy derivative can simply be calculated as the expectation value of the derivative of the Hamiltonian —as it would also be suggested by a first-order perturbation theory.

If the parameter α is one of the nuclear coordinates X_A of the molecule treated in the Born–Oppenheimer framework, then the Hellmann–Feynman theorem indicates that the force acting along that coordinate is

$$F_{X_A} = -\frac{\partial E}{\partial X_A} = -\frac{\langle\Psi|\frac{\partial \hat{H}}{\partial X_A}|\Psi\rangle}{\langle\Psi|\Psi\rangle}\,, \tag{2.47}$$

with Ψ being the electronic wave function and $\hat{H}$ the total electronic Hamiltonian (including nuclear–nuclear repulsion).

In the electronic Hamiltonian only the terms describing the Coulomb interactions of the electron–nuclear and nuclear–nuclear type depend on the nuclear coordinates. Therefore, in the framework of the Born–Oppenheimer approximation the forces acting on a given nucleus are determined by the electrostatic fields of the other nuclei and of the charge distribution of the electronic cloud, and they do not contain any explicit two-electron terms. Accordingly, in order to obtain the forces, in principle one does not need the electronic wave function, only the distribution of the electronic charge density in the space. The situation, however, is not so advantageous in practice, as we do not work with the exact wave functions and do not know the charge distributions corresponding to them. (Note that the previous conclusion regarding the central significance of the electronic density has nothing in common with the theorems on which the so-called "density functional theory"—DFT—is based.)

It may be worthwhile to show explicitly that the forces acting on a nucleus in a molecule are *exactly* the classical electrostatic forces originating from the actual charge distribution of the molecule caused by the nuclei and the electronic cloud.

Writing out (2.47) for all three directions, and assuming for simplicity that the electronic wave function Ψ is normalized ($\langle\Psi|\Psi\rangle = 1$), we can combine the equations obtained in a single vectorial equation:

$$\vec{F}_A = -\langle\Psi|\nabla_A \hat{H}|\Psi\rangle\,. \tag{2.48}$$

Here $\vec{F}_A$ is the total force acting on nucleus A. It is a sum of two components, $\vec{F}_{A,NN}$ originating from the internuclear repulsion potential energy term, V_{NN}, of the Hamiltonian, and $\vec{F}_{A,eN}$, originating from the electron–nuclear attraction term, V_{eN}.

The nuclear–nuclear repulsion energy V_{NN} does not depend on the electronic wave function, so $\nabla_A V_{NN}$ can be brought out of the integral, and the latter reduces to a factor one. Thus, the nuclear part of the force is simply

$$\vec{F}_{A,NN} = -\nabla_A \sum_{\alpha<\beta}^{N_N} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} = -\nabla_A \sum_{\substack{B=1\\(B\neq A)}}^{N_N} \frac{Z_A Z_B}{R_{AB}}\,. \tag{2.49}$$

(We use here the Born–Oppenheimer Hamiltonian written in atomic units—see Appendix VI; in the second equality in (2.49) we have taken into account that both α and β may equal A.)

We compute in a standard manner

$$\begin{aligned}
\frac{\partial}{\partial X_A}\frac{1}{R_{AB}} &= -\frac{1}{R_{AB}^2}\frac{\partial R_{AB}}{\partial X_A} \\
&= -\frac{1}{R_{AB}^2}\frac{\partial}{\partial X_A}\sqrt{(X_B-X_A)^2+(Y_B-Y_A)^2+(Z_B-Z_A)^2} \\
&= \frac{1}{R_{AB}^2}\frac{X_B-X_A}{\sqrt{(X_B-X_A)^2+(Y_B-Y_A)^2+(Z_B-Z_A)^2}} \\
&= \frac{1}{R_{AB}^3}(X_B-X_A)
\end{aligned} \tag{2.50}$$

Determining analogously the derivatives for the other two directions, one gets in vectorial form

$$\vec{F}_{A,NN} = -\sum_{\substack{B=1\\(B\neq A)}}^{N_N}\frac{Z_A Z_B}{R_{AB}^3}\vec{R}_{AB} = -\sum_{\substack{B=1\\(B\neq A)}}^{N_N}\frac{Z_A Z_B}{R_{AB}^2}\vec{R}_{AB}^{\,0}\,, \tag{2.51}$$

where $\vec{R}_{AB}^{\,0}$ is the unit vector directed from nucleus A toward nucleus B.

The derivation is quite analogous for the electron–nuclear component of the force, but the summation is now for the electrons having charge –1 and all quantities considered are under the integral:

$$\vec{F}_{A,eN} = -\langle\Psi|\nabla_A\sum_{\alpha=1}^{N_N}\sum_{i=1}^{N_e}\frac{-Z_\alpha}{r_{\alpha i}}|\Psi\rangle = \langle\Psi|\sum_{i=1}^{N_e}\frac{Z_A}{r_{Ai}^2}\vec{r}_{Ai}^{\,0}|\Psi\rangle\,. \tag{2.52}$$

The electrons are identical and the wave function Ψ is assumed to be properly antisymmetrized, so the contribution of each electron in (2.52) should be equal; we have, therefore:

$$\begin{aligned}
\vec{F}_{A,eN} &= N_e\,\langle\Psi|\frac{Z_A}{r_{A1}^2}\vec{r}_{A1}^{\,0}|\Psi\rangle \\
&= N_e\int dv_1\left[\sum_{\sigma_1}\int\int\ldots\int d\tau_2 d\tau_3\ldots d\tau_{N_e}|\Psi(1,2,\ldots,N_e)|^2\right]\frac{Z_A}{r_{A1}^2}\vec{r}_{A1}^{\,0}\,.
\end{aligned} \tag{2.53}$$

Here we denote by $\int d\tau_i$ the integration over the coordinates of electron i, including summation over the spin of that electron, whereas notation $\int dv_1$ denotes the integration over the spatial coordinates of electron 1 and $\sum_{\sigma_1}$ denotes the summation over its spin.

It is known that $|\Psi(1,2,\ldots,N_e)|^2$ gives the probability density of the system to have the coordinates $\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_{N_e}$ and spins $\sigma_1,\sigma_2,\ldots,\sigma_{N_e}$; therefore the quantity in braces in (2.53) —which is integrated over the coordinates of

electrons $2, 3, \ldots, N_e$ and summed over the spin of all electrons—gives the *spatial* probability density of finding electron 1 around the point $\vec{r}_1$ with an arbitrary spin and admitting arbitrary spatial and spin coordinates for all the other electrons. As the electrons are identical, the other electrons will have the same spatial probability distribution. Therefore, multiplying the function in braces by the number of electrons N_e, we get the spatial density $\rho(\vec{r}_1)$ of the electron distribution ("electron density"), and (2.53) can be rewritten as

$$\vec{F}_{A,eN} = \int dv_1 \rho(\vec{r}_1) \frac{Z_A}{r_{A1}^2} \vec{r}_{A1}^{\,0} \,, \tag{2.54}$$

which is really the integral of the electrostatic force acting on nucleus A and caused by the negative charge distribution $-\rho(\vec{r}_1)$. [The elementary negative charge $-\rho(\vec{r}_1)dv_1$ gives rise to the elementary force $\frac{dv_1\rho(\vec{r}_1)Z_A}{r_{A1}^2}\vec{r}_{A1}^{\,0}$; note the opposite sign as compared with (2.51).]

The differential Hellmann–Feynman theorem is fulfilled not only for the exact wave functions but also for some approximate ones determined by using the variation principle. In fact, (2.45) can be rewritten as

$$\frac{\partial E}{\partial \alpha} = \frac{1}{\langle \Psi | \Psi \rangle} \left\{ \langle \frac{\partial \Psi}{\partial \alpha} | \hat{H} - E | \Psi \rangle + \langle \Psi | \hat{H} - E | \frac{\partial \Psi}{\partial \alpha} \rangle + \langle \Psi | \frac{\partial \hat{H}}{\partial \alpha} | \Psi \rangle \right\}, \tag{2.55}$$

i.e., for the fulfillment of the Hellmann–Feynman theorem it is sufficient if the formulation in (2.13) of the variation principle is satisfied for the variation $\delta\Psi = \frac{\partial \Psi}{\partial \alpha} d\alpha$ of the wave function. Thus, for instance, the Hellmann–Feynman theorem holds for the Hartree–Fock wave function obtained by solving the respective integro-differential Hartree–Fock equations, but not if a finite basis set is used (Hartree–Fock–Roothaan method), as it is usually applied in the molecular calculations. In the latter case the basis functions "follow" the nuclei and a change of nuclear geometry causes the basis orbitals to be also shifted. In the framework of the Hartree–Fock–Roothaan calculations the variational problem is solved for a given (fixed) geometry and basis set. At the same time, the variation of a basis orbital caused by a shift of its center cannot be, in general, expanded in the original finite basis set used in that calculation. Therefore, an infinitesimal change of the geometry generates such variations of the resulting many-electron wave function, which do not belong to the class of the permitted variations for which the Hartree–Fock–Roothaan equations are derived. (Similar considerations hold for geometry dependence of basis orbitals in the framework of other variational—but also perturbational etc.—methods applying finite basis sets.) Consequently, the Hellmann–Feynman theorem would hold in the finite basis case only if the basis orbitals were not fixed to the nuclei but the positions of their centers would be optimized as independent variational parameters. This is the only case in which the variations of the

many-electron wave functions generated by the derivatives of the basis orbitals are also among the permitted "mathematical" variations.

The terms of the energy variation connected with the fact that the basis orbitals "follow" the nuclei (these terms are sometimes called "wave function forces") make much more difficult the calculation of the energy gradients, and thus the optimization of the molecular geometries, representing a very important practical field of applying quantum chemistry. As a consequence, one needs also the gradients of the two-electron integrals in order to calculate the forces—see Section 7 of Chapter 6 for details. (For non-variational, e.g. perturbational, methods, the gradients of the wave functions should be determined, as well.)

2.2. The integral Hellmann–Feynman theorem

The effect of *finite* changes of the parameter α can be, in general, obtained by integration (e.g., calculating potential from the force). Assuming that the normalization

$$\langle\Psi|\Psi\rangle = 1 \tag{2.56}$$

does not depend on α and applying the differential Hellmann–Feynman theorem, we get its *"integrated form"*:

$$\Delta E = E(\alpha_1) - E(\alpha_0) = \int_{\alpha_0}^{\alpha_1} \left(\frac{\partial E}{\partial \alpha}\right) d\alpha = \int_{\alpha_0}^{\alpha_1} \langle\Psi(\alpha)|\frac{\partial \hat{H}}{\partial \alpha}|\Psi(\alpha)\rangle d\alpha \ . \tag{2.57}$$

ΔE can also be determined by using the quantities corresponding to the limiting values (α_0 and α_1) of α only. Let us denote the different quantities corresponding to α_0 as Ψ_0, $\hat{H}_0$, and E_0, those corresponding to α_1 as Ψ_1, $\hat{H}_1$, and E_1, respectively, and let ΔE again denote the difference $E(\alpha_1) - E(\alpha_0)$. The Schrödinger equations corresponding to these two points are

$$\begin{aligned} \hat{H}_1\Psi_1 &= E_1\Psi_1 \ ; \\ \hat{H}_0\Psi_0 &= E_0\Psi_0 \ . \end{aligned} \tag{2.58}$$

We multiply the first equation of (2.58) with Ψ_0^* and the second with Ψ_1^* and integrate both equations:

$$\begin{aligned} \langle\Psi_0|\hat{H}_1|\Psi_1\rangle &= E_1\langle\Psi_0|\Psi_1\rangle \\ \langle\Psi_1|\hat{H}_0|\Psi_0\rangle &= E_0\langle\Psi_1|\Psi_0\rangle \end{aligned} \tag{2.59}$$

Taking the complex conjugate of the second equation and extracting from the first one, we get

$$\langle\Psi_0|\hat{H}_1|\Psi_1\rangle - \langle\Psi_0|\hat{H}_0|\Psi_1\rangle = E_1\langle\Psi_0|\Psi_1\rangle - E_0\langle\Psi_0|\Psi_1\rangle \tag{2.60}$$

i.e.,

$$\Delta E = \frac{\langle \Psi_0 | \hat{H}_1 - \hat{H}_0 | \Psi_1 \rangle}{\langle \Psi_0 | \Psi_1 \rangle} \ . \tag{2.61}$$

This is the integral Hellmann–Feynman theorem. (It has been derived for the exact wave functions, therefore it should be applied for approximate ones with proper care.)

The differential Hellmann–Feynman theorem can also be derived from the integral one, by considering $\hat{H}_1 = \hat{H}_0 + \delta\hat{H} = \hat{H}_0 + \frac{\partial \hat{H}}{\partial \alpha} d\alpha$, $\Psi_0 = \Psi$, $\Psi_1 = \Psi + \delta\Psi = \Psi + \frac{\partial \Psi}{\partial \alpha} d\alpha$; substituting in the expressions of the integral Hellmann–Feynman theorem and conserving terms of first order only. ($\langle \Psi | \delta\Psi \rangle = 0$ may be assumed.)

3. The virial theorem in quantum mechanics

3.1. Time dependence of a physical quantity

Let us consider the quantum mechanical expectation value of the physical quantity A:

$$\langle A \rangle = \langle \Psi | \hat{A} | \Psi \rangle \ , \tag{2.62}$$

assuming $\langle \Psi | \Psi \rangle = 1$. The time derivative of $\langle A \rangle$ is

$$\frac{d}{dt}\langle A \rangle = \langle \frac{\partial \Psi}{\partial t} | \hat{A} | \Psi \rangle + \langle \Psi | \frac{\partial \hat{A}}{\partial t} | \Psi \rangle + \langle \Psi | \hat{A} | \frac{\partial \Psi}{\partial t} \rangle \ . \tag{2.63}$$

The time-dependent Schrödinger equation is

$$i\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi \tag{2.64}$$

or

$$\frac{\partial \Psi}{\partial t} = -\frac{i}{\hbar}\hat{H}\Psi \ . \tag{2.65}$$

Substituting into (2.63) we have:

$$\frac{d}{dt}\langle A \rangle = \langle \Psi | \frac{\partial \hat{A}}{\partial t} | \Psi \rangle + \frac{i}{\hbar}\langle \hat{H}\Psi | \hat{A} | \Psi \rangle - \frac{i}{\hbar}\langle \Psi | \hat{A} | \hat{H}\Psi \rangle \ . \tag{2.66}$$

In the second term the sign "+" appeared because the coefficient has been in the "bra," so we had to perform a complex conjugation. We can transform this term further on the basis of the following considerations. The Hermitian adjoint $\hat{L}^\dagger$ of an operator $\hat{L}$ is defined as the operator for which the equality $\langle \hat{L}^\dagger \psi | \varphi \rangle = \langle \psi | \hat{L}\varphi \rangle$ holds for every function ψ and φ. The Hamiltonian $\hat{H}$ *is known* to be Hermitian (self-adjoint), i.e., $\hat{H} = \hat{H}^\dagger$. Consequently, we may move it from the "bra" to the "ket" (note that $\langle \psi | \hat{L} | \varphi \rangle \equiv \langle \psi | \hat{L}\varphi \rangle$):

$$\frac{d}{dt}\langle A \rangle = \langle \Psi | \frac{\partial \hat{A}}{\partial t} | \Psi \rangle + \frac{i}{\hbar}\langle \Psi | \hat{H}\hat{A} - \hat{A}\hat{H} | \Psi \rangle \tag{2.67}$$

i.e.,

$$\frac{d}{dt}\langle A\rangle = \langle\frac{\partial\hat{A}}{\partial t}\rangle + \frac{i}{\hbar}\langle[\hat{H};\hat{A}]\rangle \tag{2.68}$$

where

$$[\hat{H};\hat{A}] = \hat{H}\hat{A} - \hat{A}\hat{H} \tag{2.69}$$

is the "commutator" of operators $\hat{H}$ and $\hat{A}$.

If $\hat{A}$ does not depend on the time explicitly and commutes with the Hamiltonian (i.e., $[\hat{H};\hat{A}] = \hat{H}\hat{A} - \hat{A}\hat{H} = 0$) then $\langle A\rangle = const$, that is A is conserved (it is an integral of motion).

(Note that forming a commutator is a distributive operation, e.g., $[\hat{A};(\hat{B}+\hat{C})] = \hat{A}(\hat{B}+\hat{C}) - (\hat{B}+\hat{C})\hat{A} = \hat{A}\hat{B} + \hat{A}\hat{C} - \hat{B}\hat{A} - \hat{C}\hat{A} = [\hat{A};\hat{B}] + [\hat{A};\hat{C}]$.)

3.2. The virial theorem

We shall consider stationary states only and utilize that the expectation value of any physical quantity must be constant in a stationary state. (The virial theorem in classical mechanics—see Appendix V—applies to finite motions and in quantum mechanics to stationary states; these are in some sense counterparts of each other in the two theories.) Accordingly, one may consider

$$\frac{d}{dt}\langle\sum_i \vec{r}_i\hat{\vec{p}}_i\rangle = 0\,. \tag{2.70}$$

As $\frac{\partial}{\partial t}(\vec{r}_i\hat{\vec{p}}_i) = 0$, this means that

$$\langle\Psi|\,[\,\hat{H};\ \sum_i \vec{r}_i\hat{\vec{p}}_i\,]\,|\Psi\rangle = 0\,. \tag{2.71}$$

Furthermore, $\hat{\vec{p}}_i = \frac{\hbar}{i}\nabla_i$, and substituting the expression of $\hat{H}$ assuming the system to be closed, we get (omitting the common factor $\frac{\hbar}{i}$)

$$\langle\Psi|\,[\,\sum_j \frac{-\hbar^2}{2m_j}\Delta_j + V;\ \sum_i \vec{r}_i\nabla_i\,]\,|\Psi\rangle = 0\,. \tag{2.72}$$

As $[\Delta_j;\ \vec{r}_i\nabla_i] = 0$ for $j \neq i$, only one sum is to be conserved:

$$\sum_i \Big\{\langle\Psi|\,[\,\frac{-\hbar^2}{2m_i}\Delta_i;\ \vec{r}_i\nabla_i\,]\,|\Psi\rangle + \langle\Psi|\,[V;\ \vec{r}_i\nabla_i]\,|\Psi\rangle\Big\} = 0\,. \tag{2.73}$$

One has to take into account that $\frac{-\hbar^2}{2m_i}\Delta_i = \frac{1}{2m_i}\hat{\vec{p}}_i^{\,2}$ and the different components of the momentum commute with each other as well as with the coordinates

corresponding to orthogonal directions. Therefore, the commutator in the first term reduces to

$$[\frac{-\hbar^2}{2m_i}\Delta_i;\ \vec{r}_i\nabla_i] = \frac{-\hbar^2}{2m_i}\Big([\frac{\partial^2}{\partial x_i^2};\ x_i\frac{\partial}{\partial x_i}] + [\frac{\partial^2}{\partial y_i^2};\ y_i\frac{\partial}{\partial y_i}] + [\frac{\partial^2}{\partial z_i^2};\ z_i\frac{\partial}{\partial z_i}]\Big). \tag{2.74}$$

One has, using formulae for second derivatives of a product:

$$\begin{aligned}[\frac{\partial^2}{\partial x_i^2};\ x_i\frac{\partial}{\partial x_i}]\Psi &= \frac{\partial^2}{\partial x_i^2}x_i\frac{\partial\Psi}{\partial x_i} - x_i\frac{\partial}{\partial x_i}\frac{\partial^2\Psi}{\partial x_i^2} \\ &= 2\frac{\partial^2\Psi}{\partial x_i^2} + x_i\frac{\partial^3\Psi}{\partial x_i^3} - x_i\frac{\partial^3\Psi}{\partial x_i^3} = 2\frac{\partial^2\Psi}{\partial x_i^2}\end{aligned} \tag{2.75}$$

because $\frac{\partial^2}{\partial x_i^2}x_i = 0$, and $\frac{\partial}{\partial x_i}x_i = 1$. Performing analogous manipulations along the directions y and z and expanding the second commutator as

$$\begin{aligned}[V;\ \vec{r}_i\nabla_i]\,\Psi = V\vec{r}_i\nabla_i\Psi - \vec{r}_i\nabla_i(V\Psi) &= V\vec{r}_i\nabla_i\Psi - \vec{r}_i(\nabla_i V)\Psi - \vec{r}_i V\nabla_i\Psi \\ &= -\vec{r}_i(\nabla_i V)\Psi\end{aligned} \tag{2.76}$$

we arrive at

$$\sum_i\Big\{\frac{-\hbar^2}{2m_i}\langle\Psi|2\Delta_i|\Psi\rangle - \langle\Psi|\vec{r}_i(\nabla_i V)|\Psi\rangle\Big\} = 0\,. \tag{2.77}$$

The sum of first terms is twice the kinetic energy of the system, whereas—if the potential V is a homogeneous function of order k—the second transforms according to Euler's theorem (Appendix IV) as

$$\sum_i\langle\Psi|\vec{r}_i(\nabla_i V)|\Psi\rangle = \langle\Psi|\sum_i\vec{r}_i\frac{\partial V}{\partial\vec{r}_i}|\Psi\rangle = \langle\Psi|kV|\Psi\rangle\,. \tag{2.78}$$

Therefore, analogously to the the case of the classical mechanics (Appendix V) we have

$$2\langle T\rangle - k\langle V\rangle = 0\,. \tag{2.79}$$

For atoms and molecules (Coulombic interaction, $k = -1$) the virial theorem

$$\langle V\rangle = -2\langle T\rangle; \qquad E = -\langle T\rangle = \frac{\langle V\rangle}{2} < 0 \tag{2.80}$$

holds for the internal motion of the particles—the motion of the center of mass is described by a plane wave, for which the derivation does not apply (but it is of no interest, either). We have assumed in the derivation that the wave function is normalized; that means that the virial theorem holds for the bonded states i.e., for negative total energy. (This is in agreement with the fact that the expectation value of the kinetic energy is necessarily positive, as the kinetic energy can expressed through the squares of linear momenta.)

3.3. Scaling—a connection with the variation principle

Let us consider some approximate wave function

$$\Psi = \Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) \tag{2.81}$$

of a system, and let us replace all the $\vec{r}_i$ by $\vec{\rho}_i = \eta\vec{r}_i$. (It is assumed that the wave function considered is the total wave function depending on both electronic and nuclear coordinates and that the transformation is performed for both types of coordinates. Alternatively, one may consider the electronic wave function in the framework of the Born–Oppenheimer approximation for the case of a single atom with the nucleus in the origin.)

In terms of the new coordinates $\vec{\rho}_i$ one can write the following *normalized* wave function

$$\Psi_\eta = \eta^{\frac{3N}{2}}\Psi(\eta\vec{r}_1, \eta\vec{r}_2, \ldots, \eta\vec{r}_N) = \eta^{\frac{3N}{2}}\Psi(\vec{\rho}_1, \vec{\rho}_2, \ldots, \vec{\rho}_N)\ , \tag{2.82}$$

provided that $\Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$ is normalized. In fact, as $\vec{r}_i = \frac{1}{\eta}\vec{\rho}_i$,

$$dv_i = dx_i dy_i dz_i = \frac{1}{\eta^3} d\rho_{x_i} d\rho_{y_i} d\rho_{z_i} = \frac{1}{\eta^3} dv_{\rho_i}\ , \tag{2.83}$$

and we have

$$\begin{aligned}
&\int\ldots\int |\Psi_\eta|^2 dv_1 dv_2 \ldots dv_N \\
&\quad = \eta^{3N}\int\ldots\int |\Psi(\vec{\rho}_1, \vec{\rho}_2, \ldots, \vec{\rho}_N)|^2 dv_1 dv_2 \ldots dv_N \\
&\quad = \int\ldots\int |\Psi(\vec{\rho}_1, \vec{\rho}_2, \ldots, \vec{\rho}_N)|^2 dv_{\rho_1} dv_{\rho_2} \ldots dv_{\rho_N} \\
&\quad = \int\ldots\int |\Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)|^2 dv_1 dv_2 \ldots dv_N = 1\ ,
\end{aligned} \tag{2.84}$$

where it has been taken into account that the value of an integral does not depend on the notation of integration variables.

Because $\frac{1}{r} = \frac{\eta}{\rho}$, we obtain

$$\langle\Psi_\eta|\frac{1}{r}|\Psi_\eta\rangle = \eta\,\langle\Psi_\eta|\frac{1}{\rho}|\Psi_\eta\rangle\ . \tag{2.85}$$

When calculating the integral $\langle\Psi_\eta|\frac{1}{\rho}|\Psi_\eta\rangle$—and also $\langle\Psi_\eta|\frac{\partial^2}{\partial\rho_x^2}|\Psi_\eta\rangle$ as follows—the factors $\eta^{\frac{3N}{2}}$ cancel when recalculating the volume elements, analogously as was the case for the normalization integral, and in the same manner as there we may return to the original notations $\vec{r}_i$—this correspond to the case $\eta = 1$. Therefore, (2.85) can be written as

$$V(\eta) = \eta V\Big|_{\eta=1} = \eta V(1)\ . \tag{2.86}$$

Similarly, because

$$\frac{\partial}{\partial x} = \frac{\partial}{\partial \rho_x}\frac{d\rho_x}{dx} = \eta\frac{\partial}{\partial \rho_x}\ ; \qquad \frac{\partial^2}{\partial x^2} = \eta^2\frac{\partial^2}{\partial \rho_x^2} \tag{2.87}$$

we get

$$\langle\Psi_\eta|\frac{\partial^2}{\partial x^2}|\Psi_\eta\rangle = \eta^2\langle\Psi_\eta|\frac{\partial^2}{\partial \rho_x^2}|\Psi_\eta\rangle\ , \tag{2.88}$$

i.e.,

$$T(\eta) = \eta^2 T\Big|_{\eta=1} = \eta^2 T(1)\ . \tag{2.89}$$

In this way we obtain the energy as a function of η:

$$E(\eta) = T(\eta) + V(\eta) = \eta^2 T(1) + \eta V(1)\ . \tag{2.90}$$

Now we optimize according to η, considering it as a variational parameter:

$$\frac{dE}{d\eta} = 2\eta T(1) + V(1) = 0\ , \tag{2.91}$$

from which one obtains

$$\eta_{opt} = -\frac{V(1)}{2T(1)}\ . \tag{2.92}$$

We get, therefore,

$$T(\eta_{opt}) = \left[-\frac{V(1)}{2T(1)}\right]^2 T(1) = \frac{V(1)^2}{4T(1)} \tag{2.93}$$

$$V(\eta_{opt}) = -\frac{V(1)}{2T(1)}V(1) = -\frac{V(1)^2}{2T(1)} \tag{2.94}$$

and, accordingly

$$E_{opt} = -\frac{V(1)^2}{4T(1)}\ . \tag{2.95}$$

After performing the scaling the virial theorem is fulfilled:

$$\frac{V(\eta_{opt})}{T(\eta_{opt})} = -2\ . \tag{2.96}$$

3.4. The virial theorem in the Born–Oppenheimer approximation

For the case of fixed nuclei (the Born–Oppenheimer approximation), the virial theorem holds only for the equilibrium configuration of the nuclei and other stationary points of the PES. (The nuclear–nuclear repulsion is to be included in V.) This can be qualitatively (classically) explained as follows. The

force acting on the αth nucleus in the framework of the Born–Oppenheimer approximation is $\vec{F}_\alpha = -\frac{\partial E(R)}{\partial \vec{R}_\alpha}$. The nuclei would remain fixed if there were an external force of this value but of opposite direction to keep them in their position; however, then one has to take into account also the virial of the external forces (c.f. Appendix V), that is, one has in the Born–Oppenheimer approximation

$$2\langle T\rangle + \langle V\rangle + \sum_\alpha \vec{R}_\alpha \frac{\partial E(R)}{\partial \vec{R}_\alpha} = 0 \tag{2.97}$$

Accordingly, the virial theorem is fulfilled when all $\frac{\partial E(R)}{\partial \vec{R}_\alpha} = 0$: in the stationary points on the potential surface and in the case when the atoms are at infinite separation.

Let us turn to the quantum mechanical derivation of this result. Similarly to the general case, the stationary character of the wave function leads again to the equality

$$\frac{d}{dt}\langle \sum_i \vec{r}_i \hat{\vec{p}}_i\rangle = 0 \,, \tag{2.98}$$

but now the sum is over the electrons only. Exactly in the same manner as described previously, (2.98) can be transformed to

$$2\langle\Psi|\sum_i -\frac{\hbar^2}{2m_e}\Delta_i|\Psi\rangle - \langle\Psi|\sum_i \vec{r}_i \frac{\partial V}{\partial \vec{r}_i}|\Psi\rangle = 0 \,, \tag{2.99}$$

where we have taken into account that all m_is are equal with the electron mass m_e.

Analogous to the general case, the first term on the left-hand side of (2.99) is twice the kinetic energy of the electrons:

$$2\langle\Psi|\sum_i -\frac{\hbar^2}{2m_e}\,\Delta_i|\Psi\rangle = 2\langle T_e\rangle \,. \tag{2.100}$$

However, because the nuclei are considered *fixed* in the Born–Oppenheimer approximation, the general derivation is not directly applicable to the second term. We shall discuss this term in detail.

The potential energy of the molecular system consist of three components: electron–electron repulsion, electron–nuclear attraction, and nuclear–nuclear repulsion:

$$V = V_{ee} + V_{eN} + V_{NN} \,. \tag{2.101}$$

Accordingly, the expectation value $\langle\Psi|\sum_i \vec{r}_i \frac{\partial V}{\partial \vec{r}_i}|\Psi\rangle$ also becomes a sum of three terms; we shall consider them separately.

The potential energy V_{ee} of the interelectronic repulsion is a homogeneous function of the electronic coordinates; therefore the previous derivation applies for it:

$$-\langle\Psi|\sum_i \vec{r}_i \frac{\partial V_{ee}}{\partial \vec{r}_i}|\Psi\rangle = -k\langle V_{ee}\rangle = \langle V_{ee}\rangle \tag{2.102}$$

($k = -1$ in the Coulombic case). The energy V_{NN} of the internuclear repulsion does not depend at all on the electronic coordinates, the derivatives $\frac{\partial V_{NN}}{\partial \vec{r}_i}$ vanish, so it does not enter the expectation value considered. The electron–nuclear attraction V_{eN} depends on the distances $r_{i\alpha} = |\vec{r}_i - \vec{R}_\alpha|$ between the individual electrons and nuclei, and it does not represent a homogeneous function of the electronic coordinates in the case of fixed nuclei. (In the general case the coordinates of all particles were treated on equal footing; therefore we could consider the potential energy as a homogeneous function of *all* coordinates simultaneously. This is not possible in the framework of the Born–Oppenheimer approximation in which the nuclear coordinates are fixed.)

In order to illustrate the situation, let us consider the energy of interaction between the ith electron and the αth nucleus:

$$V_{i\alpha} = -\frac{Z_\alpha}{r_{i\alpha}} = \frac{-Z_\alpha}{\sqrt{(x_i - X_\alpha)^2 + (y_i - Y_\alpha)^2 + (z_i - Z_\alpha)^2}} \,. \tag{2.103}$$

By differentiating (2.103) we get

$$\frac{\partial V_{i\alpha}}{\partial x_i} = -Z_\alpha \frac{-(x_i - X_\alpha)}{[(x_i - X_\alpha)^2 + (y_i - Y_\alpha)^2 + (z_i - Z_\alpha)^2]^{3/2}} \tag{2.104}$$

and analogous expressions can be obtained for $\frac{\partial V_{i\alpha}}{\partial y_i}$ and $\frac{\partial V_{i\alpha}}{\partial z_i}$.

One can see by inspection that the sum $x_i \frac{\partial V_{i\alpha}}{\partial x_i} + y_i \frac{\partial V_{i\alpha}}{\partial y_i} + z_i \frac{\partial V_{i\alpha}}{\partial z_i}$ cannot be brought to the form $kV_{i\alpha}$. Such a transformation can actually be performed if one considers the expression

$$(x_i - X_\alpha)\frac{\partial V_{i\alpha}}{\partial x_i} + (y_i - Y_\alpha)\frac{\partial V_{i\alpha}}{\partial y_i} + (z_i - Z_\alpha)\frac{\partial V_{i\alpha}}{\partial z_i} \,. \tag{2.105}$$

In fact, substituting (2.104) and its analogues for $\frac{\partial V_{i\alpha}}{\partial y_i}$ and $\frac{\partial V_{i\alpha}}{\partial z_i}$ into (2.105) and dividing both the numerator and denominator by

$$r_{i\alpha}^2 = (x_i - X_\alpha)^2 + (y_i - Y_\alpha)^2 + (z_i - Z_\alpha)^2 \quad , \tag{2.106}$$

we get $-V_{i\alpha} = k\,V_{i\alpha}$. This result suggests how to proceed further in our derivation.

Let us consider the term $-\langle\Psi|\sum_i \vec{r}_i \frac{\partial V_{eN}}{\partial \vec{r}_i}|\Psi\rangle$ corresponding to the electron–nuclear attraction on the left-hand side of (2.99), substitute in it the actual expression of V_{eN}

$$V_{eN} = \sum_{j,\alpha} V_{j\alpha} = \sum_{j,\alpha} \frac{-Z_\alpha}{r_{j\alpha}} , \tag{2.107}$$

and take into account that $V_{j\alpha}$ depends on $\vec{r}_i$ only if $i = j$: $\frac{\partial V_{j\alpha}}{\partial \vec{r}_i} = \delta_{ij}\frac{\partial V_{i\alpha}}{\partial \vec{r}_i}$, and, therefore, $\frac{\partial V_{eN}}{\partial \vec{r}_i} = \sum_\alpha \frac{\partial V_{i\alpha}}{\partial \vec{r}_i}$. We get

$$-\langle\Psi|\sum_i \vec{r}_i \frac{\partial V_{eN}}{\partial \vec{r}_i}|\Psi\rangle = -\sum_{i,\alpha}\langle\Psi|\vec{r}_i \frac{\partial V_{i\alpha}}{\partial \vec{r}_i}|\Psi\rangle . \tag{2.108}$$

We add and subtract $\vec{R}_\alpha$ in each term:

$$-\langle\Psi|\sum_i \vec{r}_i \frac{\partial V_{eN}}{\partial \vec{r}_i}|\Psi\rangle = -\sum_{i,\alpha}\langle\Psi|(\vec{r}_i - \vec{R}_\alpha)\frac{\partial V_{i\alpha}}{\partial \vec{r}_i}|\Psi\rangle - \sum_{i,\alpha}\langle\Psi|\vec{R}_\alpha \frac{\partial V_{i\alpha}}{\partial \vec{r}_i}|\Psi\rangle . \tag{2.109}$$

$V_{i\alpha}$ is a function of $\vec{r}_{i\alpha} = \vec{r}_i - \vec{R}_\alpha = \vec{i}x_{i\alpha} + \vec{j}y_{i\alpha} + \vec{k}z_{i\alpha}$. Accordingly, we turn to the variables $x_{i\alpha}, y_{i\alpha}$, and $z_{i\alpha}$ in the first term of the right-hand side. By using the chain rule, we have

$$\frac{\partial V_{i\alpha}}{\partial x_i} = \frac{\partial V_{i\alpha}}{\partial x_{i\alpha}}\frac{\partial x_{i\alpha}}{\partial x_i} = \frac{\partial V_{i\alpha}}{\partial x_{i\alpha}} , \tag{2.110}$$

as the other variables do not depend on x_i and $\frac{\partial x_{i\alpha}}{\partial x_i} = 1$. Accordingly:

$$-\langle\Psi|\sum_i \vec{r}_i \frac{\partial V_{eN}}{\partial \vec{r}_i}|\Psi\rangle = -\sum_{i,\alpha}\langle\Psi|\vec{r}_{i\alpha}\frac{\partial V_{i\alpha}}{\partial \vec{r}_{i\alpha}}|\Psi\rangle - \sum_{i,\alpha}\vec{R}_\alpha\langle\Psi|\nabla_i V_{i\alpha}|\Psi\rangle. \tag{2.111}$$

In the last term we have utilized the identity $\frac{\partial V_{i\alpha}}{\partial \vec{r}_i} \equiv \nabla_i V_{i\alpha}$ and have moved out of the integral the vector $\vec{R}_\alpha$, which does not depend on the variables of integration (the electronic coordinates). Now, $V_{i\alpha}$ is a homogeneous function of $\vec{r}_{i\alpha}$ $(k = -1)$, so according to Euler's theorem the first term on the right-hand side can be written as $-k\langle V_{eN}\rangle = \langle V_{eN}\rangle$. In the second term we utilize the fact that $V_{i\alpha}$ is a function of $|\vec{r}_{i\alpha}| = |\vec{r}_i - \vec{R}_\alpha|$ and that in such cases the interchange of the two endpoints causes a change of the sign in differentiation: $\nabla_i V_{i\alpha} \equiv -\nabla_\alpha V_{i\alpha}$. We arrive at

$$-\langle\Psi|\sum_i \vec{r}_i \frac{\partial V_{eN}}{\partial \vec{r}_i}|\Psi\rangle = -k\langle V_{eN}\rangle + \sum_{i,\alpha}\vec{R}_\alpha\langle\Psi|\nabla_\alpha V_{i\alpha}|\Psi\rangle . \tag{2.112}$$

By substituting (2.100), (2.102), and (2.112) into (2.99) we get

$$0 = 2\langle T_e\rangle - k\langle V_{ee}\rangle - k\langle V_{eN}\rangle + \sum_{i,\alpha} \vec{R}_\alpha \langle\Psi|\nabla_\alpha V_{i\alpha}|\Psi\rangle \ . \tag{2.113}$$

(We have conserved the coefficients $-k = 1$ only to keep track of the origin of the respective terms.)

Owing to the obvious equality $\nabla_\alpha V_{i\beta} = \delta_{\alpha\beta}\nabla_\alpha V_{i\alpha}$, one can present $\nabla_\alpha V_{i\alpha}$ as $\nabla_\alpha V_{i\alpha} = \nabla_\alpha \sum_\beta V_{i\beta}$, thus $\sum_i \nabla_\alpha V_{i\alpha} = \nabla_\alpha V_{eN}$, and the last term on the right-hand side of (2.113) can be rewritten accordingly:

$$0 = 2\langle T_e\rangle - k\langle V_{ee}\rangle - k\langle V_{eN}\rangle + \sum_{\alpha} \vec{R}_\alpha \langle\Psi|\nabla_\alpha V_{eN}|\Psi\rangle \ . \tag{2.114}$$

To get the final expression, we shall exploit the fact that the potential energy of the internuclear repulsion is a homogeneous function of the nuclear coordinates ($k = -1$), i.e., according to Euler's theorem

$$\sum_\alpha \vec{R}_\alpha \frac{\partial V_{NN}}{\partial \vec{R}_\alpha} = kV_{NN} \ . \tag{2.115}$$

We bring all terms of (2.115) to the same side, utilize the fact that $\frac{\partial V_{NN}}{\partial \vec{R}_\alpha} \equiv \nabla_\alpha V_{NN}$, as well as that the internuclear repulsion does not depend on the electronic wave function, and therefore its expectation value over the electronic wave function is simply its actual (Born–Oppenheimer) value, and get

$$0 = -k\langle V_{NN}\rangle + \langle\Psi|\sum_\alpha \vec{R}_\alpha \nabla_\alpha V_{NN}|\Psi\rangle \ . \tag{2.116}$$

Adding (2.116) to (2.114) and taking into account that $V = V_{ee} + V_{eN} + V_{NN}$:

$$0 = 2\langle T_e\rangle - k\langle V\rangle + \sum_\alpha \vec{R}_\alpha \langle\Psi|\nabla_\alpha (V_{eN} + V_{NN})|\Psi\rangle \ . \tag{2.117}$$

According to the Hellmann–Feynman theorem (Section 2), if the Hamiltonian depends on some parameter p, then one has for the derivative of the energy

$$\frac{\partial E}{\partial p} = \langle\Psi|\frac{\partial \hat{H}}{\partial p}|\Psi\rangle \ . \tag{2.118}$$

There are two terms in the Born–Oppenheimer electronic Hamiltonian that depend on the nuclear coordinates, V_{eN} and V_{NN}. Therefore $\left(\nabla_\alpha \equiv \frac{\partial}{\partial \vec{R}_\alpha}\right)$

$$\frac{\partial E}{\partial \vec{R}_\alpha} = \langle\Psi|\nabla_\alpha \hat{H}|\Psi\rangle = \langle\Psi|\nabla_\alpha (V_{eN} + V_{NN})|\Psi\rangle \ , \tag{2.119}$$

where E is the total energy obtained by solving the electronic Schrödinger equation. Substituting (2.119) into (2.117) we get the expression of the virial theorem for the case of the Born–Oppenheimer approximation:

$$2\langle T_e\rangle + \langle V\rangle + \sum_\alpha \vec{R}_\alpha \frac{\partial E}{\partial \vec{R}_\alpha} = 0\,. \tag{2.120}$$

Now, $\frac{\partial E}{\partial \vec{R}_\alpha} = \nabla_\alpha E = -\vec{F}_\alpha$, where $\vec{F}_\alpha$ is the *force* acting on the αth nucleus. Therefore, one obtains the "ideal" virial ratio $\frac{\langle V\rangle}{\langle T\rangle} = -2$ in the framework of the Born–Oppenheimer approximation only for the stationary points (minima, maxima, saddle points) of the potential energy surface, where all forces $\vec{F}_\alpha$ vanish. (In addition, the cases of infinite separations may satisfy this requirement.)

At the equilibrium configuration the virial theorem holds for the exact wave function as well as for some approximate ones, e.g., for the Hartree–Fock wave function, but generally it is not fulfilled if finite basis sets are used (Hartree–Fock–Roothaan method; see Section 5.1 of Chapter 6). Checking how close the "virial ratio" $-\frac{\langle V\rangle}{\langle T\rangle}$ is to the ideal value of 2 gives a useful measure of the quality of the wave function (i.e., of the basis applied). However, the value $-\frac{\langle V\rangle}{\langle T\rangle} \cong 2$ is only a necessary but not a sufficient condition for the wave function to be good; one may obtain the exact value 2 for rather bad wave functions if they were subjected to a "scaling" procedure (see Section 3.3).

3.5. The virial theorem and chemical bonding

An important consequence of the virial theorem is that *the formation of a chemical bond is a consequence of a lowering in the potential energy* (the electrostatic effects of formation of the overlap population in the bonding region and/or the electrons approach somewhat closer the nuclei), whereas kinetic energy is increasing during the bond formation (but this increase is only half of the lowering in the potential energy). In this respect, there are recurring mistakes in the chemical literature, going back to Hellmann, who assumed that the bonding is connected with a decrease in the kinetic energy that becomes possible as the electrons are delocalized in a larger region of space as compared to free atoms, and this reduces their momenta—and thus kinetic energy—in accord with the uncertainty principle. But the situation is much more complex: the virial theorem shows that the kinetic energy undergoes a resulting *increase*, and not a decrease, during the bond formation.

The problem was investigated in much detail by Ruedenberg on simple examples like H_2^+. He considered the bond formation as being formally divided into consecutive steps: the first step is a *promotion* with a contraction of the valence orbitals. This leads to a lowering of the potential energy (the electrons

in average are closer to the nuclei) but the kinetic energy increases *more* than this potential energy lowering. (This increase of the kinetic energy is an obvious consequence of the uncertainty relation.) Then the delocalization permits a kinetic energy lowering *with respect to the promoted state*; this decrease in the kinetic energy is inevitable to provide the possibility of conserving the lower potential energy that has been a result of the promotion. However, in the final account the bond formation is connected with a decrease of the potential energy and increase of the kinetic energy in accord with the virial theorem. (Example: in the Heitler–London-type calculations on H_2 using the minimal basis of 1s orbitals with exponential radial part, the optimal exponent is $\cong 1.19$ as opposed to the exponent 1 of the free H atom, representing a significant orbital contraction during bond formation.)

Notes

Section 1.2.

Standard treatment—e. g., [1–3]. In some cases (e.g., H-atom) the eigenfunctions of the Hamiltonian form a complete set only if the *continuum* of the positive eigenvalues is also included. In such a case the summation in Eq. (2.5) should be complemented by integration over the continuum (e.g. [4, 5]); this however does not influence the essence of the derivation.

Section 1.3.

It is the present author's custom to put Eq. (2.13) at the center of any treatment of variational problem. The derivation given here was first published in [6]; the derivation using the Lagrangian multipliers is a standard one, see e.g., [7, 8].

Section 1.4.

Originally was derived in [9]. It has much more conceptual than practical importance.

Section 1.5.

This result has recently been published by Surján [10] by using a somewhat different derivation. For the Hermitization scheme discussed here the author is indebted to Kuprievich [11]. Others (e.g., [12]) perform the Hermitization of Eq. (2.31) as

$$[\hat{H} - (\hat{P}\hat{H} + \hat{H}\hat{P})]\Psi = E\Psi$$

This operator, however, does not have the meaning of a projected Hamiltonian.

Section 2.1.

The theorem was first derived by Hellmann [13] and somewhat later by Feynman [14]. The present proof [15] is a simple generalization of the derivation of the variation principle given in section 1.3.; essentially the same approach is taken in [5], where it is also discussed in detail that the Hellmann–Feynman theorem is applicable for finite basis sets only if "floating" functions are used. There exists another, very simple, proof (e.g., [2]) which explicitly utilizes the assumption that the wave function satisfies the Schrödinger equation and that its normalization does not depend on the parameter α. In fact, if $\langle\Psi|\Psi\rangle = 1 = const$ and $\hat{H}\Psi = E\Psi$, then $\langle\frac{\partial\Psi}{\partial\alpha}|\hat{H}|\Psi\rangle = E\langle\frac{\partial\Psi}{\partial\alpha}|\Psi\rangle$ and

$$\frac{\partial}{\partial\alpha}\langle\Psi|\hat{H}|\Psi\rangle = E\frac{\partial}{\partial\alpha}\langle\Psi|\Psi\rangle + \langle\Psi|\frac{\partial\hat{H}}{\partial\alpha}|\Psi\rangle = \langle\Psi|\frac{\partial\hat{H}}{\partial\alpha}|\Psi\rangle\,.$$

It may be noted that if one uses a second quantization formalism properly generalized [16] for overlapping basis sets, then the Hamiltonian is defined by the *integrals* over the basis orbitals

and the Hellmann–Feynman theorem becomes again formally applicable even for atom-centered basis orbitals [17].

Section 2.2.
Derived in [18], and utilized [19, 20] for explaining the origin of barriers to internal rotations in molecules like ethane. Also see [2].

Section 3.1.
Textbook results (e.g., [7]).

Section 3.2.
The treatment is analogous to that in [2]. A simple alternative derivation may be found in Appendix 3 of [21].

Section 3.3.
The derivation closely follows that in [22].

Section 3.4.
Present author's derivation (unpublished). Usual derivations either combine the quantum mechanical derivation with the classical virial of external forces [21], or seem to be somewhat problematic [22].

Section 3.5.
As noted by Löwdin (see a footnote in [22]), it was probably G. Hall who first called attention to the contradiction between Hellmann's simple picture based on the uncertainty principle and the virial theorem. The problem was then clarified by Ruedenberg [23].

Suggested readings

The classical book of Slater [21] gives extremely useful examples of how applying general theory to simple problems can elucidate most important aspects of chemical bonding. The monograph of Epstein [8] gives an elaborated mathematical background for most problems discussed in this chapter.

References

1 H.A. Bethe, *Intermediate Quantum Mechanics*, Benjamin, New York 1964.
2 L. Zülicke, *Quantenchemie*, Deutcher Verlag der Wissenschaften, Berlin 1973.
3 H. Eyring, J. Walter, G.E. Kimball, *Quantum Chemistry*, John Wiley & Sons, New York 1944.
4 E. Kapuy and F. Török, *Az Atomok és Molekulák Kvantumelmélete* Akadémiai Kiadó, Budapest 1975.
5 F. L. Pilar, *Elementary Quantum Chemistry*, McGraw Hill, New York 1968.
6 I. Mayer, J. Ladik and G. Biczó, Int. J. Quantum Chem. **7**, 583 (1973)
7 L.D. Landau and I. M. Lifshitz, *Kvantovaya Mekhanika*, Nauka, Moscow 1974.
8 S.T. Epstein, *The Variation Method in Quantum Chemistry*, Academic Press, New York 1974.
9 C. E. Eckart, Phys. Rev. **36**, 878 (1930)
10 P.R. Surján, Chem. Phys. Lett. **325**, 120 (2000)
11 V.A. Kuprievich, personal communication to the author, Kiev, 1971.
12 S. Fudzinaga (S. Huzinaga) *Metod Molekularnikh Orbitaley* (Russian translation from Japanese), Mir, Moscow, 1983.
13 H. Hellmann, *Einführung in die Quantenchemie*, Deuticke, Leipzig 1937.
14 R.P. Feynman, Phys. Rev. **56**, 340 (1939)

15 I. Mayer, *Fejezetek a Kvantumkémiából*, BME Mérnöktovábbképző Int., Budapest 1987.
16 I. Mayer, Int. J. Quantum Chem. **23**, 341 (1983)
17 P.R. Surján, I. Mayer and R. Poirier, J. Mol. Structure (Theochem) **170**, 1 (1988)
18 H.J. Kim and R.G. Parr, J. Chem. Phys. **41**, 2892 (1964)
19 R. E. Wyatt and R. G. Parr, J. Chem. Phys. **41**, 3262 (1964)
20 R.G. Parr in *Modern Quantum Chemistry (Istanbul Lectures)*, ed. O. Sinanoğlu, Academic Press, New York 1965.
21 J.C. Slater, *Quantum Theory of Molecules and Solids vol. 1. Electronic Structure of Molecules*, McGraw-Hill, New York 1963.
22 P.-O. Löwdin, J. Mol. Spectroscopy **3**, 46 (1959)
23 K. Ruedenberg, Rev. Mod. Phys. **34**, 326 (1962)

Chapter 3

THE LINEAR VARIATIONAL METHOD AND LÖWDIN'S ORTHOGONALIZATION SCHEMES

1. The linear variational method (Ritz method)

One selects a finite number m of predefined functions φ_i, each of which satisfies the boundary conditions of the problem. The functions φ_i depend on the same independent variables as the wave function that we seek, but do not contain any other variables or adjustable parameters. (If $m \to \infty$ and the set of functions φ_i tends to be complete, the use of the Ritz method becomes equivalent with solving exactly the Schrödinger equation.)

The functions φ_i are called the "basis functions" of the problem; we assume that they are linearly independent, so they span an m-dimensional subspace of the Hilbert space.

We approximate the wave function Ψ as a linear combination of the basis functions φ_i:

$$\Psi = \sum_{i=1}^{m} c_i \varphi_i \ , \tag{3.1}$$

and determine the coefficients c_i as to ensure the energy to be minimal:

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \min \ . \tag{3.2}$$

As the functions φ_i are fixed, performing the variation of the wave function given in form of (3.1) means the variation of the coefficients c_i:

$$c_i \to c_i + \delta c_i \tag{3.3}$$

($\delta c_i = \Delta c_i = dc_i$—a small change of c_i; in the case of an independent scalar variable there is no need to distinguish between its differential and variation.)

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

If c_i changes according to (3.3), Ψ will change as

$$\Psi \to \Psi + \delta\Psi = \sum_{i=1}^{m}(c_i + \delta c_i)\varphi_i = \underbrace{\sum_{i=1}^{m} c_i\varphi_i}_{\Psi} + \underbrace{\sum_{i=1}^{m} \delta c_i\varphi_i}_{\delta\Psi} . \tag{3.4}$$

That means that the *most general* variation of the wave function Ψ used in the linear variational method is

$$\delta\Psi = \sum_{i=1}^{m} \delta c_i\varphi_i , \tag{3.5}$$

and this is the $\delta\Psi$ that one has to substitute into the variational condition $\langle\delta\Psi|\hat{H} - E|\Psi\rangle = 0$ derived in Section 1.3. of Chapter 2. We get

$$\langle\sum_{i=1}^{m} \delta c_i\varphi_i|\hat{H} - E|\Psi\rangle = 0 . \tag{3.6}$$

The variations δc_i are arbitrary and completely independent of each other; we may, for instance, consider the case when only a single $\delta c_i \neq 0$. That means that we must require for each i separately the fulfillment of the equality

$$\langle\varphi_i|\hat{H} - E|\Psi\rangle = 0 \qquad (i = 1, 2, ..., m) . \tag{3.7}$$

We substitute $\Psi = \sum_{j=1}^{m} c_j\varphi_j$ into (3.7) and transform the left-hand side:

$$\begin{aligned}\langle\varphi_i|\hat{H} - E|\sum_{j=1}^{m} c_j\varphi_j\rangle &= \sum_{j=1}^{m}\langle\varphi_i|\hat{H} - E|\varphi_j\rangle c_j \\ &= \sum_{j=1}^{m}\Big(\langle\varphi_i|\hat{H}|\varphi_j\rangle - E\langle\varphi_i|\varphi_j\rangle\Big)\, c_j = \sum_{j=1}^{m}(H_{ij} - ES_{ij})c_j = 0 ,\end{aligned} \tag{3.8}$$

where we introduced the notations

$$H_{ij} = \langle\varphi_i|\hat{H}|\varphi_j\rangle \quad \text{and} \quad S_{ij} = \langle\varphi_i|\varphi_j\rangle . \tag{3.9}$$

The equality

$$\sum_{j=1}^{m}(H_{ij} - ES_{ij})c_j = 0 \tag{3.10}$$

must be fulfilled for each i, i.e., we have obtained a system of m linear equations. The number of unknown quantities is $m+1$: the m coefficients c_i and the energy

E. However, the normalization of the wave function $\Psi = \sum_i c_i \varphi_i$ gives an additional condition; in other words, it is sufficient to determine the $m-1$ ratios c_i/c_1 $(i = 2, 3, \ldots, m)$ instead of the m coefficients c_i $(i = 1, 2, \ldots, m)$.

As it is known from mathematics, the system of homogeneous linear algebraic equations (3.10) has a nontrivial solution (i.e., not all $c_i = 0$) if and only if its determinant vanishes:

$$\det\{H_{ij} - ES_{ij}\} = 0 \,. \tag{3.11}$$

Equations of this type (in particular if $S_{ij} = \delta_{ij}$) are often called "secular equations"—a name coming from astronomy.

Equation (3.11) represents an equation of order m for E. Substituting its solution into the original system of equations, one can, in principle, determine the $m - 1$ ratios c_i/c_1. In practice, however, one does not proceed in that manner, as there are much more effective numerical methods that circumvent the explicit solution of an equation of order m.

We introduce the m by m matrices $\mathbf{H}$ and $\mathbf{S}$ with the elements H_{ij} and S_{ij}, respectively, and collect the coefficients c_i into a column-vector $\mathbf{c}$:

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_i \\ \vdots \\ c_m \end{pmatrix} . \tag{3.12}$$

Then, using matrix notations, we get (3.10) in the form

$$(\mathbf{H} - E\mathbf{S})\mathbf{c} = 0 \,, \tag{3.13}$$

i.e.,

$$\mathbf{Hc} = E\mathbf{Sc} \,, \tag{3.14}$$

a "generalized matrix eigenvalue equation." If the basis $\{\varphi_i\}$ is orthonormalized, $S_{ij} = \langle \varphi_i | \varphi_j \rangle = \delta_{ij}$; i.e., $\mathbf{S} = \mathbf{1}$ (unit matrix), and we get a "standard" matrix eigenvalue equation

$$\mathbf{Hc} = E\mathbf{c} \,. \tag{3.15}$$

Owing to their definitions in (3.9), matrices $\mathbf{H}$ and $\mathbf{S}$ are Hermitian: $H_{ij} = H^*_{ji}$; $S_{ij} = S^*_{ji}$; in most practical cases they are also real (real Hermitian matrix = symmetric matrix).

It is known from mathematics that every Hermitian matrix of dimension m by m has m eigenvectors. The eigenvalues of a Hermitian matrix are real. The eigenvectors belonging to different eigenvalues are automatically orthogonal; those corresponding to the same (degenerate) eigenvalue can also be chosen orthogonal. That means that the eigenvectors can be used to construct a complete orthonormalized basis in the (sub)space considered. By forming the columns of a matrix $\mathbf{U}$ from the different eigenvectors of a Hermitian matrix, one gets a unitary matrix: $(\mathbf{U}^\dagger\mathbf{U})_{ij} = \sum_k (\mathbf{U}^\dagger)_{ik}U_{kj} = \sum_k U^*_{ki}U_{kj} = \delta_{ij}$, as the scalar product of the ith and jth eigenvectors.

According to the definition, the similarity transformation of matrix $\mathbf{A}$ by matrix $\mathbf{T}$ gives matrix $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$. In the particular case when $\mathbf{T}$ is unitary ($\mathbf{T}^{-1} = \mathbf{T}^\dagger$) this reduces to $\mathbf{T}^\dagger\mathbf{A}\mathbf{T}$. Finding all the eigenvalues and eigenvectors of a Hermitian matrix is a task equivalent to its *diagonalization* by a similarity transformation: by performing the similarity transformation of matrix $\mathbf{H}$ by the unitary matrix $\mathbf{U}$ built up of its eigenvectors, one gets a diagonal matrix of the eigenvalues. The reverse also holds: the columns of a matrix $\mathbf{U}$ diagonalizing the Hermitian matrix $\mathbf{H}$ are its eigenvectors. In fact let

$$\mathbf{U}^\dagger\mathbf{H}\mathbf{U} = \mathbf{D} \quad \text{with} \quad \mathbf{D} = diag(E_i)\ , \tag{3.16}$$

i.e., $D_{ij} = E_i\delta_{ij}$. By multiplying (3.16) with $\mathbf{U}$ from the left and utilizing its unitary character ($\mathbf{U}\mathbf{U}^\dagger = \mathbf{1}$), we get

$$\mathbf{H}\mathbf{U} = \mathbf{U}\mathbf{D}\ , \tag{3.17}$$

which is the matrix form of the eigenvalue equation (3.15) written down for all the eigenvectors together. Indeed, expressing (3.17) in components, we have

$$\sum_k H_{ik}U_{kj} = \sum_k U_{ik}D_{kj} = \sum_k U_{ik}E_j\delta_{kj} = E_jU_{ij}\ , \tag{3.18}$$

which is an eigenvalue equation for the jth column of matrix $\mathbf{U}$, with the eigenvalue E_j. We can consider the jth column of matrix $\mathbf{U}$ as a column-vector $\mathbf{c}^j$ with elements $c^j_i = U_{ij}$, with the superscript j indicating the column. Substituting this into (3.18), we have

$$\sum_k H_{ik}c^j_k = E_jc^j_i\ , \tag{3.19}$$

the expanded form of the eigenvalue equation (3.15) with $\mathbf{c} = \mathbf{c}^j$ and $E = E_j$.

(Note the order of matrices on the right-hand-side of (3.17): the diagonal matrix of eigenvalues stands *after* the matrix of eigenvectors. Analogously, the matrix form of the generalized eigenvalue equation (3.14), is $\mathbf{HC} = \mathbf{SCD}$.)

There are standard numerical algorithms and programs available for solving the matrix–eigenvalue problem. Their detailed discussion is out of the scope

of the present book; I shall only make a few brief remarks. First of all, in practice one avoids the direct evaluation of the secular determinant (3.11) and the solution of a high-order algebraic equation for the eigenvalue E. Instead, for "compact" matrices stored entirely in computer memory, one uses methods based on the similarity transformations performing the diagonalization of the matrix. In the classical Jacobi method (1846) one finds the eigenvectors by performing consecutive two-by-two rotations of the basis vectors. In each of such step one nonzero off-diagonal element is eliminated. However, this matrix element may (and usually will) become nonzero again when another rotation is performed, therefore the procedure should be cyclically repeated for all off-diagonal matrix elements until all of them become sufficiently small. The disadvantage of the Jacobi method is that it is rather time consuming. Nowadays one mostly uses one of the modern (middle of the twentieth century) "tridiagonalization" methods (the Givens, Householder, QR, or QL algorithms) in which one performs a similarity transformation reducing the matrix to a "tridiagonal" form, in which only the elements on the main diagonal and their first neighbors differ from zero. The eigenvalue problem of the tridiagonal matrix is then solved by a refined algebraic method developed for this special problem and the results are transformed back to the original basis.

In order to treat problems related to (very) large but "sparse" matrices (i.e., matrices having relatively few nonzero elements) one uses methods that do not require one to compute and store the whole matrix explicitly—that often would be impossible even for the largest computers. One can avoid working with the whole matrix $\mathbf{A}$ if one has a procedure that permits one to compute in some way the product-vector $\mathbf{Ax}$ for an arbitrary vector $\mathbf{x}$ and if one wishes to determine only one or a few eigenvector(s) belonging to the eigenvalue(s) with the largest absolute value(s). In this case the memory requirement can be reduced to a few vectors of length m, where m is the order of the problem, while storing matrix $\mathbf{A}$ would require m^2 storage places—or at least $\frac{m(m+1)}{2}$ for a symmetric $\mathbf{A}$. In quantum chemistry this is the technique by which the large-scale CI calculations are performed: the product $\mathbf{Hx}$ is calculated by processing the lists of one- and two-electron *integrals* entering the matrix elements (Section 6. of Chapter 5), without calculating explicitly any off-diagonal element of the Hamiltonian (CI) matrix $\mathbf{H}$.

Alternate derivations by using Lagrangian multipliers

There are several variants of deriving the secular equations in (3.10)—or their matrix form in (3.14)—by using the method of Lagrangian multipliers. It may be instructive to consider two of them in some detail.

1. In the most common derivation one expands the integrals entering the Rayleigh coefficient in (3.2) as

$$\begin{aligned} \langle\Psi|\hat{H}|\Psi\rangle &= \sum_{i,j=1}^{m} c_i^* H_{ij} c_j \\ \langle\Psi|\Psi\rangle &= \sum_{i,j=1}^{m} c_i^* S_{ij} c_j \end{aligned} \tag{3.20}$$

and considers them as functions of the scalar variables c_i. The variational problem to be solved is then

$$\sum_{i,j=1}^{m} c_i^* H_{ij} c_j = \min \tag{3.21}$$

under the condition

$$\sum_{i,j=1}^{m} c_i^* S_{ij} c_j = 1 \ . \tag{3.22}$$

One forms the auxiliary function by introducing the Lagrangian multiplier λ as

$$F' = \sum_{i,j=1}^{m} c_i^* H_{ij} c_j + \lambda\Big(\sum_{i,j=1}^{m} c_i^* S_{ij} c_j - 1\Big) \tag{3.23}$$

and one requests $\frac{\partial F'}{\partial c_i} = 0$ and $\frac{\partial F'}{\partial c_i^*} = 0$ for all $i = 1, 2, \ldots, m$. In this derivation c_i and c_i^* are considered as two independent variables; the conditions $\frac{\partial F'}{\partial c_i} = 0$ and $\frac{\partial F'}{\partial c_i^*} = 0$ lead to complex conjugate equations and need not be considered separately. (See the remark that follows.) One should also join condition (3.22) to the equations obtained—or, which is the same, request $\frac{\partial F'}{\partial \lambda} = 0$. Then, by calculating $\frac{\partial F'}{\partial c_i^*}$, one gets the system of equations

$$\begin{aligned} \sum_{j=1}^{m} H_{ij} c_j + \lambda \sum_{j=1}^{m} S_{ij} c_j &= 0 \qquad (i = 1, 2, \ldots, m) \\ \sum_{i,j=1}^{m} c_i^* S_{ij} c_j &= 1 \ . \end{aligned} \tag{3.24}$$

Multiplying the first equation with c_i^*, summing over i, and taking into account the second equation, we get $\lambda = -\sum_{i,j=1}^{m} c_i^* H_{ij} c_j = -E$, thus

$$\sum_{j=1}^{m} H_{ij} c_j - E \sum_{j=1}^{m} S_{ij} c_j = 0 \ . \tag{3.25}$$

Q.E.D.

Remark

Obviously, quantities c_i and c_i^* are not strictly independent parameters—they represent complex conjugates of each other. But there are two truly independent real variables behind them, the real and imaginary parts $a_i = \mathrm{Re}\, c_i = \frac{1}{2}(c_i + c_i^*)$ and $b_i = \mathrm{Im}\, c_i = \frac{1}{2i}(c_i - c_i^*)$, respectively. One could study the linear variational problem by using only parameters a_i and b_i, but it is cumbersome. Instead one can request $\frac{\partial E}{\partial a_i} = \frac{\partial E}{\partial b_i} = 0$ and use the chain rule to get ($c_i = a_i + ib_i$, $c_i^* = a_i - ib_i$)

$$\begin{aligned} \frac{\partial E}{\partial a_i} &= \frac{\partial E}{\partial c_i}\frac{\partial c_i}{\partial a_i} + \frac{\partial E}{\partial c_i^*}\frac{\partial c_i^*}{\partial a_i} = \frac{\partial E}{\partial c_i} + \frac{\partial E}{\partial c_i^*} = 0 \\ \frac{\partial E}{\partial b_i} &= \frac{\partial E}{\partial c_i}\frac{\partial c_i}{\partial b_i} + \frac{\partial E}{\partial c_i^*}\frac{\partial c_i^*}{\partial b_i} = i\frac{\partial E}{\partial c_i} - i\frac{\partial E}{\partial c_i^*} = 0 \end{aligned} \tag{3.26}$$

and this system is equivalent with the conditions $\frac{\partial E}{\partial c_i} = \frac{\partial E}{\partial c_i^*} = 0$, used previously.

2. One can derive directly the matrix form in (3.14) of the secular equation, by modifying the derivation of the Schrödinger equation from the variational principle, discussed in Section 1.3 of Chapter 2 to the case when the wave function Ψ has the form of expansion (3.1).

We require $\langle\Psi|\hat{H}|\Psi\rangle$ to be stationary under the constraint $\langle\Psi|\Psi\rangle = 1$. We form the auxiliary functional

$$F' = \langle\Psi|\hat{H}|\Psi\rangle + \lambda(\langle\Psi|\Psi\rangle - 1)\ , \tag{3.27}$$

request $\delta F' = 0$, and join to it the condition $\langle\Psi|\Psi\rangle = 1$ (or—which is the same—request $\frac{\partial F'}{\partial \lambda} = 0$.) We get

$$\begin{aligned} \langle\delta\Psi|\hat{H}|\Psi\rangle + \lambda\langle\delta\Psi|\Psi\rangle + c.c. &= 0 \\ \langle\Psi|\Psi\rangle &= 1\ . \end{aligned} \tag{3.28}$$

As Ψ has the form in (3.1) with predefined functions φ_i, $\delta\Psi$ has the form

$$\delta\Psi = \sum_{i=1}^{m} \delta c_i \varphi_i \tag{3.29}$$

with arbitrary coefficients $\delta c_i \to 0$. They may be combined into a vector $\delta\mathbf{c}$. Substituting expansions (3.1) and (3.29) into (3.28) and using the definitions (3.9) of the matrix elements H_{ij} and S_{ij}, we get in matrix notations

$$\begin{aligned} \delta\mathbf{c}^\dagger\mathbf{H}\mathbf{c} + \lambda\delta\mathbf{c}^\dagger\mathbf{S}\mathbf{c} + c.c. &= 0 \\ \mathbf{c}^\dagger\mathbf{S}\mathbf{c} &= 1\ . \end{aligned} \tag{3.30}$$

As $\delta\mathbf{c}$ has an arbitrary phase factor, the term written explicitly and its complex conjugate ($c.c.$) should vanish separately:

$$\begin{aligned} \delta\mathbf{c}^\dagger\mathbf{H}\mathbf{c} + \lambda\delta\mathbf{c}^\dagger\mathbf{S}\mathbf{c} &= 0 \\ \mathbf{c}^\dagger\mathbf{S}\mathbf{c} &= 1\,. \end{aligned} \tag{3.31}$$

Now, $\delta\mathbf{c}^\dagger$ is an arbitrary vector (for instance, we may consider consecutively the cases when only one $\delta c_i \neq 0$; $i = 1, 2, \ldots, m$) therefore the first equation in (3.31) can hold only if

$$\mathbf{H}\mathbf{c} + \lambda\mathbf{S}\mathbf{c} = 0\,. \tag{3.32}$$

Multiplying this equation with $\mathbf{c}^\dagger$ from the left and taking into account the second equation in (3.31), we get $\lambda = -\mathbf{c}^\dagger\mathbf{H}\mathbf{c} = -\langle\Psi|\hat{H}|\Psi\rangle = -E$, i.e.,

$$\mathbf{H}\mathbf{c} = E\mathbf{S}\mathbf{c}\,. \tag{3.33}$$

Q.E.D.

2. Löwdin's symmetric orthogonalization

2.1. Matrix $\mathbf{S}^{-1/2}$

As we have seen in the previous section, the linear variational problem leads to a *generalized* eigenvalue equation $\mathbf{H}\mathbf{c} = E\mathbf{S}\mathbf{c}$, if the basis functions φ_i are not orthonormalized. Löwdin's symmetric orthogonalization scheme represents one of the methods permitting one to reduce such a generalized eigenvalue equation to a "standard" eigenvalue equation $\mathbf{H}'\mathbf{d} = E\mathbf{d}$. However, the symmetric orthogonalization (or simply "Löwdin orthogonalization") is not only an algorithm for treating the generalized eigenvalue problem, but it has also its own conceptual importance and it is often used in different areas of quantum chemical theory. (We shall note that Löwdin introduced other orthogonalization methods, as well, the most important of which is the "canonic" orthogonalization; see Section 3.2 that follows.)

The overlap (or "metric") matrix $\mathbf{S}$ having the elements $S_{ij} = \langle\varphi_i|\varphi_j\rangle$ is obviously Hermitian, and, if the basis set $\{\varphi_i\}$ is linearly independent, it is also positive definite (see Section 3 that follows). Therefore, matrix $\mathbf{S}$ can be diagonalized by a unitary matrix $\mathbf{V}$ and all the eigenvalues $\lambda_i > 0$:

$$\mathbf{V}^\dagger\mathbf{S}\mathbf{V} = \mathbf{\Lambda}\,; \quad \mathbf{\Lambda} = diag(\lambda_i) = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots \\ 0 & \lambda_2 & 0 & \ldots \\ 0 & 0 & \lambda_3 & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{3.34}$$

Let us form the diagonal matrix $\mathbf{K} = diag\left(\frac{1}{\sqrt{\lambda_i}}\right)$. Obviously, one has $\mathbf{\Lambda}\mathbf{K}^2 = \mathbf{1}$. We subject the matrices $\mathbf{\Lambda}$ and $\mathbf{K}$ to a similarity transforma-

tion with the *inverse* of matrix $\mathbf{V}$ that diagonalizes matrix $\mathbf{S}$. (As $\mathbf{V}$ is unitary, $\mathbf{V}^{-1} = \mathbf{V}^\dagger, \mathbf{V}^{\dagger -1} = \mathbf{V}$.) We get the matrices

$$\mathbf{V}\mathbf{\Lambda}\mathbf{V}^\dagger = \mathbf{S} \tag{3.35}$$

and

$$\mathbf{Z} = \mathbf{V}\mathbf{K}\mathbf{V}^\dagger \,. \tag{3.36}$$

Matrix $\mathbf{Z}$ defined in that manner has the specific property that $\mathbf{S}\mathbf{Z}^2 = \mathbf{1}$. In fact ($\mathbf{V}$ is unitary, $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\dagger$),

$$\mathbf{S}\mathbf{Z}^2 = \mathbf{V}\mathbf{\Lambda}\underbrace{\mathbf{V}^\dagger\mathbf{V}}_{\mathbf{1}}\mathbf{K}\underbrace{\mathbf{V}^\dagger\mathbf{V}}_{\mathbf{1}}\mathbf{K}\mathbf{V}^\dagger = \mathbf{V}\underbrace{\mathbf{\Lambda}\mathbf{K}^2}_{\mathbf{1}}\mathbf{V}^\dagger = \mathbf{V}\mathbf{V}^\dagger = \mathbf{1} \,. \tag{3.37}$$

One can check similarly that $\mathbf{Z}^2\mathbf{S} = \mathbf{Z}\mathbf{S}\mathbf{Z} = \mathbf{1}$. Therefore matrix $\mathbf{Z}$ can be considered as the $-\frac{1}{2}$ th power of matrix $\mathbf{S}$, and is usually denoted so:

$$\mathbf{S}^{-1/2} = \mathbf{Z} = \mathbf{V}\,diag\left\{\frac{1}{\sqrt{\lambda_i}}\right\}\mathbf{V}^\dagger \tag{3.38}$$

As it can immediately be seen, matrix $\mathbf{S}^{-1/2}$ is Hermitian $(\mathbf{S}^{-1/2})^\dagger = \mathbf{S}^{-1/2}$, because $\mathbf{V}$ is unitary and $\mathbf{S}$ is a Hermitian positive definite matrix, thus in (3.38) all $\lambda_i > 0$.

2.2. The $\mathbf{S}^{-1/2}$ transformation

Now, let us consider the generalized eigenvalue equation

$$\mathbf{H}\mathbf{c} = E\mathbf{S}\mathbf{c} \,. \tag{3.39}$$

We multiply (3.39) by $\mathbf{S}^{-1/2}$ from the left and insert the unit matrix written as $\mathbf{1} = \mathbf{S}^{-1/2}\mathbf{S}^{1/2}$ after matrix $\mathbf{H}$. (Matrix $\mathbf{S}^{1/2} = (\mathbf{S}^{-1/2})^{-1} = \mathbf{Z}^{-1}$; as will be seen, there is no need to actually compute it for treating the generalized eigenvalue problem.) We get

$$\underbrace{\mathbf{S}^{-1/2}\mathbf{H}\mathbf{S}^{-1/2}}_{\mathbf{H}'}\underbrace{\mathbf{S}^{1/2}\mathbf{c}}_{\mathbf{d}} = E\underbrace{\mathbf{S}^{1/2}\mathbf{c}}_{\mathbf{d}} \,. \tag{3.40}$$

By denoting $\mathbf{H}' = \mathbf{S}^{-1/2}\mathbf{H}\mathbf{S}^{-1/2}$ (Hermitian) and $\mathbf{d} = \mathbf{S}^{1/2}\mathbf{c}$, we get the "standard" eigenvalue problem

$$\mathbf{H}'\mathbf{d} = E\mathbf{d} \,. \tag{3.41}$$

The eigenvalues E coincide with the eigenvalues of the original problem in (3.39), whereas the eigenvectors of the latter can be obtained by the obvious relationship

$$\mathbf{c} = \mathbf{S}^{-1/2}\mathbf{d} \,. \tag{3.42}$$

An alternate treatment of the generalized eigenvalue equations is provided by the so-called Cholesky decomposition. In this case one does not consider matrix $\mathbf{S}$ as the product $\mathbf{S}^{1/2}\mathbf{S}^{1/2}$, but rather as $\mathbf{S} = \mathbf{L}\mathbf{L}^{\dagger}$, where $\mathbf{L}$ is a lower triangle matrix (i.e., it has non-zero elements only on and below the main diagonal); the transformation is similar to that described previously, only $\mathbf{S}^{-1/2}$ is replaced by $\mathbf{L}^{-1}$ or $\mathbf{L}^{-1\dagger}$. The Cholesky decomposition, however, is of interest only from the numerical point of view but has no role in quantum chemical theory.

2.3. The Löwdin basis

The $\mathbf{S}^{-1/2}$ transformation described previously is equivalent to a change of basis set when treating the linear variational problem. Instead of the non-orthogonal set $\{\varphi_i\}$ of basis functions, we introduce the "Löwdin-orthogonalized" (or "symmetrically orthogonalized") basis $\{\psi_j\}$ defined as

$$\psi_j = \sum_{k=1}^{m} S_{kj}^{-1/2}\varphi_k \; ; \quad j = 1, 2, \ldots, m \; . \tag{3.43}$$

Here we have introduced the shorthand $S_{kj}^{-1/2}$ instead of the cumbersome $(\mathbf{S}^{-1/2})_{kj}$.

If the basis $\{\varphi_i\}$ is linearly independent, then matrix $\mathbf{S}^{-1/2}$ is not singular. Then the functions ψ_j span *the same subspace* as the original φ_is but are orthonormalized:

$$\begin{aligned} \langle\psi_i|\psi_j\rangle &= \langle\sum_k S_{ki}^{-1/2}\varphi_k|\sum_l S_{lj}^{-1/2}\varphi_l\rangle \qquad (3.44)\\ &= \sum_{k,l} S_{ki}^{-1/2*}\langle\varphi_k|\varphi_l\rangle S_{lj}^{-1/2} = \sum_{k,l} S_{ik}^{-1/2} S_{kl} S_{lj}^{-1/2} = (\mathbf{S}^0)_{ij} = \delta_{ij} \; , \end{aligned}$$

where we have taken into account that the coefficients originating from the "bra" should be subjected to a complex conjugation and that $S_{ki}^{-1/2*} = S_{ik}^{-1/2}$, because $\mathbf{S}^{-1/2}$ is Hermitian.

The matrix elements of the Hamiltonian $\hat{H}$ in the basis $\{\psi_j\}$ are

$$\langle\psi_i|\hat{H}|\psi_j\rangle = \langle\sum_k S_{ki}^{-1/2}\varphi_k|\hat{H}|\sum_l S_{lj}^{-1/2}\varphi_l\rangle = \sum_{k,l} S_{ik}^{-1/2} H_{kl} S_{lj}^{-1/2} = H'_{ij} \; . \tag{3.45}$$

As the sets of basis functions $\{\varphi_i\}$ and $\{\psi_j\}$ span the same subspace, any function $\sum_k c_k\varphi_k$ can also be expanded in terms of ψ_js. It follows from the definition in (3.43) that functions φ_k can, in turn, be expressed as

$$\varphi_k = \sum_l S_{lk}^{+1/2}\psi_l \quad (k = 1, 2, \ldots, m) \; . \tag{3.46}$$

Therefore, one has

$$\sum_k c_k \varphi_k = \sum_k c_k \sum_l S_{lk}^{1/2} \psi_l = \sum_{k,l} (S_{lk}^{1/2} c_k) \psi_l = \sum_l d_l \psi_l \ , \tag{3.47}$$

where vector $\mathbf{d} = \mathbf{S}^{1/2}\mathbf{c}$ is the same as defined previously.

The previous results mean that, when turning to the basis $\{\psi_j\}$, the different matrices and vectors transform as

$$\begin{aligned} \mathbf{H} &\rightarrow \mathbf{H}' = \mathbf{S}^{-1/2}\mathbf{H}\mathbf{S}^{-1/2} \\ \mathbf{S} &\rightarrow \mathbf{1} \\ \mathbf{c} &\rightarrow \mathbf{d} \end{aligned} \tag{3.48}$$

Substituting these expressions into (3.15), we conclude that the solution of the linear variational problem in the orthonormalized basis $\{\psi_j\}$ leads to the "standard" eigenvalue equation

$$\mathbf{H}'\mathbf{d} = E\mathbf{d} \ . \tag{3.49}$$

This equation is the same as that we have obtained previously by the "$S^{-1/2}$ transformation" of the generalized eigenvalue equation (3.14). This is in full agreement with the obvious requirement that the function realizing a stationary value of the energy functional in a given subspace must not depend on the selection of the basis in that subspace; the value of the functional should be invariant as well.

2.4. The stationary property of Löwdin's symmetric orthogonalization scheme

In 1950 Löwdin introduced his symmetric orthogonalization method, mainly as an algorithm that can be utilized in numerical work. However, it became clear very soon that this scheme has much more general importance. In 1957 Carlson and Keller proved that the symmetrically orthogonalized functions have the following remarkable property: among all the possible functions obtained by orthogonalizing a given set of nonorthogonal ones, the symmetrically orthogonalized functions are the closest in the least squares sense (*vide infra*) to the original nonorthogonal functions.

The result that the Löwdin-orthogonalized orbitals resemble the original ones as much as possible gives an independent importance to the symmetric orthogonalization. In 1965 Fischer-Hjalmars showed that the two-electron integrals that are neglected in the theories applying the so-called "Zero Differential Overlap" (ZDO) approximation (actually she discussed the Pariser–Parr–Pople π-electron SCF method) are really small if the basis is assumed to be a Löwdin-orthogonalized counterpart of the original AO basis. This observation can serve

as a basis of justification for numerous semi-empirical quantum chemical methods (CNDO, MNDO, etc.) and many simplified models of solid state physics, in all of which the overlap of the basis orbitals is neglected. One may consider that a "Löwdin basis" has been implicitly introduced in these models and all the expressions are pertinent to the latter. (These methods usually apply other approximations, as well.)

We shall consider the stationarity condition mentioned and show that they are fulfilled by the Löwdin-orthogonalized functions.[1]

Let $\{\varphi_i\}$ be the set of N nonorthogonal but linearly independent basis functions, and $\mathbf{S}$ their overlap matrix with the elements $S_{ij} = \langle\varphi_i|\varphi_j\rangle$, and let $\{\psi_i\}$ be the orthonormalized functions we are seeking. We consider the case when the sets $\{\varphi_i\}$ and $\{\psi_i\}$ span the same subspace; then the orthogonal functions can be expanded in the terms of the original ones as

$$\psi_i = \sum_k C_{ki}\varphi_k \ . \tag{3.50}$$

The orthonormalization of the functions ψ_i means that

$$\langle\psi_i|\psi_j\rangle \ = \ \langle\sum_k C_{ki}\varphi_k|\sum_l C_{lj}\varphi_l\rangle \ = \ \sum_{k,l} C^*_{ki}S_{kl}C_{lj} = \delta_{ij} \tag{3.51}$$

or, in matrix form, introducing the matrix $\mathbf{C}$ of the expansion coefficients and its Hermitian adjoint $\mathbf{C}^\dagger$:

$$\mathbf{C}^\dagger\mathbf{S}\mathbf{C} = \mathbf{1} \ . \tag{3.52}$$

We admit the possibility of complex functions and complex C_{ij} coefficients. As the set $\{\varphi_i\}$ is linearly independent, matrices $\mathbf{S}$ and $\mathbf{C}$ are not singular. We can therefore multiply (3.52) with $\mathbf{C}^{-1}\mathbf{S}^{-1}$ from the right and with $\mathbf{C}$ from the left, as to obtain the orthogonality condition (3.51) as

$$\mathbf{C}\mathbf{C}^\dagger = \mathbf{S}^{-1} \ . \tag{3.53}$$

Now we shall turn to the stationarity condition. We wish to require the two sets to be as close as possible in the least squares sense, that is

$$\sum_i \int |\psi_i - \varphi_i|^2 d\tau \ = \ \sum_i \langle\psi_i - \varphi_i|\psi_i - \varphi_i\rangle \ = \ min \ , \tag{3.54}$$

where $\int d\tau$ denotes the integration over all the variables on which the functions φ_i, ψ_i depend. Expanding the integral $\langle\psi_i - \varphi_i|\psi_i - \varphi_i\rangle$ in (3.54) and taking into account that $\langle\psi_i|\psi_i\rangle = 1$, owing to the condition of the orthonormalization, as

[1]The following proof is reproduced from my paper "On Löwdin's Method of Symmetric Orthogonalization," I. Mayer, Int. J. Quantum Chem. **90**, 63 (2002); Copyright 2002 Wiley Periodicals.

well as that $\langle\varphi_i|\varphi_i\rangle = const$, we can formulate our condition as the following variational problem: one should find the expansion matrix **C** of (3.50) providing

$$\sum_i \Big\{\langle\varphi_i|\psi_i\rangle + \langle\psi_i|\varphi_i\rangle\Big\} = max \tag{3.55}$$

under the conditions of orthonormalization $\langle\psi_i|\psi_j\rangle = \delta_{ij}$.

One could, in principle, apply the standard method of Lagrangian multipliers to solve the constrained variation problem, but it seems simpler, and in particular much more transparent, to consider explicitly different *specified variations* of the functions ψ_i, and require the variation of the functional in the left-hand side of (3.55) to vanish under all the *permitted* variations, i.e., variations not violating the orthonormalization constraint:

$$\sum_i \Big\{\langle\varphi_i|\delta\psi_i\rangle + \langle\delta\psi_i|\varphi_i\rangle\Big\} = 0 \tag{3.56}$$

for all *sets* of $\delta\psi_i$s that conserve $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. As usual, this gives the necessary conditions of the extremum.

Variations of the phase

The simplest variation of a function is the variation of its phase, when the function ψ_i is replaced by

$$\psi_i + \delta\psi_i = e^{i\eta}\psi_i \tag{3.57}$$

with $\eta \to 0$ being an arbitrary complex variational parameter. Expanding the exponent in series ($e^{i\eta} = 1 + i\eta + \ldots$), one can readily see that the first-order variation of ψ_i is

$$\delta\psi_i = i\eta\psi_i \ . \tag{3.58}$$

The variation of the phase (3.57) does not violate the orthonormalization conditions at all, and they are conserved by (3.58) in the first order of η. Substituting (3.58) into (3.56) we get

$$\langle\varphi_i|i\eta\psi_i\rangle + \langle i\eta\psi_i|\varphi_i\rangle = 0, \tag{3.59}$$

which (as $\eta \to 0$, but $\eta \neq 0$) reduces to

$$\langle\varphi_i|\psi_i\rangle = \langle\psi_i|\varphi_i\rangle \ . \tag{3.60}$$

That means that integral $\langle\psi_i|\varphi_i\rangle$ is real. In components one has

$$\langle\varphi_i|\sum_l C_{li}\varphi_l\rangle = \langle\sum_l C_{li}\varphi_l|\varphi_i\rangle \ , \tag{3.61}$$

i.e.,

$$\sum_l S_{il}C_{li} = \sum_l C^*_{li}S_{li} \ . \tag{3.62}$$

In matrix notations this can be written as

$$(\mathbf{SC})_{ii} = (\mathbf{C}^\dagger\mathbf{S})_{ii}. \tag{3.63}$$

Variations of rotation type

The variation of type $\delta\psi_i = \eta\psi_j$ $(i \neq j)$ is not a permitted one because it violates the orthonormalization constraints. However, the *simultaneous variation*

$$\begin{aligned} \delta\psi_i &= \eta\psi_j \\ \delta\psi_j &= -\eta^*\psi_i \end{aligned} \tag{3.64}$$

is a permitted one, as in its case the orthonormalization condition is conserved in the first order of parameter η. In the case of the simultaneous variation of two functions ψ_i and ψ_j, the condition (3.56) reduces to

$$\langle\varphi_i|\delta\psi_i\rangle + \langle\delta\psi_i|\varphi_i\rangle + \langle\varphi_j|\delta\psi_j\rangle + \langle\delta\psi_j|\varphi_j\rangle = 0\ . \tag{3.65}$$

Substituting the variations (3.64) into (3.65) we get

$$\eta(\langle\varphi_i|\psi_j\rangle - \langle\psi_i|\varphi_j\rangle) + c.c. = 0\ . \tag{3.66}$$

Because η contains an arbitrary phase factor, (3.66) is fulfilled for arbitrary ηs if and only if both the quantity written explicitly and its complex conjugate $(c.c.)$ are separately equal to zero:

$$\langle\varphi_i|\psi_j\rangle - \langle\psi_i|\varphi_j\rangle = 0\ . \tag{3.67}$$

Substituting the expansion (3.50) of the functions we have

$$\langle\varphi_i|\sum_l C_{lj}\varphi_l\rangle - \langle\sum_l C_{li}\varphi_l|\varphi_j\rangle = 0\ , \tag{3.68}$$

i.e.,

$$\sum_l S_{il}C_{lj} = \sum_l C^*_{li}S_{lj}\ , \tag{3.69}$$

or, in matrix notations:

$$(\mathbf{SC})_{ij} = (\mathbf{C}^\dagger\mathbf{S})_{ij}\ . \tag{3.70}$$

Condition (3.63) should be fulfilled for all functions ψ_i, and condition (3.70) should be fulfilled for all pairs of functions ψ_i, ψ_j; that means that we have proven that all the elements of matrices $\mathbf{SC}$ and $\mathbf{C}^\dagger\mathbf{S}$ should be equal, i.e.:

$$\mathbf{SC} = \mathbf{C}^\dagger\mathbf{S}\ . \tag{3.71}$$

Multiplying this equation with $\mathbf{C}$ from the right and utilizing (3.52), we get

$$\mathbf{S}\mathbf{C}^2 = \mathbf{1} \tag{3.72}$$

or

$$\mathbf{C}^2 = \mathbf{S}^{-1} . \tag{3.73}$$

Comparing (3.73) with (3.53) we see that matrix $\mathbf{C}$ should be Hermitian ($\mathbf{C} = \mathbf{C}^\dagger$), which means that it should represent the *Hermitian square root* of matrix $\mathbf{S}^{-1}$:

$$\mathbf{C} = \mathbf{S}^{-1/2} . \tag{3.74}$$

Q.E.D.

2.5. Löwdin orthogonalization: A two-dimensional example

It may be instructive to go through explicitly the different steps discussed in Sections 2.1 and 2.3 for the simplest case of Löwdin orthogonalization of two overlapping functions.

Let us consider two normalized functions φ_1 and φ_2 with the overlap $\langle\varphi_1|\varphi_2\rangle = S$. Then the overlap matrix is

$$\mathbf{S} = \begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix} . \tag{3.75}$$

Symmetry considerations indicate (it is also easy to check directly) that the normalized eigenvectors of matrix $\mathbf{S}$ are the vectors

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{3.76}$$

with the eigenvalues $\lambda_1 = 1 + S$ and $\lambda_2 = 1 - S$, respectively. This means that the unitary matrix diagonalizing $\mathbf{S}$ is

$$\mathbf{V} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} . \tag{3.77}$$

This matrix has the interesting property that it is not only unitary but also Hermitian.

Matrix $\mathbf{\Lambda}$ is accordingly

$$\mathbf{\Lambda} = \begin{pmatrix} 1+S & 0 \\ 0 & 1-S \end{pmatrix} \tag{3.78}$$

and matrix $\mathbf{K}$ is

$$\mathbf{K} = \begin{pmatrix} \frac{1}{\sqrt{1+S}} & 0 \\ 0 & \frac{1}{\sqrt{1-S}} \end{pmatrix} \tag{3.79}$$

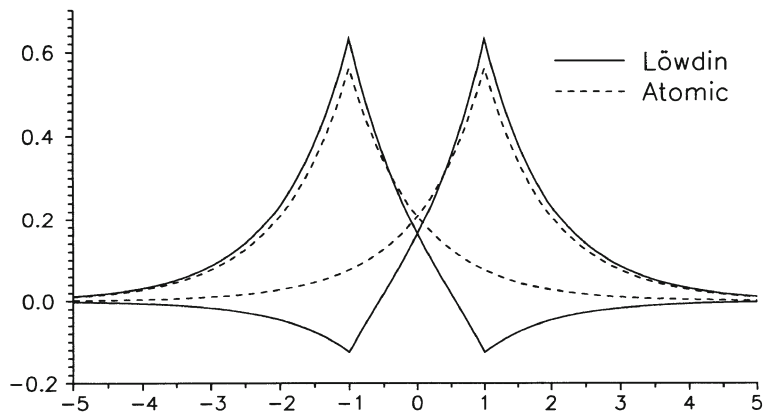
and we readily get the matrix $\mathbf{S}^{-1/2}$ as

$$\begin{aligned}
\mathbf{S}^{-1/2} &= \mathbf{V}\mathbf{K}\mathbf{V}^{\dagger} \\
&= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{1+S}} & 0 \\ 0 & \frac{1}{\sqrt{1-S}} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} \frac{1}{\sqrt{1+S}} + \frac{1}{\sqrt{1-S}} & \frac{1}{\sqrt{1+S}} - \frac{1}{\sqrt{1-S}} \\ \frac{1}{\sqrt{1+S}} - \frac{1}{\sqrt{1-S}} & \frac{1}{\sqrt{1+S}} + \frac{1}{\sqrt{1-S}} \end{pmatrix}
\end{aligned} \tag{3.80}$$

Accordingly, the Löwdin-orthogonalized functions ψ_1 and ψ_2 are

$$\begin{aligned}
\psi_1 &= \frac{1}{2}\left\{\left(\frac{1}{\sqrt{1+S}} + \frac{1}{\sqrt{1-S}}\right)\varphi_1 + \left(\frac{1}{\sqrt{1+S}} - \frac{1}{\sqrt{1-S}}\right)\varphi_2\right\} \\
\psi_2 &= \frac{1}{2}\left\{\left(\frac{1}{\sqrt{1+S}} - \frac{1}{\sqrt{1-S}}\right)\varphi_1 + \left(\frac{1}{\sqrt{1+S}} + \frac{1}{\sqrt{1-S}}\right)\varphi_2\right\} .
\end{aligned} \tag{3.81}$$

One can see that $\psi_1 \to \varphi_1$ and $\psi_2 \to \varphi_2$ if $S \to 0$.



The figure shows the exact atomic $1s$ orbitals of two hydrogen atoms placed at the distance of 2 a.u. (larger than the equilibrium distance of ca. 1.4. a.u. in the H_2 molecule) and their Löwdin-orthogonalized counterparts (S is as large as 0.5865.) It can be seen that the Löwdin orbitals do not differ very much from their overlapping counterparts in the respective "own" atomic region and have a negative cusp on another atom, which provides the resulting orthogonality of the Löwdin orbitals.

It may be of interest to study the dependence of the coefficients of the Löwdin-orthogonalized orbitals on the overlap integral S. The orbitals in (3.81) can be

written as

$$\begin{aligned} \psi_1 &= \alpha\varphi_1 + \beta\varphi_2 \\ \psi_2 &= \beta\varphi_1 + \alpha\varphi_2 \end{aligned} \tag{3.82}$$

with

$$\begin{aligned} \alpha &= \frac{1}{2}\left(\frac{1}{\sqrt{1+S}} + \frac{1}{\sqrt{1-S}}\right) \\ \beta &= \frac{1}{2}\left(\frac{1}{\sqrt{1+S}} - \frac{1}{\sqrt{1-S}}\right) \end{aligned} \tag{3.83}$$

The table displays α and β for some selected interatomic distances (S values)

R (a.u)	S	α	β
1.5	0.7252	1.3345	–0.5731
2	0.5865	1.1745	–0.3806
3	0.3485	1.0500	–0.1889

In the case of two functions, one may have also a specific look at Löwdin orthogonalization. Let us first consider two orthonormalized functions ψ_1 and ψ_2, and their normalized sum and difference:

$$\begin{aligned} g &= \frac{1}{\sqrt{2}}(\psi_1 + \psi_2) \\ u &= \frac{1}{\sqrt{2}}(\psi_1 - \psi_2) \end{aligned} \tag{3.84}$$

Functions g and u are orthogonal to each other.

The initial orthogonal functions ψ_1 and ψ_2 can be, in turn, recovered as the normalized sum of functions g and u:

$$\begin{aligned} \psi_1 &= \frac{1}{\sqrt{2}}(g + u) \\ \psi_2 &= \frac{1}{\sqrt{2}}(g - u)\ . \end{aligned} \tag{3.85}$$

In the case when we have two normalized but nonorthogonal functions φ_1 and φ_2, having the overlap integral S, then we can again form their normalized sum and difference, but the normalization coefficients will depend on S:

$$\begin{aligned} g &= \frac{1}{\sqrt{2(1+S)}}(\varphi_1 + \varphi_2) \\ u &= \frac{1}{\sqrt{2(1-S)}}(\varphi_1 - \varphi_2)\ . \end{aligned} \tag{3.86}$$

Functions u and g are again orthogonal.

Now, we may again form the normalized sum and difference of g and u, but in this case we do not recover the initial overlapping functions φ_1 φ_2 but get their Löwdin-orthogonalized counterparts ψ_1 and ψ_2:

$$\begin{aligned} \psi_1 &= \frac{1}{\sqrt{2}}(g+u) = \alpha\varphi_1 + \beta\varphi_2 \\ \psi_2 &= \frac{1}{\sqrt{2}}(g-u) = \beta\varphi_1 + \alpha\varphi_2 \end{aligned} \tag{3.87}$$

with the values of α and β defined in (3.83).

3. Linear independence of the basis and Löwdin's canonic orthogonalization

3.1. Eigenvalues of the overlap matrix: A measure for the linear independence of the basis

Let us consider m functions φ_i with an overlap matrix $\mathbf{S}$ having the elements $S_{kl} = \langle\varphi_k|\varphi_l\rangle$, and let $\mathbf{c}^i$ be an eigenvector of $\mathbf{S}$, corresponding to the eigenvalue λ_i:

$$\mathbf{S}\mathbf{c}^i = \lambda_i \mathbf{c}^i \ . \tag{3.88}$$

Let $\mathbf{c}^i$ be normalized to unity:

$$\mathbf{c}^{i\dagger}\mathbf{c}^i = \sum_k |c_k^i|^2 = 1 \ . \tag{3.89}$$

(In this section all the summations run from 1 to m.) Furthermore, let us form the function

$$\psi_i = \sum_k c_k^i \varphi_k \tag{3.90}$$

and determine its overlap with itself (the square of its norm):

$$\begin{aligned} \langle\psi_i|\psi_i\rangle &= \langle \sum_j c_j^i\varphi_j | \sum_k c_k^i\varphi_k\rangle = \sum_{j,k} c_j^{i*} S_{jk} c_k^i \\ &= \sum_j c_j^{i*}\left(\sum_k S_{jk}c_k^i\right) = \sum_j c_j^{i*}\lambda_i c_j^i = \lambda_i \sum_j |c_j^i|^2 = \lambda_i \ , \end{aligned} \tag{3.91}$$

where we have utilized the fact $\mathbf{c}^i$ is a normalized eigenvector of $\mathbf{S}$, therefore $\sum_k S_{jk}c_k^i = \lambda_i c_j^i$ and $\sum_j |c_j^i|^2 = 1$.

It follows from the result given in (3.91) that

a) The eigenvalues of the overlap matrix **S** are not negative because they represent the square of the norm $\langle\psi_i|\psi_i\rangle$ of a function ψ_i. If the basis is linearly independent, all $\lambda_i > 0$. (For a linearly independent basis the equality $\langle\psi_i|\psi_i\rangle = 0$ is possible only if all $c_j^i = 0$.) In other words, the overlap matrix **S** is "positive definite" if the basis set is linearly independent and "positive semidefinite" if it is linearly dependent.

b) One may consider the lowest eigenvalue of the matrix **S** as a measure of the linear independence of the basis used. To see this, we shall prove the following theorem.

Theorem

Let $\psi = \sum_j d_j\varphi_j$ be a linear combination of the basis functions φ_j with an arbitrary coefficient vector **d** normalized to unity ($\sum_j |d_j|^2 = 1$). Then the value of the integral $\langle\psi|\psi\rangle$ is limited by the lowest and largest eigenvalues of the overlap matrix **S**; i.e., it satisfies the inequalities

$$\lambda_{\min} \leq \langle\psi|\psi\rangle \leq \lambda_{\max} \,. \tag{3.92}$$

Proof

S is a Hermitian matrix and its eigenvectors form a complete orthonormalized set in the linear space of the m-dimensional vectors. Therefore, one may always write

$$\mathbf{d} = \sum_i \eta_i \mathbf{c}^i \tag{3.93}$$

with $\sum_i |\eta_i|^2 = 1$. Substituting expansion (3.93) of **d** into the integral $\langle\psi|\psi\rangle$ we get

$$\begin{aligned}
\langle\psi|\psi\rangle &= \langle\sum_j d_j\varphi_j|\sum_k d_k\varphi_k\rangle = \sum_{j,k} d_j^* S_{jk} d_k \\
&= \mathbf{d}^\dagger\mathbf{S}\mathbf{d} = \sum_{i,j} \eta_i^* \mathbf{c}^{i\dagger}\mathbf{S}\mathbf{c}^j\eta_j = \sum_{i,j} \eta_i^*\eta_j\lambda_j\mathbf{c}^{i\dagger}\mathbf{c}^j \qquad (3.94)\\
&= \sum_{i,j} \eta_i^*\eta_j\lambda_j\delta_{ij} = \sum_i |\eta_i|^2\lambda_i \,,
\end{aligned}$$

where we have taken into account that $\mathbf{S}\mathbf{c}^j = \lambda_j\mathbf{c}^j$ and $\mathbf{c}^{i\dagger}\mathbf{c}^j = \delta_{ij}$.

Because $\lambda_i \leq \lambda_{\max}$, $\sum_i |\eta_i|^2\lambda_i \leq \lambda_{\max}\sum_i |\eta_i|^2 = \lambda_{\max}$ and, similarly, $\lambda_i \geq \lambda_{\min}$, therefore one has $\sum_i |\eta_i|^2\lambda_i \geq \lambda_{\min}$.

Q.E.D.

If the basis exhibits an approximate linear dependence, then the lowest λ_i value becomes very small. Owing to rounding errors, one may consider it practically zero beyond some value (say $10^{-8} - 10^{-10}$, depending on the computer used). The presence of such an approximate linear dependence can cause trouble in the calculations, e.g., as the result of division with a number very close to 0 or of computing small differences of very large numbers. In such situations one has to omit from the basis those functions ψ_i that are nearly linear combinations of some others. These are the functions built up according to (3.90) by using those eigenvectors of **S** that correspond to the smallest eigenvalues (eigenvalues smaller than some threshold). In such cases the solutions of the generalized eigenvalue equations (Section 1) can be best computed by using Löwdin's "canonic orthogonalization"—a procedure different from the symmetric orthogonalization discussed in Section 2.

3.2. Löwdin's canonic orthogonalization

Let us renormalize the functions in (3.90) formed by using those eigenvectors of the overlap matrix **S** that correspond to a nonzero eigenvalue (to an eigenvalue larger than the selected threshold). Owing to (3.91), we get the functions

$$\psi_i' = \frac{1}{\sqrt{\lambda_i}}\psi_i = \frac{1}{\sqrt{\lambda_i}}\sum_j c_j^i \varphi_j \ , \tag{3.95}$$

that are normalized to unity and form an orthonormalized basis in the space spanned by the original functions $\{\varphi_j\}$:

$$\begin{aligned}
\langle \psi_i' | \psi_j' \rangle &= \frac{1}{\sqrt{\lambda_i \lambda_j}} \langle \sum_k c_k^j \varphi_k | \sum_l c_l^i \varphi_l \rangle \\
&= \frac{1}{\sqrt{\lambda_i \lambda_j}} \sum_{k,l} c_k^{j*} S_{kl} c_l^i = \frac{1}{\sqrt{\lambda_i \lambda_j}} \sum_k c_k^{j*} \sum_l S_{kl} c_l^i \qquad (3.96) \\
&= \frac{\lambda_i}{\sqrt{\lambda_i \lambda_j}} \sum_k c_k^{j*} c_k^i = \delta_{ij} \ ,
\end{aligned}$$

where we utilized the equality $\sum_l S_{kl} c_l^i = \lambda_i c_k^i$, and the fact that the eigenvectors $\mathbf{c}^i$ of the Hermitian matrix **S** are orthonormalized. (In this section the summations again run from 1 to m, except if indicated otherwise.)

The procedure of forming the orthonormal vectors $\{\psi_i'\}$ is called Löwdin's "canonic orthogonalization" of functions $\{\varphi_j\}$.

Let us define matrix **Q** consisting of the expansion coefficients of the functions ψ_i' in terms of the original basis $\{\varphi_j\}$ by forming the ith column from the

coefficients of the ith function ψ_i':

$$Q_{ji} = \frac{1}{\sqrt{\lambda_i}} c_j^i \,. \tag{3.97}$$

If the basis $\{\varphi_j\}$ is (strictly or approximately) linearly dependent, and we have, therefore, omitted one or more functions ψ_i corresponding to eigenvectors of $\mathbf{S}$ with zero (small) eigenvalues, then matrix $\mathbf{Q}$ is rectangular—it has as many columns as the number of linearly independent functions ψ_i'. If the number of original basis functions φ_j is m, and n of them are linearly independent, then matrix $\mathbf{Q}$ is an $m \times n$ matrix. (If there is no linear dependence, then $n = m$, of course.)

If there is a linear dependence, then the expansion of different functions in terms of the basis $\{\varphi_j\}$ is not unique. In fact, if the basis is linearly dependent, then one can form some vanishing combination $\sum_j p_j \varphi_j = 0$ of the basis functions φ_j. If some function Ψ can be expanded as $\Psi = \sum_j q_j \varphi_j$, then by adding to it such a vanishing combination we find that *the same* Ψ can be written also as $\Psi = \sum_j (q_j + p_j)\varphi_j = \sum_j q_j' \varphi_j$ with $q_j' \neq q_j$. In the same time, the same function has a unique expansion in the orthonormalized (thus linearly independent) basis $\{\psi_i'\} : \Psi = \sum_{j=1}^{n} r_j \psi_j'$. Substituting here the expansion (3.95) of the vector ψ_j', we get, taking into account (3.97)

$$\begin{aligned} \Psi &= \sum_{j=1}^{n} r_j \psi_j' = \sum_{j=1}^{n} r_j \frac{1}{\sqrt{\lambda_j}} \sum_{k=1}^{m} c_k^j \varphi_k = \sum_{j=1}^{n} r_j \sum_{k=1}^{m} Q_{kj} \varphi_k \\ &= \sum_{k=1}^{m} \left(\sum_{j=1}^{n} Q_{kj} r_j \right) \varphi_k = \sum_{k=1}^{m} (\mathbf{Qr})_k \varphi_k = \sum_{k=1}^{m} q_k \varphi_k \,. \end{aligned} \tag{3.98}$$

Consequently, one possible vector $\mathbf{q}$ is

$$\mathbf{q} = \mathbf{Qr} \,. \tag{3.99}$$

The $n \times n$ matrix $\mathbf{Q}^\dagger\mathbf{S}\mathbf{Q}$ is a unit matrix. In fact,

$$\begin{aligned}(\mathbf{Q}^\dagger\mathbf{S}\mathbf{Q})_{ij} &= \sum_{k,l}(\mathbf{Q}^\dagger)_{ik}S_{kl}Q_{lj} = \sum_{k,l} Q^*_{ki}S_{kl}Q_{lj} \\ &= \sum_{k,l}\frac{1}{\sqrt{\lambda_i}}c_k^{i*}S_{kl}\frac{1}{\sqrt{\lambda_j}}c_l^j = \frac{1}{\sqrt{\lambda_i\lambda_j}}\sum_k c_k^{i*}\sum_l S_{kl}c_l^j \\ &= \frac{\lambda_j}{\sqrt{\lambda_i\lambda_j}}\sum_k c_k^{i*}c_k^j = \frac{\lambda_j}{\sqrt{\lambda_i\lambda_j}}\delta_{ij} = \delta_{ij}\ , \qquad (3.100)\end{aligned}$$

where we utilized the fact that $\mathbf{c}^i$ and $\mathbf{c}^j$ are orthonormalized eigenvectors of $\mathbf{S}$. (The result in (3.100) also follows from the fact that the functions $\psi_i{}'$ are orthonormalized.)

Let us multiply the generalized eigenvalue equation

$$\mathbf{H}\mathbf{c} = \varepsilon\mathbf{S}\mathbf{c} \qquad (3.101)$$

by $\mathbf{Q}^\dagger$ from the left:

$$\mathbf{Q}^\dagger\mathbf{H}\mathbf{c} = \varepsilon\mathbf{Q}^\dagger\mathbf{S}\mathbf{c}\ . \qquad (3.102)$$

If the basis is linearly dependent, then vector $\mathbf{c}$ representing the solution of equations (3.101) and (3.102) we are looking for is not uniquely defined; however all possible functions $\sum_i c_i\varphi_i$ have the same expansion $\sum_j d_j\psi_j{}'$ with respect of the orthonormalized basis $\{\psi_i{}'\}$. As we have seen previously, in such a situation we can choose vector $\mathbf{c}$ as $\mathbf{c} = \mathbf{Q}\mathbf{d}$. (Of course, we may write vector $\mathbf{c}$ in the form $\mathbf{c} = \mathbf{Q}\mathbf{d}$ also in the case when basis $\{\varphi_i\}$ is linearly independent. In that case the transformation between the sets $\{\varphi_i\}$ and $\{\psi_i{}'\}$ is not singular, and one can also calculate the inverse of the transformation (3.99), i.e., the inverse of matrix $\mathbf{Q}$: one has $\mathbf{Q}^\dagger\mathbf{S}\mathbf{Q} = \mathbf{1}$ according to (3.100), from which $\mathbf{Q}^{-1} = \mathbf{Q}^\dagger\mathbf{S}$. This means that vectors $\mathbf{c}$ and $\mathbf{d}$ determine each other in a mutually unique manner if the basis is linearly independent.)

By substituting $\mathbf{c} = \mathbf{Q}\mathbf{d}$ into (3.102) we get

$$\mathbf{Q}^\dagger\mathbf{H}\mathbf{Q}\mathbf{d} = \varepsilon\mathbf{Q}^\dagger\mathbf{S}\mathbf{Q}\mathbf{d}\ , \qquad (3.103)$$

i.e., by denoting the Hermitian matrix $\mathbf{Q}^\dagger\mathbf{H}\mathbf{Q}$ by $\mathbf{H}'$ and taking into account that $\mathbf{Q}^\dagger\mathbf{S}\mathbf{Q} = \mathbf{1}$ according to (3.100):

$$\mathbf{H}'\mathbf{d} = \varepsilon\mathbf{d}\ , \qquad (3.104)$$

which is a "standard" eigenvalue problem.

It is easy to see that matrix $\mathbf{H}'$ is the matrix of the Hamiltonian $\hat{H}$ in the basis of the orthonormalized functions $\psi_i{}'$, from which the linearly dependent vectors

of the original basis $\{\varphi_i\}$ (if any) have been omitted. After the eigenvectors of (3.104) have been found, they can be transformed back to the original basis $\{\varphi_i\}$ by the relationship $\mathbf{c} = \mathbf{Qd}$; of course, one can obtain no more eigenvectors than the number n of the linearly independent basis functions.

Notes

Section 1.

Standard problem, included in all books on quantum chemistry, from [1] to [2]. (In the oldest books only the case of orthonormal basis functions was considered.) The derivation Eq.s (3.7)–(3.10) has a close relationship to the scheme called Galerkin (or Galiorkin) method in mathematics [3] or method of momenta in quantum chemistry [4, 5]. In the Galerkin method one looks for an approximate solution of a functional (e.g. differential) equation $\hat{L}f = 0$ by requesting its projections on different "weight functions" φ_i to vanish: $\int \varphi_i^* \hat{L} f dx = 0$. If one approximates the function f as a linear combination of a set of functions $\{\varphi_i\}$ and *the same* functions are used as weight functions, then the Galiorkin method is equivalent to the linear variational problem and leads to the secular equation (3.10). (The "coupled cluster approach" to electron correlation problem—see Section 3.3 of Chapter 8—represents a typical method of momenta.)

Section 2.

Löwdin's symmetric orthogonalization has been introduced in [6]. Its stationary property was first proved in [7]; the proof in Section 2.4. belongs to the present author [8]. The importance of the Löwdin orthogonalization for understanding the success of ZDO methods was pointed out in [9].

Löwdin's symmetric orthogonalization is used in most different areas of quantum chemistry—from *a posteriori* "natural bonding orbitals" analysis of [10–12] to the theory of intermolecular interactions [13– 15] or development of second quantized formalism for non-orthogonal basis sets [16,17] etc.

Section 3.

Löwdin introduced his "canonic orthogonalization" in [18] (also see [19]). The exclusion of the nearly linear dependencies is a general practice in every standard quantum chemical program (e.g., [20, 21]). This may cause some "difficult to understand" minor discontinuities in the computed potential surfaces, if a large basis set is used.

Suggested readings

Wilkinson's monograph [22] represents the classical work about the algorithms used in practical numerical solution of eigenvalue problems. (Also see, e.g., [23].) Davidson [24] discusses the problem from the specific quantum chemical standpoint (including his own algorithm for large matrices). In most practical cases one can use with success the standard linear algebraic program packages, e.g., "Lapack" [25].

For a brief survey of the non-orthogonality problem and different orthogonalization schemes in spirit of Löwdin we refer the reader to [26].

References

1 H. Eyring, J. Walter, G.E. Kimball, *Quantum Chemistry*, John Wiley & Sons, New York 1944.

2 T. Veszprémi and M. Fehér, *Quantum Chemistry: Fundamentals to applications*, Kluwer Academic/Plenum New York, 1999.

3 G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, MacGraw-Hill, New York 1961.

4 E. Szondy and T. Szondy, Acta Phys. Hung. **20**, 253 (1966)

5 S. F. Boys, Proc. Roy. Soc. **A309**, 195 (1969)

6 P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950)

7 B.C. Carlson and J. M. Keller, Phys. Rev. **105**, 102 (1957)

8 I. Mayer, Int. J. Quantum Chem. **90**, 63 (2002)

9 I. Fischer-Hjalmars, in *Modern Quantum Chemistry (Istanbul Lectures)*, ed. O. Sinanoğlu, Academic Press, New York 1965.

10 J.P. Foster and F. Weinhold, J. Am. Chem. Soc. **102**, 7211 (1980)

11 A.E. Reed, R. B. Weinstock and F. Weinhold, J. Chem. Phys. **83**, 735 (1985)

12 A.E. Reed, L. A. Curtiss and F. Weinhold, Chem. Rev. **88**, 899 (1988)

13 V. Kvasnička, V. Laurinc and I. Hubač, Phys. Rev. A. **10**, 2016 (1974).

14 V. Laurinc, V. Lukeś and S. Biskupič, Theor. Chem. Acc. **99**, 53 (1989).

15 V. Lukeś, V. Laurinc and S. Biskupič, Int. J. Quantum Chem. **75**, 81 (1999)

16 V. Kvasnička, Chem. Phys. Lett. **51**, 165 (1977)

17 I. Mayer, Int. J. Quantum Chem. **23**, 341 (1983)

18 P.-O. Löwdin, Adv. Phys. **5**, 1 (1966)

19 P.-O. Löwdin, J. appl. Phys. Suppl. **33**, 251 (1962)

20 Program "HONDO 2000", M. Dupuis, A. Marques and E.R. Davidson. (Based on "HONDO 95.3" of the same authors, Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN 47405.)

21 Program "Gaussian 98", M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.

22 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford 1965.

23 B. P. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs 1980.

24 E. R. Davidson, *Matrix Eigenvector Methods* in *Methods of Computational Molecular Physics* (ed. G.H.F. Diercksen and S. Wilson) Reidel, Dordrecht 1983.

25 See the web-site *http://www.netlib.org/lapack*

26 G. del Re, *The Non-Orthogonality Problem and Orthogonalization procedures* in *Quantum Science, Methods and Structure*, Plenum, New York 1976.

Chapter 4

PERTURBATIONAL METHODS

1. Nondegenerate Rayleigh–Schrödinger perturbation theory

1.1. The problem

Let us assume that we know the solutions (i.e., the orthonormalized eigenfunctions Ψ_i^0 and the corresponding eigenvalues E_i^0) of some "unperturbed" Schrödinger equation

$$\hat{H}^0 \Psi_i^0 = E_i^0 \Psi_i^0 \ . \tag{4.1}$$

We wish to utilize this knowledge for an approximate treatment of a "perturbed" Schrödinger equation containing a Hamiltonian $\hat{H}$ that is considered "close" to $\hat{H}^0$:

$$\hat{H}\Psi = (\hat{H}^0 + \hat{V})\Psi = E\Psi \ . \tag{4.2}$$

The perturbed and "unperturbed" Hamiltonians $\hat{H}$ and $\hat{H}^0$, respectively, are "close" to each other, if $\hat{V}$ is "small." One can express the smallness of $\hat{V}$ by introducing the notation

$$\hat{V} = \lambda \hat{W} \ , \tag{4.3}$$

where operator $\hat{W}$ has the same order of magnitude $\mathcal{O}(1)$ as $\hat{H}^0$, whereas λ is a (small) parameter of perturbation. We get

$$(\hat{H}^0 + \lambda \hat{W})\Psi = E\Psi \ . \tag{4.4}$$

In perturbation theory (PT) λ is usually considered as a small continuous parameter (formally $\lambda \to 0$), and both the wave function and the energy are

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

expanded in power series according to λ:

$$\Psi_i = \sum_{l=0}^{\infty} \lambda^l \Psi_i^{(l)} \tag{4.5}$$

and

$$E_i = \sum_{k=0}^{\infty} \lambda^k E_i^{(k)} \ . \tag{4.6}$$

There are problems in which λ is really a continuous variable, the value of which may be freely changed. This is the case, for instance, when one considers the effect of an external electromagnetic field on the system studied, which can be regulated in the experimental setup. However, in most cases of quantum chemical interest λ is not really a variable, and considering it as being a continuous parameter represents only a mathematical idealization. Nevertheless, if λ is truly small, the systematization of the results by the powers of λ is often very useful.

Care must be taken as even a weak perturbation may change the *overall character* of the spectrum of the Hamiltonian. For instance, if there is a hole of finite depth on the potential surface, then there may be stationary states describing a charged particle trapped in that hole. However, if we introduce an arbitrarily weak uniform external electric field $\vec{\mathcal{E}}$, which corresponds to an electrostatic potential $U = -\vec{\mathcal{E}}\vec{r}$, then at sufficiently large distances (in a direction depending on the direction of the field) the potential energy will be lower than that at the original minimum. This will change the character of the particle's motion from a bounded to an unbounded one; accordingly the spectrum of the Hamiltonian will change from a discrete to a continuous one. This does not mean that we may not find perturbed states of bounded type; they, however, will be not true stationary states, but only quasistationary ones ("resonances") decaying as a consequence of "tunnel effect" with a characteristic time depending on the height and width of the resulting potential barrier. (A similar picture of tunneling through a high but finite barrier is used in the theory of radioactivity to describe the phenomenon of α-decay.)

If the perturbation is not connected with a variable external field, then the actual physical problem defines only the total perturbed Hamiltonian $\hat{H}$ but not the unperturbed $\hat{H}^0$. Therefore, the decomposition of $\hat{H}$ into unperturbed part $\hat{H}^0$ and perturbation $\hat{V}$ is more or less arbitrary. Usually the selection of the unperturbed problem is performed according to (mathematical) convenience. This selection is, of course, crucial from the point of view of the convergence properties of the perturbational expansions. It is to be noted in this connection that the actual perturbational series often exhibit only so-called "asymptotic convergence," which means that the perturbational approximations of the exact

solutions improve until some (finite) order of PT, but the perturbational series eventually diverges. However, this possibility is usually disregarded when developing the perturbational formalism; it is essentially based on the tacit assumption that all the series considered are absolutely convergent—for instance, one may multiply them term by term. (The point is that the perturbational series are convergent for sufficiently small $|\lambda|$ values; the radius of convergence, however, may or may not include the actual "physical" values of λ.) Accordingly, we shall deal with this problem in some detail only when considering the size–consistency problem.

It may be noted that many authors do not use two different notations $\hat{V}$ and $\hat{W}$ for the perturbation operator, differing by the factor λ. Instead, they formally introduce the factor λ to operator $\hat{V}$ in (4.2), perform the expansions according to the powers of λ, and substitute $\lambda = 1$ in the final expressions obtained. This procedure is fully equivalent with ours; however we feel the use of the pair of different operators connected by (4.3) is more transparent.

By substituting the expansions (4.5) and (4.6) into the Schrödinger equation (4.4), one gets

$$(\hat{H}^0 + \lambda\hat{W}) \sum_{l=0}^{\infty} \lambda^l \Psi_i^{(l)} = \sum_{k=0}^{\infty} \lambda^k E_i^{(k)} \sum_{l=0}^{\infty} \lambda^l \Psi_i^{(l)} \,. \tag{4.7}$$

In (4.7) parameter λ is considered (at least formally) as an independent continuous variable; therefore, its different powers must have equal coefficients on the both sides.

Considering the terms with λ^0 (i.e., those that do not contain λ and do not vanish, therefore, as $\lambda \to 0$) one gets

$$\hat{H}^0 \Psi_i^{(0)} = E_i^{(0)} \Psi_i^{(0)} \,, \tag{4.8}$$

i.e., we recover the unperturbed problem

$$\Psi_i^{(0)} = \Psi_i^0 \,; \quad E_i^{(0)} = E_i^0 \tag{4.9}$$

as it could be expected. ($\hat{H} \to \hat{H}^0$ if $\lambda \to 0$.)

It is worthwhile to separate out the zero-order terms in (4.5) and (4.6), and write

$$\Psi_i = \Psi_i^0 + \sum_{l=1}^{\infty} \lambda^l \Psi_i^{(l)} \tag{4.10}$$

and

$$E_i = E_i^0 + \sum_{k=1}^{\infty} \lambda^l E_i^{(k)} \,. \tag{4.11}$$

Equivalently, one may write

$$\begin{aligned} \Psi_i &= \Psi_i^0 + \sum_{l=1}^{\infty} \psi_i^{(l)} \\ E_i &= E_i^0 + \sum_{k=1}^{\infty} \varepsilon_i^{(k)} \end{aligned} \tag{4.12}$$

where

$$\psi_i^{(l)} = \lambda^l \Psi_i^{(l)} \; ; \quad \varepsilon_i^{(k)} = \lambda^k E_i^{(k)} \; . \tag{4.13}$$

(As noted previously, one may avoid introducing the intermediate notations by putting formally $\lambda = 1$ *after* the expansions are completed.)

1.2. "Algebraic" expansion

One can expand every function $\Psi_i^{(l)}$ in terms of the solutions Ψ_k^0 of the unperturbed problem (4.1):

$$\Psi_i^{(l)} = \sum_{\substack{k \\ (k \neq i)}} c_{ik}^{(l)} \Psi_k^0 \qquad (l \geq 1) \; . \tag{4.14}$$

We have introduced the restriction $k \neq i$ on the summation, indicating that there is no need to include Ψ_i^0 in the expansion of the wave function corrections $\Psi_i^{(l)}$. This is the case because if it were included, it could be combined with the zero-order (unperturbed) wave function, which eventually would influence the normalization of the wave function only. The normalization, however, may be chosen arbitrarily. The wave function written out in the form of (4.10) and (4.14) is not normalized to unity but has the form:

$$\Psi_i = \Psi_i^0 + \chi_i \tag{4.15}$$

where

$$\langle \Psi_i | \Psi_i^0 \rangle = 1 \; ; \quad \langle \Psi_i^0 | \chi_i \rangle = 0 \; . \tag{4.16}$$

The normalization of this type is usually called "intermediate" (or "correlational") normalization.

By substituting the expansions (4.10), (4.11), and (4.14) into the perturbed Schrödinger equation $(H^0 + \lambda \hat{W}) \Psi_i = E_i \Psi_i$, we may write

$$(\hat{H}^0 + \lambda\hat{W})\left[\Psi_i^0 + \sum_{l=1}^{\infty}\lambda^l \sum_{\substack{q \\ (q\neq i)}} c_{iq}^{(l)}\Psi_q^0\right] = \left[E_i^0 + \sum_{k=1}^{\infty}\lambda^k E_i^{(k)}\right]\left[\Psi_i^0 + \sum_{l=1}^{\infty}\lambda^l \sum_{\substack{q \\ (q\neq i)}} c_{iq}^{(l)}\Psi_q^0\right] . \tag{4.17}$$

We multiply (4.17) by Ψ_i^{0*} and integrate, by taking into account that

$$\langle\Psi_p^0|\hat{H}^0|\Psi_q^0\rangle = E_q^0\delta_{pq} \tag{4.18}$$

and, in particular, $\langle\Psi_i^0|\hat{H}^0|\Psi_q^0\rangle = 0$ for $q \neq i$. We get

$$E_i^0 + \lambda\langle\Psi_i^0|\hat{W}|\Psi_i^0\rangle + \lambda\langle\Psi_i^0|\hat{W}|\sum_{l=1}^{\infty}\lambda^l \sum_{\substack{q \\ (q\neq i)}} c_{iq}^{(l)}\Psi_q^0\rangle = E_i^0 + \sum_{k=1}^{\infty}\lambda^k E_i^{(k)} . \tag{4.19}$$

The zero-order terms cancel on the two sides, in agreement with (4.8). By introducing the notation

$$W_{ij} = \langle\Psi_i^0|\hat{W}|\Psi_j^0\rangle \tag{4.20}$$

we can write

$$\sum_{k=1}^{\infty}\lambda^k E_i^{(k)} = \lambda W_{ii} + \sum_{l=1}^{\infty}\lambda^{l+1} \sum_{\substack{q \\ (q\neq i)}} c_{iq}^{(l)} W_{iq} . \tag{4.21}$$

On the right-hand side we can introduce a new summation index $k = l + 1$, so $l = k - 1$; as l runs from 1 to ∞, k runs from 2 to ∞:

$$\sum_{k=1}^{\infty}\lambda^k E_i^{(k)} = \lambda W_{ii} + \sum_{k=2}^{\infty}\lambda^k \sum_{\substack{q \\ (q\neq i)}} c_{iq}^{(k-1)} W_{iq} . \tag{4.22}$$

We request the coefficients at the different powers of λ to be equal on the two sides. Considering λ^1, we get

$$E_i^{(1)} = W_{ii} \tag{4.23}$$

and for λ^k ($k \geq 2$)

$$E_i^{(k)} = \sum_{\substack{q \\ (q\neq i)}} c_{iq}^{(k-1)} W_{iq} . \tag{4.24}$$

That means that the kth correction of the energy is determined by the $(k-1)$th correction of the wave function (also see Section 1.4).

In order to get the corrections of the wave function, we multiply (4.17) by one of the functions Ψ_j^{0*} $(j \neq i)$ and integrate:

$$\lambda W_{ji} + E_j^0 \sum_{l=1}^{\infty} \lambda^l c_{ij}^{(l)} + \sum_{l=1}^{\infty} \lambda^{l+1} \sum_{\substack{q \\ (q \neq i)}} c_{iq}^{(l)} W_{jq} = \left[E_i^0 + \sum_{k=1}^{\infty} \lambda^k E_i^{(k)} \right] \sum_{l=1}^{\infty} \lambda^l c_{ij}^{(l)} . \tag{4.25}$$

By introducing the summation index $n = k + l$ (it runs from 2 to ∞) and replacing, accordingly, l by $n - k$, we get

$$(E_j^0 - E_i^0) \sum_{l=1}^{\infty} \lambda^l c_{ij}^{(l)} = \sum_{n=2}^{\infty} \lambda^n \sum_{k=1}^{n-1} E_i^{(k)} c_{ij}^{(n-k)} - \\ -\lambda W_{ij} - \sum_{l=1}^{\infty} \lambda^{l+1} \sum_{\substack{q \\ (q \neq i)}} c_{iq}^{(l)} W_{jq} , \tag{4.26}$$

i.e., by introducing the summation index $k = l + 1$ in the last term:

$$\sum_{l=1}^{\infty} \lambda^l c_{ij}^{(l)} = \frac{1}{E_j^0 - E_i^0} \left[\sum_{n=2}^{\infty} \lambda^n \sum_{k=1}^{n-1} E_i^{(k)} c_{ij}^{(n-k)} - \lambda W_{ji} \right. \\ \left. - \sum_{k=2}^{\infty} \lambda^k \sum_{\substack{q \\ (q \neq i)}} c_{iq}^{(k-1)} W_{jq} \right] . \tag{4.27}$$

By comparing the coefficients of λ^1 on the two sides we get

$$c_{ij}^{(1)} = -\frac{W_{ji}}{E_j^0 - E_i^0} . \tag{4.28}$$

This result indicates that the Rayleigh–Schrödinger PT is expected to be well applicable in those cases in which these fractions are small. (More precisely, one should consider these fractions multiplied by λ, i.e., the fractions $\frac{V_{ij}}{(E_j^0 - E_i^0)}$ containing the matrix elements of the "original" perturbation operator $\hat{V} = \lambda \hat{W}$.)

For a general $l \geq 2$ value the comparison of the coefficient at λ^l on the two sides of (4.27) gives

$$c_{ij}^{(l)} = \frac{1}{E_j^0 - E_i^0} \left[\sum_{k=1}^{l-1} E_i^{(k)} c_{ij}^{(l-k)} - \sum_{\substack{q \\ (q \neq i)}} c_{iq}^{(l-1)} W_{jq} \right] . \tag{4.29}$$

This result, together with (4.24) indicates that the Rayleigh–Schrödinger PT can be considered as a recursion procedure. (This permits the perturbational series to be treated explicitly up to very high orders for some relatively small problems.)

By using (4.24) and (4.28) we get the explicit expression for the second-order energy correction:

$$E_i^{(2)} = \sum_{\substack{j \\ (j \neq i)}} \left(-\frac{W_{ji}}{E_j^0 - E_i^0} \right) W_{ij} = - \sum_{\substack{j \\ (j \neq i)}} \frac{|W_{ij}|^2}{E_j^0 - E_i^0} . \tag{4.30}$$

It can be seen that the second-order energy correction of a ground state is always negative (because then all $E_j^0 > E_i^0$).

In appendix VIII we have collected the final explicit formulae of the Rayleigh–Schrödinger PT up to fourth order in energy and third order in wave function. (Also see the formulae in Section 1.4.)

It is important to stress that a finite order perturbational expansion of the energy does not provide any upper (or lower) limit of the exact energy. The perturbational series often (but not necessarily) converges in an alternating fashion. Also, the sum of the first- and second-order energy corrections often "overshoots," i.e., gives an energy value that is lower than the exact one—but this is not necessarily the case. Of course, one obtains an upper limit of the energy if one computes the expectation value of the Hamiltonian with a wave function obtained by using perturbational theory. As discussed in Section 1.4, this expectation value is expected (much) more accurate than the perturbational wave function. In fact, the wave function up to nth order has an error proportional to $\varepsilon = \lambda^{n+1}$, therefore the error of the energy expectation value should be proportional to $\varepsilon^2 = \lambda^{2n+2}$. This is a formulation of Wigner's "$2n + 1$ theorem"; it deserves to be considered more in detail in Section 1.4.

Compact matrix notations

The previous results can be presented very compactly by using matrix notations. We introduce the unit vector $\mathbf{e}^i$, all elements of which are zero, except the ith, which is 1: and the vectors $\mathbf{c}^{i(l)}$ the elements of which are the expansion coefficients $c_{ij}^{(l)}$ for $j \neq i$ and 0 for $j = i$. Furthermore, we define the diagonal matrix $\mathbf{\Delta}^i$, with $\frac{1}{E_j - E_i}$ in the jth diagonal position, except the ith, which is

equal to zero. It is easy to check explicitly that by using these notations, (4.24), (4.28), and (4.29) can be compactly written as follows:

$$\begin{aligned} E_i^{(k)} &= \mathbf{e}^{i\dagger}\mathbf{W}\mathbf{c}^{i(k-1)} \\ \mathbf{c}^{i(1)} &= -\mathbf{\Delta}^i\mathbf{W}\mathbf{e}^i \\ \mathbf{c}^{i(l)} &= \mathbf{\Delta}^i\left[\sum_{k=1}^{l-1} E_i^{(k)}\mathbf{c}^{i(l-k)} - \mathbf{W}\mathbf{c}^{i(l-1)}\right] \end{aligned} \tag{4.31}$$

1.3. The use of the reduced resolvent in the Rayleigh–Schrödinger perturbation theory

The reduced resolvent

Let us define a resolution of identity in the space of functions $|\Psi_k^0\rangle$ as

$$\hat{P} + \hat{Q} = 1 \tag{4.32}$$

with

$$\hat{P} = |\Psi_i^0\rangle\langle\Psi_i^0| \tag{4.33}$$

and

$$\hat{Q} = 1 - \hat{P} = \sum_{\substack{k \\ (k\neq i)}} |\Psi_k^0\rangle\langle\Psi_k^0| \,. \tag{4.34}$$

Here $|\Psi_i^0\rangle$ is the ith solution of the unperturbed problem $\hat{H}^0|\Psi_i^0\rangle = E_i^0|\Psi_i^0\rangle$, the perturbation of which we are actually considering. We assume that it is nondegenerate. Furthermore, let us write down the spectral resolution of the operator $\hat{H}^0 - E_i^0$ (more rigorously, of the operator $\hat{H}^0 - E_i^0\hat{1}$) in the terms of the eigenvectors $|\Psi_l^0\rangle$ of operator $\hat{H}^0$:

$$\hat{H}^0 - E_i^0 = \sum_l (E_l^0 - E_i^0)|\Psi_l^0\rangle\langle\Psi_l^0| \,. \tag{4.35}$$

Let us define operator $\hat{R}^0$ via its spectral resolution:

$$\hat{R}^0 = \sum_{\substack{k \\ (k\neq i)}} \frac{1}{E_k^0 - E_i^0}|\Psi_k^0\rangle\langle\Psi_k^0| \,. \tag{4.36}$$

The restriction at the summation guarantees that the denominator does not become zero. (E_i^0 has been assumed to be nondegenerate.)

Let us determine the operator–product $\hat{R}^0(\hat{H}^0 - E_i^0)$ by using (4.35) and (4.36):

$$\begin{aligned}\hat{R}^0(\hat{H}^0 - E_i^0) &= \sum_{\substack{k \\ (k\neq i)}} \frac{1}{E_k^0 - E_i^0}|\Psi_k^0\rangle\langle\Psi_k^0| \sum_l (E_l^0 - E_i^0)|\Psi_l^0\rangle\langle\Psi_l^0| \\ &= \sum_{\substack{k \\ (k\neq i)}} \sum_l \frac{E_l^0 - E_i^0}{E_k^0 - E_i^0}|\Psi_k^0\rangle \underbrace{\langle\Psi_k^0|\Psi_l^0\rangle}_{\delta_{kl}}\langle\Psi_l^0| = \sum_{\substack{k \\ (k\neq i)}} |\Psi_k^0\rangle\langle\Psi_k^0| = \hat{Q}\,.\end{aligned} \tag{4.37}$$

This result indicates that in the subspace orthogonal to $|\Psi_i^0\rangle$ the operator $\hat{R}^0$ behaves like the inverse of the operator $\hat{H}^0 - E_i^0$; hence comes its name "reduced resolvent."

Corrections of the wave function

Let us consider the perturbational expansions of the wave function and of the energy

$$\begin{aligned}|\Psi_i\rangle &= |\Psi_i^0\rangle + \sum_{j=1}^{\infty} \lambda^j |\Psi_i^{(j)}\rangle \\ E_i &= E_i^0 + \sum_{k=1}^{\infty} \lambda^k E_i^{(k)}\end{aligned} \tag{4.38}$$

and use the intermediate normalization

$$\begin{aligned}\langle\Psi_i^0|\Psi_i^0\rangle &= 1 \\ \langle\Psi_i^0|\Psi_i^{(j)}\rangle &= 0 \qquad j \geq 1\,.\end{aligned} \tag{4.39}$$

$|\Psi_i^0\rangle$ is a solution of the unperturbed problem, i.e., $\hat{H}^0|\Psi_i^0\rangle = E_i^0|\Psi_i^0\rangle$; $\langle\Psi_i^0|\hat{H}^0 = E_i^0\langle\Psi_i^0|$. The intermediate normalization (4.39) leads, therefore, to the equalities

$$\langle\Psi_i^0|\hat{H}^0|\Psi_i^0\rangle = E_i^0 \tag{4.40}$$

and

$$\langle\Psi_i^0|\hat{H}^0|\Psi_i^{(j)}\rangle = 0\,. \tag{4.41}$$

We substitute the series expansion (4.38) into the Schrödinger equation:

$$\begin{aligned}&(\hat{H}^0 + \lambda\hat{W})\left[|\Psi_i^0\rangle + \sum_{j=1}^{\infty} \lambda^j |\Psi_i^{(j)}\rangle\right] \\ &\qquad = \left[E_i^0 + \sum_{k=1}^{\infty} \lambda^k E_i^{(k)}\right]\left[|\Psi_i^0\rangle + \sum_{l=1}^{\infty} \lambda^l |\Psi_i^{(l)}\rangle\right]\end{aligned} \tag{4.42}$$

and regroup the terms as

$$(\hat{H}^0 - E_i^0)\left[|\Psi_i^0\rangle + \sum_{j=1}^{\infty} \lambda^j |\Psi_i^{(j)}\rangle\right] \tag{4.43}$$

$$= \left[-\lambda\hat{W} + \sum_{k=1}^{\infty} \lambda^k E_i^{(k)}\right]\left[|\Psi_i^0\rangle + \sum_{l=1}^{\infty} \lambda^l |\Psi_i^{(l)}\rangle\right] .$$

The first term on the left-hand side vanishes because $(\hat{H}^0 - E_i^0)|\Psi_i^0\rangle = 0$. On the right-hand side we introduce a new summation index $j = l + 1$ (then $l = j - 1$) in the term containing the product of $-\lambda\hat{W}$ with the sum over l, whereas in the term containing the product of the two sums we turn to summation indices $j = k + l$ and k, instead of l and k summing independently (then $l = j - k$). We arrive at

$$\begin{aligned}(\hat{H}^0 - E_i^0)\sum_{j=1}^{\infty} \lambda^j |\Psi_i^{(j)}\rangle &= -\lambda\hat{W}|\Psi_i^0\rangle - \sum_{j=2}^{\infty} \lambda^j \hat{W}|\Psi_i^{(j-1)}\rangle \\ &+ \sum_{k=1}^{\infty} \lambda^k E_i^{(k)}|\Psi_i^0\rangle + \sum_{j=2}^{\infty}\sum_{k=1}^{j-1} \lambda^j E_i^{(k)}|\Psi_i^{(j-k)}\rangle .\end{aligned} \tag{4.44}$$

Let us now apply operator $\hat{R}^0$ to both sides of (4.44), utilizing the relationship (4.37) as well as the fact that $\hat{R}^0|\Psi_i^0\rangle = 0$. Also, it follows from the intermediate normalization (4.39) that $\hat{Q}|\Psi_i^{(j)}\rangle = |\Psi_i^{(j)}\rangle$ for any $j \geq 1$. We obtain

$$\begin{aligned}\sum_{j=1}^{\infty} \lambda^j |\Psi_i^{(j)}\rangle &= -\lambda\hat{R}^0\hat{W}|\Psi_i^0\rangle - \sum_{j=2}^{\infty} \lambda^j \hat{R}^0\hat{W}|\Psi_i^{(j-1)}\rangle \\ &+ \sum_{j=2}^{\infty}\sum_{k=1}^{j-1} \lambda^j E_i^{(k)}\hat{R}^0|\Psi_i^{(j-k)}\rangle .\end{aligned} \tag{4.45}$$

Now, we shall request the coefficients at the different powers of λ on the two sides of (4.44) to be equal. Comparing the coefficients at λ^1, we get

$$|\Psi_i^{(1)}\rangle = -\hat{R}^0\hat{W}|\Psi_i^0\rangle . \tag{4.46}$$

For λ^j ($j \geq 2$) we have:

$$|\Psi_i^{(j)}\rangle = -\hat{R}^0\hat{W}|\Psi_i^{(j-1)}\rangle + \sum_{k=1}^{j-1} E_i^{(k)}\hat{R}^0|\Psi_i^{(j-k)}\rangle . \tag{4.47}$$

It is easy to see the close analogy between these expression and those written in the compact matrix notations at the at the end of the Section 1.2; the same holds for the energy corrections discussed as follows.

By substituting the definition (4.36) of $\hat{R}^0$ into (4.46) and (4.47), we can easily obtain the expression for the coefficients $c_{ip}^{(j)}$ describing the expansion of the different wave function corrections $|\Psi_i^{(j)}\rangle$ in terms of the unperturbed functions, $|\Psi_i^{(j)}\rangle = \sum_{\substack{p \\ (p\neq i)}} c_{ip}^{(j)} |\Psi_p^0\rangle$. In fact

$$|\Psi_i^{(1)}\rangle = -\hat{R}^0\hat{W}|\Psi_i^0\rangle = -\sum_{\substack{j \\ (j\neq i)}} \frac{|\Psi_j^0\rangle\langle\Psi_j^0|\hat{W}|\Psi_i^0\rangle}{E_j^0 - E_i^0} = \sum_{\substack{j \\ (j\neq i)}} \frac{-W_{ji}}{E_j^0 - E_i^0}|\Psi_j^0\rangle \,, \tag{4.48}$$

which is the well-known expression. Similarly, for $j \geq 2$ we have

$$\begin{aligned}
|\Psi_i^{(j)}\rangle &= \sum_{\substack{p \\ (p\neq i)}} c_{ip}^{(j)}|\Psi_p^0\rangle = -\hat{R}^0\hat{W}|\Psi_i^{(j-1)}\rangle + \sum_{k=1}^{j-1} E_i^{(k)}\hat{R}^0|\Psi_i^{(j-k)}\rangle \\
&= -\sum_{\substack{p \\ (p\neq i)}} \frac{|\Psi_p^0\rangle\langle\Psi_p^0|}{E_p^0 - E_i^0}\,\hat{W} \sum_{\substack{l \\ (l\neq i)}} c_{il}^{(j-1)}|\Psi_l^0\rangle \\
&\qquad + \sum_{k=1}^{j-1} E_i^{(k)} \sum_{\substack{p \\ (p\neq i)}} \frac{|\Psi_p^0\rangle\langle\Psi_p^0|}{E_p^0 - E_i^0} \sum_{\substack{l \\ (l\neq i)}} c_{il}^{(j-k)}|\Psi_l^0\rangle \qquad (4.49) \\
&= -\sum_{\substack{p \\ (p\neq i)}} \sum_{\substack{l \\ (l\neq i)}} |\Psi_p^0\rangle \frac{W_{pl}c_{il}^{(j-1)}}{E_p^0 - E_i^0} \\
&\qquad + \sum_{\substack{p \\ (p\neq i)}} \sum_{\substack{l \\ (l\neq i)}} |\Psi_p^0\rangle\delta_{pl}\frac{1}{E_p^0 - E_i^0}\sum_{k=1}^{j-1} E_i^{(k)}c_{il}^{(j-k)} \\
&= \sum_{\substack{p \\ (p\neq i)}} |\Psi_p^0\rangle \frac{1}{E_p^0 - E_i^0}\left[-\sum_{\substack{l \\ (l\neq i)}} W_{pl}c_{il}^{(j-1)} + \sum_{k=1}^{j-1} E_i^{(k)}c_{ip}^{(j-k)}\right].
\end{aligned}$$

Thus we have recovered the expression of $c_{ip}^{(j)}$ derived in the previous section in an "algebraic" manner.

The energy corrections

Let us multiply (4.42) with $\langle\Psi_i^0|$, by taking into account the relationships (4.40), (4.41)

$$E_i^0 + \lambda\langle\Psi_i^0|\hat{W}|\Psi_i^0\rangle + \sum_{j=1}^{\infty}\lambda^{j+1}\langle\Psi_i^0|\hat{W}|\Psi_i^{(j)}\rangle = E_i^0 + \sum_{k=1}^{\infty}\lambda^k E_i^{(k)} \quad (4.50)$$

In the sum on the left-hand side we turn to the summation index $k = j + 1$ (so $j = k - 1$):

$$\lambda\langle\Psi_i^0|\hat{W}|\Psi_i^0\rangle + \sum_{k=2}^{\infty}\lambda^k\langle\Psi_i^0|\hat{W}|\Psi_i^{(k-1)}\rangle = \sum_{k=1}^{\infty}\lambda^k E_i^{(k)} \,. \quad (4.51)$$

Requesting the coefficients at the different powers of λ on the two sides to be equal, we get for λ^1

$$E_i^{(1)} = \langle\Psi_i^0|\hat{W}|\Psi_i^0\rangle = W_{ii} \quad (4.52)$$

—the well-known result. For λ^k ($k \geq 2$) the result is

$$E_i^{(k)} = \langle\Psi_i^0|\hat{W}|\Psi_i^{(k-1)}\rangle \,. \quad (4.53)$$

By substituting here the expansion of $|\Psi_i^{(k-1)}\rangle$ in terms of the zero-order functions, we recover the "algebraic" result in (4.24):

$$E_i^{(k)} = \langle\Psi_i^0|\hat{W}|\sum_{\substack{p\\(p\neq i)}} c_{ip}^{(k-1)}\Psi_p^0\rangle = \sum_{\substack{p\\(p\neq i)}} c_{ip}^{(k-1)}W_{ip} \,. \quad (4.54)$$

Invariance

According to the previous derivations, the perturbational corrections of the wave function and of the energy can be generated by using two operators: the perturbation $\hat{W}$ and the reduced resolvent $\hat{R}^0$. This permits to prove easily that the results given by the Rayleigh–Schrödinger perturbation theory are invariant under unitary transformations between degenerate zero-order solutions $|\Psi_k^0\rangle$, if any. (However, we have to stress that the ith solution $|\Psi_i^0\rangle$, the perturbation of which we are actually considering, has been assumed nondegenerate, so the degeneracy may involve only some other zero-order functions. The case when one deals with a degenerate unperturbed wave function will be considered later, in Section 3)

Operator $\hat{W}$ is an entity independent of the transformations of the basis function, so we have only to consider the reduced resolvent $\hat{R}^0$. Inspection of its definition permits to conclude that it is invariant, as well. In fact, let us

assume that there are m zero-order solutions $\Psi_l^0, \Psi_{l+1}^0, ..., \Psi_{l+m-1}^0$ belonging to the same unperturbed energy E_l^0. Then these m states will contribute to the reduced resolvent (4.36) by

$$\sum_{k=l}^{l+m-1} \frac{1}{E_l^0 - E_i^0} |\Psi_k^0\rangle\langle\Psi_k^0| = \frac{1}{E_l^0 - E_i^0} \sum_{k=l}^{l+m-1} |\Psi_k^0\rangle\langle\Psi_k^0| \,. \tag{4.55}$$

In the right-hand side we see the projector into the degenerate subspace; it is invariant under the unitary transformations between the orbitals involved (see Appendix VII).

1.4. Wigner's $2n + 1$ theorem

1.4.1. The accuracy of the energy expectation value calculated by the nth order perturbational wave function

Theorem:

The expectation value of the energy calculated by using the nth order perturbational wave function is accurate up to terms of order $2n + 1$, inclusively. (It differs from the exact energy by terms of order $2n + 2$ and higher.)

Proof:

The exact ground state wave function Ψ_0 can be written as

$$\Psi_0 = \Psi_0^{[n]} + \chi^{(n+1)} \tag{4.56}$$

where

$$\Psi_0^{[n]} = \sum_{l=0}^{n} \lambda^l \Psi_0^{(l)} \tag{4.57}$$

is the perturbational wave function obtained up to the nth order of PT inclusively, whereas $\chi^{(n+1)}$ is its error:

$$\chi^{(n+1)} = \sum_{k=n+1}^{\infty} \lambda^k \Psi_0^{(k)} \sim \mathcal{O}(\lambda^{n+1}) \,. \tag{4.58}$$

One can rewrite (4.56) as

$$\Psi_0^{[n]} = \Psi_0 - \chi^{(n+1)} \,. \tag{4.59}$$

The expectation value of energy calculated by using $\Psi_0^{[n]}$ is, therefore,

$$E = \frac{\langle\Psi_0^{[n]}|\hat{H}|\Psi_0^{[n]}\rangle}{\langle\Psi_0^{[n]}|\Psi_0^{[n]}\rangle} = \frac{\langle\Psi_0 - \chi^{(n+1)}|\hat{H}|\Psi_0 - \chi^{(n+1)}\rangle}{\langle\Psi_0 - \chi^{(n+1)}|\Psi_0 - \chi^{(n+1)}\rangle} \tag{4.60}$$

Now, Ψ_0 is the exact wave function, so

$$\hat{H}^0\Psi_0 = E_0\Psi_0 \ , \tag{4.61}$$

where E_0 is the *exact* energy. We also have, owing to (4.58) and (4.61)

$$\begin{aligned}
\langle\chi^{(n+1)}|\hat{H}|\Psi^0\rangle &= E_0\langle\chi^{(n+1)}|\Psi^0\rangle \\
\langle\Psi_0|\hat{H}|\chi^{(n+1)}\rangle &= E_0\langle\Psi_0|\chi^{(n+1)}\rangle \\
\langle\chi^{(n+1)}|\hat{H}|\chi^{(n+1)}\rangle &= \mathcal{O}(2n+2) \\
\langle\chi^{(n+1)}|\chi^{(n+1)}\rangle &= \mathcal{O}'(2n+2)
\end{aligned} \tag{4.62}$$

where $\mathcal{O}(2n+2)$, $\mathcal{O}'(2n+2)$, etc. denote quantities of order λ^{2n+2}. (There is no reason nor need to assume $\chi^{(n+1)}$ to be orthogonal to Ψ_0.)

Substituting the results from (4.61) and (4.62) into (4.60) we get

$$\begin{aligned}
E &= \frac{E_0(\langle\Psi_0|\Psi_0\rangle - \langle\chi^{(n+1)}|\Psi_0\rangle - \langle\Psi_0|\chi^{(n+1)}\rangle) + \mathcal{O}(2n+2)}{\langle\Psi_0|\Psi_0\rangle - \langle\chi^{(n+1)}|\Psi_0\rangle - \langle\Psi_0|\chi^{(n+1)}\rangle + \mathcal{O}'(2n+2)} \\
&= E_0 + \mathcal{O}''(2n+2) \ .
\end{aligned} \tag{4.63}$$

Q.E.D.

1.4.2. Computing the exact energy corrections up to order $2n+1$ by using the first n wave function corrections

Let, as usual

$$\hat{H} = \hat{H}^0 + \lambda\hat{W} \tag{4.64}$$

$$\Psi_i = \Psi_i^0 + \sum_{l=1}^{\infty}\lambda^l\Psi_i^{(l)} \tag{4.65}$$

$$E_i = E_i^0 + \sum_{j=1}^{\infty}\lambda^j E_i^{(j)} \ , \tag{4.66}$$

where $\hat{H}^0\Psi_i^0 = E_i^0\Psi_i^0$, and we use the intermediate normalization $\langle\Psi_i^0|\Psi_i^0\rangle = 1$; $\langle\Psi_i^0|\Psi_i^{(l)}\rangle = 0$ $(l \geq 1)$.

We substitute expressions (4.64)–(4.66) into the Schrödinger equation $\hat{H}\Psi_i = E_i\Psi_i$. As in the following derivations every expression is related to the ith level, we shall omit subscripts i for the sake of simplicity. We have

$$(\hat{H}^0 + \lambda\hat{W})\left[\Psi^0 + \sum_{l=1}^{\infty}\lambda^l\Psi^{(l)}\right] = \left[E^0 + \sum_{k=1}^{\infty}\lambda^k E^{(k)}\right]\left[\Psi^0 + \sum_{j=1}^{\infty}\lambda^j\Psi^{(j)}\right] . \tag{4.67}$$

Let us collect the coefficients at λ^p ($p \geq 1$) on both sides of (4.67). In the product of the two sums on the right-hand side we should take into account all terms for which $k + j = p$. Therefore, we may write $k = p - j$ and should perform a summation for $j = 1$ to $j = p - 1$. We get

$$\hat{H}^0 \Psi^{(p)} + \hat{W} \Psi^{(p-1)} = E^{(p)} \Psi^0 + E^0 \Psi^{(p)} + \sum_{j=1}^{p-1} E^{(p-j)} \Psi^{(j)} \ . \tag{4.68}$$

(The last term is absent, of course, if $p = 1$.)

By multiplying (4.68) with Ψ^{0*} and integrating, we get the known result

$$E^{(p)} = \langle \Psi^0 | \hat{W} | \Psi^{(p-1)} \rangle \ . \tag{4.69}$$

We are going to show that one can reduce the order $p - 1$ in the "ket" on the right-hand side of (4.69) at the expense of the appearance of some wave function correction in the "bra" as well as of some additional terms containing lower order wave function and energy corrections.

Let us bring $E^0 \Psi^{(p)}$ to the left-hand-side in (4.68):

$$(\hat{H}^0 - E^0) \Psi^{(p)} + \hat{W} \Psi^{(p-1)} = E^{(p)} \Psi^0 + \sum_{j=1}^{p-1} E^{(p-j)} \Psi^{(j)} \ . \tag{4.70}$$

This is a general result, valid for every $p \geq 1$. (In the case of $p = 1$, the last sum should be omitted.)

Let us multiply relationship (4.70) by $\Psi^{(1)*}$ and integrate ($\langle \Psi^{(1)} | \Psi^0 \rangle = 0$):

$$\langle \Psi^{(1)} | \hat{H}^0 - E^0 | \Psi^{(p)} \rangle + \langle \Psi^{(1)} | \hat{W} | \Psi^{(p-1)} \rangle = \sum_{j=1}^{p-1} E^{(p-j)} \langle \Psi^{(1)} | \Psi^{(j)} \rangle \ . \tag{4.71}$$

Let us write out (4.70) also for $p = 1$ (the sum is to be omitted):

$$(\hat{H}^0 - E^0) \Psi^{(1)} + \hat{W} \Psi^0 = E^{(1)} \Psi^0 \ , \tag{4.72}$$

multiply by $\Psi^{(p)*}$ and integrate ($\langle \Psi^{(p)} | \Psi^0 \rangle = 0$):

$$\langle \Psi^{(p)} | \hat{H}^0 - E^0 | \Psi^{(1)} \rangle + \langle \Psi^{(p)} | \hat{W} | \Psi^0 \rangle = 0 \ . \tag{4.73}$$

Taking the complex conjugate of (4.73) and regrouping the terms,

$$\langle \Psi^{(1)} | \hat{H}^0 - E^0 | \Psi^{(p)} \rangle = - \langle \Psi^0 | \hat{W} | \Psi^{(p)} \rangle \ . \tag{4.74}$$

By substituting (4.74) into (4.71) we get

$$- \langle \Psi^0 | \hat{W} | \Psi^{(p)} \rangle + \langle \Psi^{(1)} | \hat{W} | \Psi^{(p-1)} \rangle = \sum_{j=1}^{p-1} E^{(p-j)} \langle \Psi^{(1)} | \Psi^{(j)} \rangle \tag{4.75}$$

that is, taking into account (4.69):

$$E^{(p+1)} = \langle \Psi^0 | \hat{W} | \Psi^{(p)} \rangle = \langle \Psi^{(1)} | \hat{W} | \Psi^{(p-1)} \rangle - \sum_{j=1}^{p-1} E^{(p-j)} \langle \Psi^{(1)} | \Psi^{(j)} \rangle \quad (4.76)$$

Equation (4.76) indicates that we do not need any more $\Psi^{(p)}$ for computing $E^{(p+1)}$: the wave function correction of highest order occurring in the right-hand side of (4.76) is $\Psi^{(p-1)}$.

We may repeat this procedure. We write out (4.70) for $p-1$ instead of p. (As in the derivation of (4.70) it was assumed $p \geq 1$, here we should request $p \geq 2$.)

$$(\hat{H}^0 - E^0)\Psi^{(p-1)} + \hat{W}\Psi^{(p-2)} = E^{(p-1)}\Psi^0 + \sum_{j=1}^{p-2} E^{(p-1-j)}\Psi^{(j)} \ . \quad (4.77)$$

By multiplying (4.77) with $\Psi^{(2)*}$ and integrating, we get $(\langle \Psi^{(2)} | \Psi^0 \rangle = 0)$:

$$\langle \Psi^{(2)} | \hat{H}^0 - E^0 | \Psi^{(p-1)} \rangle + \langle \Psi^{(2)} | \hat{W} | \Psi^{(p-2)} \rangle = \sum_{j=1}^{p-2} E^{(p-1-j)} \langle \Psi^{(2)} | \Psi^{(j)} \rangle \ . \quad (4.78)$$

Let us write out (4.70) also for $p = 2$:

$$(\hat{H}^0 - E^0)\Psi^{(2)} + \hat{W}\Psi^{(1)} = E^{(2)}\Psi^0 + E^{(1)}\Psi^{(1)} \ , \quad (4.79)$$

multiply it by $\Psi^{(p-1)*}$ and integrate $(\langle \Psi^{(p-1)} | \Psi^0 \rangle = 0$, because $p \geq 2)$. We get

$$\langle \Psi^{(p-1)} | \hat{H}^0 - E^0 | \Psi^{(2)} \rangle + \langle \Psi^{(p-1)} | \hat{W} | \Psi^{(1)} \rangle = E^{(1)} \langle \Psi^{(p-1)} | \Psi^{(1)} \rangle \ . \quad (4.80)$$

By taking the complex conjugate of (4.80) and rearranging we have

$$\langle \Psi^{(2)} | \hat{H}^0 - E^0 | \Psi^{(p-1)} \rangle = -\langle \Psi^{(1)} | \hat{W} | \Psi^{(p-1)} \rangle + E^{(1)} \langle \Psi^{(1)} | \Psi^{(p-1)} \rangle \ . \quad (4.81)$$

By substituting (4.81) into (4.78) and regrouping we get

$$\begin{aligned} \langle \Psi^{(1)} | \hat{W} | \Psi^{(p-1)} \rangle &= \langle \Psi^{(2)} | \hat{W} | \Psi^{(p-2)} \rangle + E^{(1)} \langle \Psi^{(1)} | \Psi^{(p-1)} \rangle \\ &\quad - \sum_{j=1}^{p-2} E^{(p-1-j)} \langle \Psi^{(2)} | \Psi^{(j)} \rangle \ . \end{aligned} \quad (4.82)$$

We substitute (4.82) into the right-hand side of (4.76):

$$\begin{aligned} E^{(p+1)} &= \langle \Psi^0 | \hat{W} | \Psi^{(p)} \rangle = \langle \Psi^{(2)} | \hat{W} | \Psi^{(p-2)} \rangle + E^{(1)} \langle \Psi^{(1)} | \Psi^{(p-1)} \rangle \\ &\quad - \sum_{j=1}^{p-2} E^{(p-1-j)} \langle \Psi^{(2)} | \Psi^{(j)} \rangle - \sum_{j=1}^{p-1} E^{(p-j)} \langle \Psi^{(1)} | \Psi^{(j)} \rangle \ . \end{aligned} \quad (4.83)$$

In the last sum we separate out the term $j = p - 1$, which cancels the second term on the right-hand side, and we combine the remaining two sums into one:

$$\langle\Psi^0|\hat{W}|\Psi^{(p)}\rangle = \langle\Psi^{(2)}|\hat{W}|\Psi^{(p-2)}\rangle - \sum_{j=1}^{p-2}\sum_{l=1}^{2} E^{(p+1-l-j)}\langle\Psi^{(l)}|\Psi^{(j)}\rangle \ . \quad (4.84)$$

Considering (4.76) and (4.84), we may formulate the following:

Proposition:

$$\langle\Psi^0|\hat{W}|\Psi^{(p)}\rangle = \langle\Psi^{(k)}|\hat{W}|\Psi^{(p-k)}\rangle - \sum_{j=1}^{p-k}\sum_{l=1}^{k} E^{(p+1-l-j)}\langle\Psi^{(l)}|\Psi^{(j)}\rangle \quad (4.85)$$

for every k in the range $1 < k < p$.

Proof by induction:

We assume that the equality (4.85) is valid for some value k, and repeat again the procedure performed already twice.

Let us write out equality (4.70), substituting the value $p - k$ instead of p. (We assume $p > k$; $p - k > 0$.) We get

$$(\hat{H}^0 - E^0)\Psi^{(p-k)} + \hat{W}\Psi^{(p-k-1)} = E^{(p-k)}\Psi^0 + \sum_{j=1}^{p-k-1} E^{(p-k-j)}\Psi^{(j)} \ . \quad (4.86)$$

Multiplying with $\Psi^{(k+1)*}$ and integrating $(\langle\Psi^{(k+1)}|\Psi^0\rangle = 0)$:

$$\langle\Psi^{(k+1)}|\hat{H}^0 - E^0|\Psi^{(p-k)}\rangle + \langle\Psi^{(k+1)}|\hat{W}|\Psi^{(p-k-1)}\rangle = \sum_{j=1}^{p-k-1} E^{(p-k-j)}\langle\Psi^{(k+1)}|\Psi^{(j)}\rangle \quad (4.87)$$

Now, we write out (4.70) also for $p = k + 1$,

$$(\hat{H}^0 - E^0)\Psi^{(k+1)} + \hat{W}\Psi^{(k)} = E^{(k+1)}\Psi^0 + \sum_{j=1}^{k} E^{(k+1-j)}\Psi^{(j)} \quad (4.88)$$

multiply with $\Psi^{(p-k)*}$ and integrate $(\langle\Psi^{(p-k)}|\Psi^0\rangle = 0$, as $p - k > 0)$:

$$\langle\Psi^{(p-k)}|\hat{H}^0 - E^0|\Psi^{(k+1)}\rangle + \langle\Psi^{(p-k)}|\hat{W}|\Psi^{(k)}\rangle = \sum_{j=1}^{k} E^{(k+1-j)}\langle\Psi^{(p-k)}|\Psi^{(j)}\rangle \ . \quad (4.89)$$

We take the complex conjugate of (4.89)

$$\langle\Psi^{(k+1)}|\hat{H}^0-E^0|\Psi^{(p-k)}\rangle+\langle\Psi^{(k)}|\hat{W}|\Psi^{(p-k)}\rangle = \sum_{j=1}^{k} E^{(k+1-j)}\langle\Psi^{(j)}|\Psi^{(p-k)}\rangle \ , \tag{4.90}$$

extract (4.87) from (4.90), and move one of the terms to the right-hand side:

$$\begin{aligned}\langle\Psi^{(k)}|\hat{W}|\Psi^{(p-k)}\rangle &= \langle\Psi^{(k+1)}|\hat{W}|\Psi^{(p-k-1)}\rangle + \sum_{j=1}^{k} E^{(k+1-j)}\langle\Psi^{(j)}|\Psi^{(p-k)}\rangle \\ &\quad - \sum_{j=1}^{p-k-1} E^{(p-k-j)}\langle\Psi^{(k+1)}|\Psi^{(j)}\rangle \ .\end{aligned} \tag{4.91}$$

Now, we substitute this expression obtained for $\langle\Psi^{(k)}|\hat{W}|\Psi^{(p-k)}\rangle$ into (4.85):

$$\begin{aligned}\langle\Psi^{0}|\hat{W}|\Psi^{(p)}\rangle &= \langle\Psi^{(k+1)}|\hat{W}|\Psi^{(p-k-1)}\rangle + \sum_{j=1}^{k} E^{(k+1-j)}\langle\Psi^{(j)}|\Psi^{(p-k)}\rangle \\ &\quad - \sum_{j=1}^{p-k-1} E^{(p-k-j)}\langle\Psi^{(k+1)}|\Psi^{(j)}\rangle \\ &\quad - \sum_{j=1}^{p-k}\sum_{l=1}^{k} E^{(p+1-l-j)}\langle\Psi^{(l)}|\Psi^{(j)}\rangle \ .\end{aligned} \tag{4.92}$$

In the double sum we separate out the term $j = p - k$, corresponding to the upper limit of the outer summation $[p + 1 - l - (p - k) = k + 1 - l]$:

$$\begin{aligned}\langle\Psi^{0}|\hat{W}|\Psi^{(p)}\rangle &= \langle\Psi^{(k+1)}|\hat{W}|\Psi^{(p-k-1)}\rangle + \sum_{j=1}^{k} E^{(k+1-j)}\langle\Psi^{(j)}|\Psi^{(p-k)}\rangle \\ &\quad - \sum_{j=1}^{p-k-1} E^{(p-k-j)}\langle\Psi^{(k+1)}|\Psi^{(j)}\rangle \\ &\quad - \sum_{j=1}^{p-k-1}\sum_{l=1}^{k} E^{(p+1-l-j)}\langle\Psi^{(l)}|\Psi^{(j)}\rangle \\ &\quad - \sum_{l=1}^{k} E^{(k+1-l)}\langle\Psi^{(l)}|\Psi^{(p-k)}\rangle \ .\end{aligned} \tag{4.93}$$

The first and last sums on the right-hand side cancel (they differ only in the notation of the summation index); the remaining two sums can be combined

into one: the former gives the additional term $l = k + 1$ to the latter $[p + 1 - (k + 1) - j = p - k - j]$. We get

$$\begin{aligned}\langle\Psi^0|\hat{W}|\Psi^{(p)}\rangle &= \langle\Psi^{(k+1)}|\hat{W}|\Psi^{[p-(k+1)]}\rangle \\ &\quad - \sum_{j=1}^{p-(k+1)}\sum_{l=1}^{k+1} E^{(p+1-l-j)}\langle\Psi^{(l)}|\Psi^{(j)}\rangle\,. \qquad (4.94)\end{aligned}$$

That means that assuming the validity of (4.85) for some value k, we have derived it for $k + 1$. Because, as we have seen, (4.85) is valid for $k = 1$ and $k = 2$, it must be valid for every $k < p$.
Q.E.D.

As $\langle\Psi^0|\hat{W}|\Psi^{(p)}\rangle = E^{(p+1)}$, we may write (4.85) as

$$E^{(p+1)} = \langle\Psi^{(k)}|\hat{W}|\Psi^{(p-k)}\rangle - \sum_{j=1}^{p-k}\sum_{l=1}^{k} E^{p+1-l-j}\langle\Psi^{(l)}|\Psi^{(j)}\rangle \qquad (4.95)$$

and one may choose any value of k in the range $0 < k < p$. (If $k = 0$ or $k = p$, then the sum on the right-hand side should be omitted—the limits of summation also become meaningless in these cases.)

Expressions for the energy corrections $E^{(2n)}$ and $E^{(2n+1)}$

Let us assume that in (4.95) $p + 1 = 2n$ (an even number), i.e., $p = 2n - 1$ (odd), and let us choose $k = n - 1$. Then $p - k = 2n - 1 - (n - 1) = n$, and

$$E^{(2n)} = \langle\Psi^{(n-1)}|\hat{W}|\Psi^{(n)}\rangle - \sum_{j=1}^{n}\sum_{l=1}^{n-1} E^{(2n-l-j)}\langle\Psi^{(l)}|\Psi^{(j)}\rangle\,. \qquad (4.96)$$

Now, let us assume that $p+1 = 2n+1$ (an odd number), i.e., $p = 2n$ (even), and let us choose $k = n$. We get ($p - k = 2n - n = n$):

$$E^{(2n+1)} = \langle\Psi^{(n)}|\hat{W}|\Psi^{(n)}\rangle - \sum_{j=1}^{n}\sum_{l=1}^{n} E^{(2n+1-l-j)}\langle\Psi^{(l)}|\Psi^{(j)}\rangle\,. \qquad (4.97)$$

One can see that the highest order wave function correction entering (4.96) and (4.97) is $\Psi^{(n)}$. We do need energy corrections of higher orders, as well, but if we have got the wave function corrections $\Psi^{(1)}, \Psi^{(2)}, \ldots, \Psi^{(n)}$, then the necessary energy corrections can be obtained successively, one by one, by the repeated use of (4.96) and (4.97), without using any wave function correction

of order higher than $\Psi^{(n)}$:

$$
\begin{aligned}
n=0: \quad E^{(1)} &= \langle \Psi^0 | \hat{W} | \Psi^0 \rangle \\
n=1: \quad E^{(2)} &= \langle \Psi^0 | \hat{W} | \Psi^{(1)} \rangle \\
E^{(3)} &= \langle \Psi^{(1)} | \hat{W} | \Psi^{(1)} \rangle - E^{(1)} \langle \Psi^{(1)} | \Psi^{(1)} \rangle \qquad (4.98) \\
n=2: \quad E^{(4)} &= \langle \Psi^{(1)} | \hat{W} | \Psi^{(2)} \rangle - \sum_{j=1}^{2} \sum_{l=1}^{1} E^{(4-l-j)} \langle \Psi^{(l)} | \Psi^{(j)} \rangle \\
&= \langle \Psi^{(1)} | \hat{W} | \Psi^{(2)} \rangle - E^{(2)} \langle \Psi^{(1)} | \Psi^{(1)} \rangle - E^{(1)} \langle \Psi^{(1)} | \Psi^{(2)} \rangle \\
E^{(5)} &= \langle \Psi^{(2)} | \hat{W} | \Psi^{(2)} \rangle - \sum_{j=1}^{2} \sum_{l=1}^{2} E^{(5-l-j)} \langle \Psi^{(l)} | \Psi^{(j)} \rangle \\
&= \langle \Psi^{(2)} | \hat{W} | \Psi^{(2)} \rangle - E^{(3)} \langle \Psi^{(1)} | \Psi^{(1)} \rangle - E^{(2)} \langle \Psi^{(2)} | \Psi^{(1)} \rangle \\
&\quad - E^{(2)} \langle \Psi^{(1)} | \Psi^{(2)} \rangle - E^{(1)} \langle \Psi^{(2)} | \Psi^{(2)} \rangle
\end{aligned}
$$

and so on.

2. Variational–perturbational method: The Hylleraas functional

Let us again consider the perturbed Schrödinger equation $(\hat{H}^0 + \lambda \hat{W})\Psi = E\Psi$, and the series expansion of the zero (lowest) energy level according to the powers of the perturbational parameter λ. (As all quantities belong to the same ground state, we shall omit the subscript "0".)

$$E = E^0 + \lambda E^{(1)} + \lambda^2 E^{(2)} + \ldots \qquad (4.99)$$

As it is discussed in Section 1.1, λ can be considered—at least formally—a continuous variable.

The expansion of the energy as a power series of λ must be unique. That means that the expansion coefficients $E^0, E^{(1)}, E^{(2)}, \ldots$ must not depend on the way we derive them. (The convergence of the perturbational series is assumed, of course.) We may obtain the expansion in (4.99) not only by the methods discussed in the previous sections, but also by using the following technique: we calculate the expectation value of the energy for the exact wave function presented as an infinite series in terms of λ, and then re-expand this expectation value as a power series of λ.

These considerations indicate that (4.99) may also be considered as the series expansion of the expectation value of the energy, pertinent to the exact wave function; the latter represents a function of both the parameter λ and of the perturbational expansion coefficients.

According to the variation principle, the ground-state energy represents the lowest possible expectation value of the energy, $E = min$. Therefore, the

energy as a function of the wave function expansion coefficients should be minimal for any value of λ, for which the perturbational series converges. As λ is, at least formally, a continuous variable, the energy represented in (4.99) should be minimal for all values of λ. This is possible only if each $E^{(i)}$ separately reaches a minimum or is a constant. That means that for a given λ, one can consider the perturbational expansion coefficients of the wave function as unknown quantities and determine them by minimizing the perturbational energy corrections $E^{(i)}$ one by one. (In the limit of $\lambda \to 0$ the higher order energy corrections proportional to some λ^j are negligible as compared with any lower order ones proportional to λ^i, where $i < j$. Therefore, one should start with minimizing the lowest order energy corrections and proceed with the higher order ones by keeping fixed the expansion coefficients already determined.)

In the case of E^0 and $E^{(1)}$ there are no free wave function parameters, but $E^{(2)}$ depends on the first-order wave function (wave function correction). Therefore, the true first-order wave function is that for which the second-order energy contribution reaches its minimum.

It is to be noted that all the energy corrections of odd order should be constant in the sense that they are determined uniquely by the wave function parameters optimized for the energy corrections of lower (even) orders. Otherwise, obtaining a minimum of the term $\lambda^{2i+1}E^{(2i+1)}$ for a positive λ value, we would have a maximum for a negative one, in contradiction with the variation principle. Of course, this conclusion is in line with Wigner's "$2n+1$" theorem discussed in Section 1.4.

Now, we shall turn to the optimization of the first-order wave function. Let us consider the wave function $\Psi^0 + \lambda\chi$, where Ψ^0 is the zero-order wave function satisfying the zero-order Schrödinger equation $\hat{H}^0\Psi^0 = E^0\Psi^0$, and χ is the first-order wave function that is to be optimized. ($\hat{H}^0$ is Hermitian, therefore E^0 is real—nothing else needs to be assumed at this point.)

We expand the expectation value of the energy up to terms of second order (inclusively):

$$
\begin{aligned}
E &= \frac{\langle \Psi^0 + \lambda\chi | \hat{H}^0 + \lambda\hat{W} | \Psi^0 + \lambda\chi \rangle}{\langle \Psi^0 + \lambda\chi | \Psi^0 + \lambda\chi \rangle} \qquad (4.100) \\
&= \Big\{ \langle \Psi^0 | \hat{H}^0 | \Psi^0 \rangle + \lambda \langle \chi | \hat{H}^0 | \Psi^0 \rangle + \lambda \langle \Psi^0 | \hat{W} | \Psi^0 \rangle + \lambda \langle \Psi^0 | \hat{H}^0 | \chi \rangle \\
&\quad + \lambda^2 \langle \chi | \hat{W} | \Psi^0 \rangle + \lambda^2 \langle \chi | \hat{H}^0 | \chi \rangle + \lambda^2 \langle \Psi^0 | \hat{W} | \chi \rangle + \mathcal{O}(\lambda^3) \Big\} \\
&\quad \times \left\{ \langle \Psi^0 | \Psi^0 \rangle \left[1 + \lambda \frac{\langle \chi | \Psi^0 \rangle + \langle \Psi^0 | \chi \rangle}{\langle \Psi^0 | \Psi^0 \rangle} + \lambda^2 \frac{\langle \chi | \chi \rangle}{\langle \Psi^0 | \Psi^0 \rangle} \right] \right\}^{-1} .
\end{aligned}
$$

We utilize the zero-order equation and the known expansion $(1+x)^{-1} = 1 - x + x^2 - \ldots$ and get, keeping only terms up to second order:

$$\begin{aligned} E &= \frac{1}{\langle\Psi^0|\Psi^0\rangle}\Big\{E^0\langle\Psi^0|\Psi^0\rangle + \lambda E^0\left(\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle\right) \\ &\quad +\lambda\langle\Psi^0|\hat{W}|\Psi^0\rangle + \lambda^2\left(\langle\chi|\hat{W}|\Psi^0\rangle + \langle\Psi^0|\hat{W}|\chi\rangle\right) + \lambda^2\langle\chi|\hat{H}^0|\chi\rangle\Big\} \\ &\quad \times\Big\{1 - \lambda\frac{\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle}{\langle\Psi^0|\Psi^0\rangle} + \lambda^2\frac{(\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle)^2}{\langle\Psi^0|\Psi^0\rangle^2} \\ &\quad -\lambda^2\frac{\langle\chi|\chi\rangle}{\langle\Psi^0|\Psi^0\rangle} + \ldots\Big\} + \mathcal{O}(3)\,. \end{aligned} \tag{4.101}$$

By performing the multiplications, keeping again only terms up to second order, we have

$$\begin{aligned} E &= \frac{1}{\langle\Psi^0|\Psi^0\rangle}\Big\{E^0\langle\Psi^0|\Psi^0\rangle + \lambda E^0\left(\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle\right) \\ &\quad +\lambda\langle\Psi^0|\hat{W}|\Psi^0\rangle + \lambda^2\left(\langle\chi|\hat{W}|\Psi^0\rangle + \langle\Psi^0|\hat{W}|\chi\rangle + \langle\chi|\hat{H}^0|\chi\rangle\right) \\ &\quad -\lambda\frac{E^0\langle\Psi^0|\Psi^0\rangle(\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle)}{\langle\Psi^0|\Psi^0\rangle} - \lambda^2\frac{E^0(\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle)^2}{\langle\Psi^0|\Psi^0\rangle} \\ &\quad -\lambda^2\frac{\langle\Psi^0|\hat{W}|\Psi^0\rangle(\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle)}{\langle\Psi^0|\Psi^0\rangle} \\ &\quad +\lambda^2\frac{E^0\langle\Psi^0|\Psi^0\rangle(\langle\chi|\Psi^0\rangle + \langle\Psi^0|\chi\rangle)^2}{\langle\Psi^0|\Psi^0\rangle^2} - \lambda^2E^0\langle\chi|\chi\rangle\Big\} + \mathcal{O}(3)\,. \end{aligned} \tag{4.102}$$

The second term in the braces cancels with the fifth, and the sixth with the eight. Taking into account that the first-order energy is

$$E^{(1)} = \frac{\langle\Psi^0|\hat{W}|\Psi^0\rangle}{\langle\Psi^0|\Psi^0\rangle}\,, \tag{4.103}$$

equation (4.102) reduces to

$$\begin{aligned} E &= E^0 + \lambda E^{(1)} + \frac{\lambda^2}{\langle\Psi^0|\Psi^0\rangle}\Big[\langle\chi|\hat{W} - E^{(1)}|\Psi^0\rangle \\ &\quad +\langle\Psi^0|\hat{W} - E^{(1)}|\chi\rangle + \langle\chi|\hat{H}^0 - E^0|\chi\rangle\Big] + \mathcal{O}(3)\,. \end{aligned} \tag{4.104}$$

That means

$$E = E^0 + \lambda E^{(1)} + \lambda^2 J_2 + \mathcal{O}(3)\,, \tag{4.105}$$

where, in the case of Ψ^0 normalized to unity

$$J_2 = \langle\chi|\hat{W} - E^{(1)}|\Psi^0\rangle + \langle\Psi^0|\hat{W} - E^{(1)}|\chi\rangle + \langle\chi|\hat{H}^0 - E^0|\chi\rangle \quad (4.106)$$

is the (second-order) Hylleraas functional, the minimum of which gives the second-order perturbational energy correction

$$E^{(2)} = \min(J_2) \ . \quad (4.107)$$

Equation (4.107) represents a variational problem. The necessary condition of the minimum is that the variation of J_2 should vanish:

$$\delta J_2 = 0 \ . \quad (4.108)$$

Performing the variation of J_2 in (4.106), we get

$$\langle\delta\chi|\hat{W} - E^{(1)}|\Psi^0\rangle + \langle\delta\chi|\hat{H}^0 - E^0|\chi\rangle + c.c. = 0 \ . \quad (4.109)$$

As $\delta\chi$ may be complex and contains an arbitrary phase factor, (4.109) can be fulfilled for every $\delta\chi$ only if the term written explicitly and its complex conjugate (*c.c.*) vanish separately:

$$\langle\delta\chi|\hat{W} - E^{(1)}|\Psi^0\rangle + \langle\delta\chi|\hat{H}^0 - E^0|\chi\rangle = 0 \ . \quad (4.110)$$

If we use the intermediate normalization $\langle\Psi^0|\Psi^0\rangle = 1$; $\langle\Psi^0|\chi\rangle = 0$, and expand the first-order wave function χ in terms of the eigenvectors Ψ_i^0 $(i \neq 0)$ of the unperturbed Hamiltonian

$$\chi = \sum_{i=1} c_i \Psi_i^0 \ , \quad (4.111)$$

then the permitted variations $\delta\chi$ are the functions of the type

$$\delta\chi = \eta\Psi_j^0 \ ; \qquad j = 1, 2, \ldots \quad (4.112)$$

where $\eta \to 0$ is an arbitrary variational parameter (we could also denote it δc_j as used in Section 1 of Chapter 3). By substituting (4.112) into (4.110) and omitting the common factor η^*, we get

$$\langle\Psi_j^0|\hat{W} - E^{(1)}|\Psi^0\rangle + \langle\Psi_j^0|\hat{H}^0 - E^0|\sum_{i=1} c_i\Psi_i^0\rangle = 0 \ . \quad (4.113)$$

Taking into account that $\langle\Psi_j^0|\Psi^0\rangle = 0$ $(j \neq 0)$, $\hat{H}^0\Psi_i^0 = E_i^0\Psi_i^0$, and $\langle\Psi_j^0|\Psi_i^0\rangle = \delta_{ij}$, (4.113) readily reduces to

$$W_{j0} + (E_j^0 - E^0)c_j = 0 \ , \quad (4.114)$$

where $W_{j0} = \langle \Psi_j^0 | \hat{W} | \Psi^0 \rangle$; *i.e.*, we get

$$c_j = -\frac{W_{j0}}{E_j^0 - E^0} \tag{4.115}$$

which is the well-known formula for the coefficients of the first-order wave function in the Rayleigh–Schrödinger PT.

The previous result means that

$$\chi = \sum_{j=1} \frac{-W_{j0}}{E_j^0 - E^0} |\Psi_j^0\rangle \tag{4.116}$$

and this should be substituted into (4.106) in order to get the second-order energy correction:

$$\begin{aligned} J_2 &= \langle \sum_{j=1} \frac{-W_{j0}}{E_j^0 - E^0} \Psi_j^0 | \hat{W} - E^{(1)} | \Psi^0 \rangle \\ &+ \langle \Psi^0 | \hat{W} - E^{(1)} | \sum_{j=1} \frac{-W_{j0}}{E_j^0 - E^0} \Psi_j^0 \rangle + \\ &+ \langle \sum_{i=1} \frac{-W_{i0}}{E_i^0 - E^0} \Psi_i^0 | \hat{H}^0 - E^0 | \sum_{j=1} \frac{-W_{j0}}{E_j^0 - E^0} \Psi_j^0 \rangle \, . \end{aligned} \tag{4.117}$$

We utilize the facts that $W_{j0}^* = W_{0j}$, $\langle \Psi_j^0 | \Psi^0 \rangle = 0$, and that in the third term $(\hat{H}^0 - E^0)\Psi_j^0 = (E_j^0 - E^0)\Psi_j^0$, which cancels the last denominator. We get:

$$\begin{aligned} J_2 &= -2 \sum_{j=1} \frac{|W_{j0}|^2}{E_j^0 - E^0} + \sum_{i,j} \frac{W_{0i} W_{j0}}{E_i^0 - E^0} \underbrace{\langle \Psi_i^0 | \Psi_j^0 \rangle}_{\delta_{ij}} \\ &= -\sum_{j=1} \frac{|W_{j0}|^2}{E_j^0 - E^0} \end{aligned} \tag{4.118}$$

This is the well-known second-order energy formula.

The main advantage of the variational–perturbational theory is not in the independent rederivation of the expressions of the usual PT formalism but in the possibility of treating, at least approximately, the problems in which one is unable to use the standard perturbational expansions. This is the case if one does not know all the eigenvalues and eigenvectors of the unperturbed Hamiltonian $\hat{H}^0$. In such a case one assumes that the best approximate first-order wave function is that which results in the lowest J_2 value.

It may be of interest to note that the form of (4.106) is such that the first-order wave function χ is not necessarily requested to be orthogonal to Ψ^0. As it is

trivial to check, the value of J_2 does not change, if one replaces χ by $\chi + \alpha\Psi^0$, where α is an arbitrary coefficient. That means that the desired minimal value of J_2 can be obtained by performing the optimization without imposing any restrictions on χ, but then χ is not determined uniquely. The correct first-order wave function is obtained, however, only if the requirement $\langle\Psi^0|\chi\rangle = 0$ is also added.

3. Degenerate Rayleigh–Schrödinger perturbation theory

The formulas discussed in the previous sections become inapplicable if the unperturbed state considered belongs to a degenerate eigenvalue of the unperturbed Hamiltonian, because some of the denominators become zero.

Let us consider the case when the eigenvalue E_k^0 of the unperturbed Hamiltonian has an f-fold degeneracy

$$\hat{H}^0\Psi_{k,i}^0 = E_k^0\Psi_{k,i}^0 \; ; \quad i = 1, 2, \ldots, f \; , \tag{4.119}$$

where we denote by $\Psi_{k,i}^0$ the f orthonormalized solutions of the unperturbed problem, belonging to the eigenvalue E_k^0.

Let the perturbed problem be, as usual

$$(\hat{H}^0 + \lambda\hat{W})\Psi = E\Psi \; . \tag{4.120}$$

Expanding Ψ and E as power series of λ, we have

$$(\hat{H}^0 + \lambda\hat{W})\sum_{l=0}^{\infty}\lambda^l\Psi_k^{(l)} = \sum_{j=0}^{\infty}\lambda^j E_k^{(j)}\sum_{l=0}^{\infty}\lambda^l\Psi_k^{(l)} \; . \tag{4.121}$$

We shall request the coefficients at the different powers of λ to be equal on the two sides of (4.121).

For λ^0 we get

$$\hat{H}^0\Psi_k^0 = E_k^0\Psi_k^0 \; , \tag{4.122}$$

which means that $E_k^{(0)} = E_k^0$ and (4.122) is satisfied by arbitrary linear combinations

$$\Psi_k^{(0)} = \sum_{i=1}^{f} c_i\Psi_{k,i}^0 \; . \tag{4.123}$$

It is important to stress that *the coefficients c_i cannot be determined by studying the zero-order problem alone:* it is the actual perturbation $\hat{W}$ that determines which vector of the degenerate subspace is the appropriate zero-order wave function.

For the coefficients at λ^1 we have

$$\hat{W}\Psi_k^{(0)} + \hat{H}^0\Psi_k^{(1)} = E_k^{(1)}\Psi_k^{(0)} + E_k^{(0)}\Psi_k^{(1)} \ . \tag{4.124}$$

Regrouping the terms of (4.124) and substituting the sum (4.123) for the $\Psi_k^{(0)}$:

$$(\hat{W} - E_k^{(1)})\sum_{i=1}^{f} c_i\Psi_{k,i}^0 = -(\hat{H}^0 - E_k^0)\Psi_k^{(1)} \ . \tag{4.125}$$

Now, we multiply (4.125) by one of the functions $\Psi_{k,m}^{0*}$ and integrate—the two terms on the right-hand side cancel due to the Hermiticity of $\hat{H}^0$ and the equality (4.119):

$$\langle\Psi_{k,m}^0|\hat{W} - E_k^{(1)}|\sum_{i=1}^{f} c_i\Psi_{k,i}^0\rangle = 0 \ ; \quad m = 1, 2, \ldots, f \ . \tag{4.126}$$

By introducing the notation

$$W_{mi}^{(k)} = \langle\Psi_{k,m}^0|\hat{W}|\Psi_{k,i}\rangle \tag{4.127}$$

we get

$$\sum_{i=1}^{f} W_{mi}^{(k)} c_i = E_k^{(1)} c_m \ ; \qquad m = 1, 2, \ldots, f \ . \tag{4.128}$$

We introduce the following matrix notations: $\mathbf{W}^{(k)}$ is the $f \times f$ matrix built up of the matrix elements of $\hat{W}$ in the degenerate subspace, $\mathbf{c}$ is the f-dimensional vector of the coefficients c_i. Then (4.128) can be written as the *eigenvalue equation*

$$\mathbf{W}^{(k)}\mathbf{c} = E_k^{(1)}\mathbf{c} \ . \tag{4.129}$$

This equation has f different orthonormalized solutions that diagonalize the perturbation in the degenerate subspace. Usually (but not necessarily) the first-order energies are not degenerate anymore.

According to the previous discussion, in order to find the correct zero-order functions, one has to solve the linear variational problem (c.f. Section 1 of Chapter 3) for the total perturbed Hamiltonian in the subspace spanned by the degenerate eigenvectors of the unperturbed problem. Obviously, this scheme remains applicable also in the "quasidegenerate" case, when the unperturbed problem has several close—not necessarily strictly degenerate—states, separated by a considerable energy "gap" from the remaining ones, although in that

case such a procedure is not quite strict from a formal perturbational point of view.

In the case of degenerate zero-order problems one considers only rarely the higher order corrections of the wave function and of the energy. It may be of interest to note that the first-order wave function and the second order energy can be determined by neglecting completely the remaining $f-1$ states of the subspace that was originally degenerate. Essentially the formulas of the non-degenerate PT become applicable by using the selected eigenvector of (4.129) as the unperturbed state: it interacts with the states in the orthogonal complement of the originally degenerate subspace but, because of the diagonalization performed, not with the other $f-1$ states in that subspace.

4. Brillouin–Wigner perturbation theory

Historically this theory was considered an alternative of the Rayleigh–Schrödinger PT, as it may give faster convergence—but at the expense of an additional iteration procedure. The iterations are necessary because in this framework the energy we are looking for also occurs on the right-hand side of the perturbational equations. Nowadays this scheme is not in general use because of the size–consistency problems discussed in the next section.

In this approach one does not apply any formal perturbational parameter λ, and the wave function and the energy are searched in an iterative fashion. (The iterations for the energy mentioned previously represent something additional to the iterational character of the derivation by which the formulas themselves are obtained.)

We write the perturbed Schrödinger equation in the form

$$(\hat{H}^0 + \hat{V})\,|\Psi^0 + \varphi\rangle = E\,|\Psi^0 + \varphi\rangle\ , \tag{4.130}$$

where $|\Psi^0\rangle$ is the solution of the unperturbed Schrödinger equation $\hat{H}^0|\Psi^0\rangle = E^0|\Psi^0\rangle$, and the "intermediate" normalization is used: $\langle\Psi^0|\Psi^0\rangle = 1$; $\langle\Psi^0|\varphi\rangle = 0$. (We assume that $|\Psi^0\rangle$ represents the lowest energy eigenstate of $\hat{H}^0$, i.e., $|\Psi^0\rangle = |\Psi^0_0\rangle$; for simplicity we shall omit subscript 0 everywhere.)

By multiplying (4.130) with $\langle\Psi^0|$, one gets readily

$$E = \langle\Psi^0|\hat{H}^0 + \hat{V}|\Psi^0 + \varphi\rangle = E^0 + V_{00} + \langle\Psi^0|\hat{V}|\varphi\rangle\ , \tag{4.131}$$

because

$$\langle\Psi^0|\hat{H}^0|\varphi\rangle = E^0\langle\Psi^0|\varphi\rangle = 0\ . \tag{4.132}$$

The Schrödinger equation (4.130) can be regrouped as

$$(E - \hat{H}^0)|\varphi\rangle = (\hat{H}^0 - E)|\Psi^0\rangle + \hat{V}|\Psi^0 + \varphi\rangle\ . \tag{4.133}$$

We introduce the projector

$$\hat{Q} = 1 - |\Psi^0\rangle\langle\Psi^0| \tag{4.134}$$

on the orthogonal complement to $|\Psi^0\rangle$ and apply it to (4.133). We get

$$\hat{Q}(E - \hat{H}^0)|\varphi\rangle = \hat{Q}\hat{V}|\Psi^0 + \varphi\rangle \ , \tag{4.135}$$

because, as it is easy to see, $\hat{Q}(\hat{H}^0 - E)|\Psi^0\rangle = \hat{Q}(E^0 - E)|\Psi^0\rangle = 0$. Furthermore, $|\varphi\rangle$, is in the orthogonal complement to $|\Psi^0\rangle$ and the same holds for $\hat{H}^0|\varphi\rangle$ according to (4.132). Therefore $(E - \hat{H}^0)|\varphi\rangle$ is orthogonal to $|\Psi^0\rangle$, and

$$\hat{Q}(E - \hat{H}^0)|\varphi\rangle = (E - \hat{H}^0)|\varphi\rangle \ , \tag{4.136}$$

thus the operator $\hat{Q}$ may be omitted on the left-hand side of (4.135):

$$(E - \hat{H}^0)|\varphi\rangle = \hat{Q}\hat{V}|\Psi^0 + \varphi\rangle \ . \tag{4.137}$$

Now, we introduce the inverse operator $(E - \hat{H}^0)^{-1}$ (the resolvent corresponding to $\hat{H}^0$). It has the spectral resolution in terms of the eigenvectors and eigenvalues of $\hat{H}^0$:

$$(E - \hat{H}^0)^{-1} = \sum_{j=0} \frac{|\Psi_j^0\rangle\langle\Psi_j^0|}{E - E_j^0} \ . \tag{4.138}$$

(We assume that E does not equal to either of the eigenvalues of $\hat{H}^0$.) Applying this operator to (4.137), we have

$$|\varphi\rangle = (E - \hat{H}^0)^{-1}\hat{Q}\hat{V}|\Psi^0 + \varphi\rangle \ . \tag{4.139}$$

This is an exact equality; we may try to solve it by iterations, utilizing the fact, that if $\hat{V}$ is a small perturbation, then $|\varphi\rangle$ is a small correction. One gets the first approximation to $|\varphi\rangle$ by omitting it on the right-hand side (i.e., by using $|\varphi\rangle = 0$ as the zeroth approximation). This leads to

$$|\varphi^{(1)}\rangle = (E - \hat{H}^0)^{-1}\hat{Q}\hat{V}|\Psi^0\rangle \ . \tag{4.140}$$

By substituting (4.140) in the right-hand side of (4.139) one gets the second approximation

$$|\varphi^{(2)}\rangle = (E - \hat{H}^0)^{-1}\hat{Q}\hat{V}|\Psi^0\rangle + [(E - \hat{H}^0)^{-1}\hat{Q}\hat{V}]^2|\Psi^0\rangle \tag{4.141}$$

which may be repeated, giving the series

$$\begin{aligned} |\varphi\rangle \ &= \ (E - \hat{H}^0)^{-1}\hat{Q}\hat{V}|\Psi^0\rangle + [(E - \hat{H}^0)^{-1}\hat{Q}\hat{V}]^2|\Psi^0\rangle \\ &\quad + [(E - \hat{H}^0)^{-1}\hat{Q}\hat{V}]^3|\Psi^0\rangle + \dots \ . \end{aligned} \tag{4.142}$$

The second-order Brillouin–Wigner energy correction is obtained by substituting the first-order wave function (4.140) in the last term on the right-hand side of (4.131):

$$E^{(2)} = \langle\Psi^0|\hat{V}|\varphi^{(1)}\rangle = \langle\Psi^0|\hat{V}(E - \hat{H}^0)^{-1}\hat{Q}\hat{V}|\Psi^0\rangle \ . \tag{4.143}$$

If one stops at second order, then E should be substituted by $E^0 + V_{00} + E^{(2)}$ on the right-hand side, as well. By substituting (4.138) into (4.143) (taking also into account that $\langle\Psi^0|\hat{Q} = 0)$,

$$E^{(2)} = \langle\Psi^0|\hat{V}\sum_{j=0}\frac{|\Psi_j^0\rangle\langle\Psi_j^0|}{E-E_j^0}\hat{Q}\hat{V}|\Psi^0\rangle = \langle\Psi^0|\hat{V}\sum_{j=1}\frac{|\Psi_j^0\rangle\langle\Psi_j^0|}{E-E_j^0}\hat{V}|\Psi^0\rangle$$
$$= \sum_{j=1}\frac{|V_{0j}|^2}{E-E_j^0}\,. \tag{4.144}$$

That means that the total Brillouin–Wigner energy up to second order is

$$E = E^0 + V_{00} + \sum_{j=1}\frac{|V_{0j}|^2}{E-E_j^0}\,. \tag{4.145}$$

This result differs from the second-order Rayleigh–Schrödinger formula by the presence of the unknown energy $E = E^0 + V_{00} + E^{(2)}$ on the right-hand side (instead of simply E^0). Therefore, (4.145) should be solved by iterations.

Quite similarly to (4.145), one can obtain the explicit formula for the wave function up to the first order:

$$|\Psi\rangle = |\Psi^0\rangle + \sum_{i=1}\frac{V_{i0}}{E-E_i^0}|\Psi_i^0\rangle \tag{4.146}$$

which again differs from the respective Rayleigh–Schrödinger formula by replacing E^0 with E.

Example

Let us consider the simplest example of a system having only two unperturbed levels, Ψ_0^0 and Ψ_1^0, and let us assume the diagonal matrix elements of the perturbation to be zero ($V_{00} = V_{11} = 0, V_{01} \neq 0$). The Hamiltonian matrix in this case is

$$\mathbf{H} = \mathbf{H}^0 + \mathbf{V} = \begin{pmatrix} E^0 & V_{01} \\ V_{10} & E_1^0 \end{pmatrix}, \tag{4.147}$$

i.e., the secular equation is

$$\begin{vmatrix} E^0 - E & V_{01} \\ V_{10} & E_1^0 - E \end{vmatrix} = (E-E^0)(E-E_1^0) - |V_{01}|^2 = 0\,. \tag{4.148}$$

The second-order Brillouin–Wigner energy formula in this case is

$$E = E^0 + \frac{|V_{01}|^2}{E-E_1^0} \tag{4.149}$$

as the sum over j in (4.145) is reduced to a single term. It is easy to see that (4.149) can be rearranged to recover the exact secular equation (4.148). This result indicates that in this particular case the second-order Brillouin–Wigner perturbational energy coincides with the exact one. It can easily be checked that in this very special model the Brillouin–Wigner series terminates after the first wave function correction, as a consequence of the two-level character of the system and of the equality $V_{11} = 0$.

4.1. The size–consistency problem

The second-order Brillouin–Wigner perturbation theory gives the exact solution for the simplest two-level model problem. However, the situation becomes quite different if one has two noninteracting two-level systems A and B, each of which is described separately by a Hamiltonian matrix like (4.147). In this case the "supersystem" AB has four levels, Ψ_{00}, Ψ_{01}, Ψ_{10} and Ψ_{11}, where $\Psi_{ij} = \Psi_i^A \Psi_j^B$. (The basis set of AB is a "direct product" of the basis sets of A and B.) The Hamiltonian is $\hat{H} = \hat{H}^A + \hat{H}^B$, where $\hat{H}^A$ and $\hat{H}^B$ act only on the functions of the subsystems A and B, respectively. Therefore,

$$\begin{aligned} \langle \Psi_{ij} | \hat{H} | \Psi_{kl} \rangle &= \langle \Psi_i^A \Psi_j^B | \hat{H}^A + \hat{H}^B | \Psi_k^A \Psi_l^B \rangle \\ &= H_{ik}^A \delta_{jl} + \delta_{ik} H_{jl}^B \ . \end{aligned} \tag{4.150}$$

It is, in general, true that the Hamiltonian matrix $\mathbf{H}$ of a system consisting of two noninteracting subsystems A and B is

$$\mathbf{H} = \mathbf{H}^A \otimes \mathbf{1}^B + \mathbf{1}^A \otimes \mathbf{H}^B \ , \tag{4.151}$$

where $\mathbf{H}^A$ and $\mathbf{H}^B$ are the Hamiltonian matrices of the separated systems A and B, respectively, $\mathbf{1}^A$ and $\mathbf{1}^B$ are the unit matrices of dimensions equal to those of matrices $\mathbf{H}^A$ and $\mathbf{H}^B$, respectively, and symbol $\mathbf{X} \otimes \mathbf{Y}$ denotes the direct product of matrices $\mathbf{X}$ and $\mathbf{Y}$. (The properties of direct products of matrices are briefly described in Appendix IX.)

The explicit form of the Hamiltonian matrix of the system consisting of two noninteracting two-level subsystems A and B is

$$\begin{array}{cc} & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \quad \leftarrow \textit{levels} \\ \mathbf{H} = & \begin{pmatrix} E_0^A + E_0^B & V_{01}^B & V_{01}^A & 0 \\ V_{10}^B & E_0^A + E_1^B & 0 & V_{01}^A \\ V_{10}^A & 0 & E_1^A + E_0^B & V_{01}^B \\ 0 & V_{10}^A & V_{10}^B & E_1^A + E_1^B \end{pmatrix} \begin{array}{c} \downarrow \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \end{array} \ . \tag{4.152}$$

It can be seen by direct calculation that the eigenvalues of this Hamiltonian matrix are the sums $\varepsilon_i^A + \varepsilon_j^B$ $(i, j = 1, 2)$, where $\varepsilon_i^A, \varepsilon_j^B$ are the exact energies

of the separated subsystems. To check this, it is sufficient to form the direct products

$$\mathbf{c} = \mathbf{c}^{Ai} \otimes \mathbf{c}^{Bj} = \begin{pmatrix} c_0^{Ai} c_0^{Bj} \\ c_0^{Ai} c_1^{Bj} \\ c_1^{Ai} c_0^{Bj} \\ c_1^{Ai} c_1^{Bj} \end{pmatrix} \tag{4.153}$$

of the individual solutions

$$\mathbf{c}^{Ai} = \begin{pmatrix} c_0^{Ai} \\ c_1^{Ai} \end{pmatrix} \quad \text{and} \quad \mathbf{c}^{Bj} = \begin{pmatrix} c_0^{Bj} \\ c_1^{Bj} \end{pmatrix} \tag{4.154}$$

of the subsystems A and B, respectively, substitute it into the eigenvalue equation $\mathbf{Hc} = \varepsilon\mathbf{c}$, and group the terms appropriately. (Of course, one can also use the formulae for direct products, given in Appendix IX.)

The second-order Brillouin–Wigner energy formula for the supersystem AB described by the Hamiltonian matrix (4.152) is, according to (4.145)

$$E = E_0^A + E_0^B + \frac{|V_{01}^A|^2}{E - (E_1^A + E_0^B)} + \frac{|V_{01}^B|^2}{E - (E_0^A + E_1^B)} \tag{4.155}$$

which *does not recover* the exact eigenvalue $\varepsilon_1^A + \varepsilon_2^B$ of the Hamiltonian (4.152); accordingly, it is not the sum of the second-order Brillouin–Wigner energies of the subsystems A and B. This means that the second-order Brillouin–Wigner energy does not satisfy the very important requirement of "size consistency." (Some authors use the term "size extensivity".)

Size consistency is of utmost importance if relative energies (energy changes) are considered. When one considers the interaction between two or more subsystems of a larger system, and increases the distance between them, then the energy will tend to the sum of the energies of the isolated subsystems. Therefore a calculation of the supersystem in which the subsystems are put at an infinite distance from each other and do not interact, must give exactly the same energy as the sum of the energies that one obtains by performing separate calculations for the individual subsystems at the same level of theory. Methods satisfying this requirement are called size consistent (size extensive). A method that is not size consistent can hardly give any meaningful interaction energy value.

It is important to stress in this connection the difference with the behavior of the Rayleigh–Schrödinger perturbational method. Considering again the off-diagonal matrix elements of (4.152) as the perturbation, the second-order Rayleigh–Schrödinger energy formula gives

$$E = E_0^A + E_0^B - \frac{|V_{01}^A|^2}{E_1^A - E_0^A} - \frac{|V_{01}^B|^2}{E_1^B - E_0^B} \,. \tag{4.156}$$

This is the sum of the respective second-order Rayleigh–Schrödinger energies calculated for the isolated systems A and B. Neither of these second-order Rayleigh–Schrödinger energies equals the respective exact energy, but they obey size consistency and, therefore, can be used to calculate energy differences. Size consistency holds for Rayleigh–Schrödinger perturbational theory at any order and for arbitrary systems.

The order-by-order size consistency of the Rayleigh–Schrödinger perturbation theory is worthy of detailed investigation. The next section is devoted to this problem.

5. Size consistency of the Rayleigh–Schrödinger perturbation theory

We shall consider the size consistency of the Rayleigh–Schrödinger perturbation theory from two different aspects. First we show that the order-by-order size consistency must hold as a formal consequence of the fact that we are using power series, then we shall consider how this size consistency manifests itself at the level of the actual perturbational formulae.

5.1. Formal considerations based on the properties of power series

Let us first describe some qualitative argumentation indicating that Rayleigh–Schrödinger perturbation theory gives size consistent results if the perturbational series are (absolutely) convergent.

The energy of a system AB consisting of two noninteracting subsystems A and B is obviously the simple sum of the individual subsystem energies:

$$E^{AB}(\lambda) = E^{A}(\lambda) + E^{B}(\lambda) \ , \tag{4.157}$$

where we have indicated that these energies are considered functions of the same perturbational parameter λ. In the Rayleigh–Schrödinger perturbation theory each of these energies is expanded into a series according to the powers of λ—c.f. (4.11). Therefore the sum of the series obtained for $E^{A}(\lambda)$ and for $E^{B}(\lambda)$, respectively, equals the series for $E^{AB}(\lambda)$. This equality can hold for arbitrary λ values only if the coefficients at every power λ^{l} are the same on the two sides:

$$E^{AB(l)} = E^{A(l)} + E^{B(l)} \ . \tag{4.158}$$

This means the order-by-order size consistency of the Rayleigh–Schrödinger perturbational expansion. (The Brillouin–Wigner perturbation theory does not give the energy as a simple power series of λ; it is not surprising, therefore, that it is not size consistent. However, the power series character of the expansion is a sufficient but not a necessary condition for a method to be size consistent:

the energy obtained in the framework of the "coupled cluster approach"—see Section 3.4 of Chapter 8—is not a simple power series but is size consistent.)

Now we shall give also a more detailed demonstration of equality (4.158), which permits one to conclude that the order-by-order size consistency of the Rayleigh–Schrödinger perturbational expansion holds even in the cases in which the perturbational series does not eventually converge. The perturbational expansions often represent so-called "asymptotical series," which means that the results approach the exact solution up to a given order of the perturbational expansion but then start to diverge. In practice we perform finite-order perturbational calculations and usually cannot determine whether the series is convergent or asymptotical. Therefore, size consistency is an important issue even in the general case for which one cannot postulate convergence.

In perturbation theory we decompose each Hamiltonian into its unperturbed part and perturbation:

$$\begin{aligned}
\hat{H}^A &= \hat{H}^{A0} + \hat{V}^A = \hat{H}^{A0} + \lambda \hat{W}^A \\
\hat{H}^B &= \hat{H}^{A0} + \hat{V}^B = \hat{H}^{B0} + \lambda \hat{W}^B \\
\hat{H}^{AB} &= \hat{H}^A + \hat{H}^B = \hat{H}^{A0} + \hat{H}^{B0} + \lambda(\hat{W}^A + \hat{W}^B) = \hat{H}^{AB0} + \lambda \hat{W}^{AB} .
\end{aligned} \tag{4.159}$$

As discussed in Section 1.1, the perturbation parameter λ can be considered, at least formally, a continuous variable; we may always scale $\hat{W}^A$ and $\hat{W}^B$ as to have the same parameter λ for $\hat{H}^A$ and $\hat{H}^B$—and thus also for $\hat{H}^{AB}$. As the Hamiltonians (4.159) depend on λ, the same holds also for the energies, as indicated on both sides of (4.157). It is known from mathematics that if one can expand the energies $E^A(\lambda)$ and $E^B(\lambda)$ into (absolutely) convergent power series of λ, then the sum of these series is also convergent and converges to the sum $E^A(\lambda) + E^B(\lambda)$. It follows from the uniqueness of the power series that there is no other power series expansion of $E^{AB}(\lambda)$ than that given by the sum of the expansions obtained from $E^A(\lambda)$ and $E^B(\lambda)$. The expansion of the energies in the framework of the Rayleigh–Schrödinger perturbation theory gives nothing else than these series expansions of $E^A(\lambda)$, $E^B(\lambda)$, and $E^{AB}(\lambda)$; therefore the perturbational contributions to $E^{AB}(\lambda)$ must be equal *order by order* to the sum of the respective contributions to $E^A(\lambda)$ and $E^B(\lambda)$, i.e., (4.158) must be fulfilled for every l.

If $E(\lambda)$ is a "well-behaved" (i.e., continuous and repeatedly differentiable) function in the vicinity of $\lambda = 0$, then it can be expanded into a power series (MacLaurin's series) with finite coefficients. In turn, every power series with finite coefficients has a nonzero radius of convergence; therefore, for every perturbational problem of the type discussed, all the three series in question must converge for sufficiently small λ values, and then their coefficients obey (4.158). But the equality (4.158) does not contain the parameter λ. Therefore, the order-by-order size consistency of the Rayleigh–Schrödinger perturbation theory holds irrespective of whether the series eventually converge *at the given* λ value. It follows from these considerations that size consistency holds also for the alternate treatment mentioned in Section 1.1, in which one does not use any intermediate notations $\hat{V} = \lambda \hat{W}$, but first formally introduces parameter λ and then substitutes $\lambda = 1$ after the series expansion had been performed.

It is to be noted that size consistency for two noninteracting subsystems (A and B) implies also size consistency for an arbitrary number of such subsystems ($A, B, C \ldots$): first one establishes size consistency for $AB = A + B$, then—treating AB as a single subsystem—for $ABC = AB + C$ and so on.

5.2. Size consistency of the perturbational expansions

Now we shall consider the size consistency (extensivity) problem at the level of actual expressions obtained for the perturbational wave function and energy corrections.

We are dealing with a supersystem AB consisting of two noninteracting subsystems A and B. The Hamiltonian of the supersystem is the sum of the two subsystem Hamiltonians:

$$\hat{H}^{AB} = \hat{H}^{A} + \hat{H}^{B} \tag{4.160}$$

and the wave function Ψ^{AB} is defined in the Hilbert space representing a "direct product space" of the Hilbert spaces of the two subsystems. This means that the Hilbert space of the supersystem can be referred to a basis $\{\Psi_I\}$, each element of which represents a product of basis functions used for the individual subsystems:

$$\Psi_I^{AB} = \Psi_{ij}^{AB} = \Psi_i^A \Psi_j^B \ , \tag{4.161}$$

where subscript I represents the pair of subscripts i and j; this is denoted as $I = (ij)$.

As the subsystems do not interact, the order in which the Hamiltonian $\hat{H}^A$ and the function Ψ_j^B enter the expressions may be interchanged, and the same holds for the Hamiltonian $\hat{H}^B$ and the function Ψ_i^A. In particular, if Ψ_i^A and Ψ_j^B are eigenfunctions of $\hat{H}^A$ and $\hat{H}^B$, respectively, then we have, in agreement with the expectations

$$\begin{aligned} \hat{H}^{AB}\Psi_I^{AB} &= (\hat{H}^A + \hat{H}^B)\Psi_i^A\Psi_j^B = (\hat{H}^A\Psi_i^A)\Psi_j^B + \Psi_i^A\hat{H}^B\Psi_j^B \\ &= (E_i^A + E_j^B)\Psi_i^A\Psi_j^B = (E_i^A + E_j^B)\Psi_I^{AB} = E_I^{AB}\Psi_I^{AB} \end{aligned} \tag{4.162}$$

(c.f. with the discussion in Section 1.1 of Chapter 1).

Here we are discussing a general formalism without specifying the physical nature of the systems considered. If the two subsystems contain electrons (and/or other fermions) then the wave functions should be antisymmetric with respect of their interchanges. Obviously, the wave functions of type (4.161) do not obey the antisymmetry requirement with respect to interchanges of electrons *between* the two subsystems. However, as it will be discussed in Section 4.2 of Chapter 5, the antisymmetrization of the wave function between noninteracting subsystems is immaterial, so we may neglect this point as far as strictly noninteracting subsystems are considered.

Now, if we have the decompositions

$$\hat{H}^A = \hat{H}^{A0} + \lambda\hat{W}^A \tag{4.163}$$

and

$$\hat{H}^B = \hat{H}^{B0} + \lambda\hat{W}^B \tag{4.164}$$

then the supersystem Hamiltonian is

$$\hat{H}^{AB} = \hat{H}^{AB0} + \lambda \hat{W}^{AB} \tag{4.165}$$

with

$$\hat{H}^{AB0} = \hat{H}^{A0} + \hat{H}^{B0} \tag{4.166}$$

and

$$\hat{W}^{AB} = \hat{W}^{A} + \hat{W}^{B} \,. \tag{4.167}$$

In order to simplify the notations, we shall denote the states of subsystems A and B, which are actually considered, as the 0th states. These may be the ground states in which we are mostly interested, but that is not necessarily the case—we may always select the numbering of states in such a manner. We shall omit subscript "0" whenever the perturbed wave functions and energies (energy contributions) are considered.

The Rayleigh–Schrödinger perturbational series of the energies and wave functions of the individual subsystems are

$$\begin{aligned} E^{A} &= E_0^{A0} + \sum_{l=1}^{\infty} \lambda^l E^{A(l)} \\ E^{B} &= E_0^{B0} + \sum_{l=1}^{\infty} \lambda^l E^{B(l)} \end{aligned} \tag{4.168}$$

and

$$\begin{aligned} \Psi^{A} &= \Phi_0^{A} + \sum_{l=1}^{\infty} \lambda^l \sum_{i=1} c_i^{A(l)} \Phi_i^{A} \\ \Psi^{B} &= \Phi_0^{B} + \sum_{l=1}^{\infty} \lambda^l \sum_{j=1} c_j^{B(l)} \Phi_j^{B} \end{aligned} \tag{4.169}$$

respectively, where we have introduced the notations Φ_i^A and Φ_j^B for the solutions of the unperturbed Schrödinger equations of the subsystems.

The state of the supersystem AB corresponding to Ψ^A and Ψ^B will be assigned the notation $\Psi^{AB} = \Psi^A \Psi^B$. The product of the unperturbed wave functions $\Phi_i^A \Phi_j^B$ may be denoted as

$$\Phi_I^{AB} = \Phi_{ij}^{AB} = \Phi_i^A \Phi_j^B \,. \tag{4.170}$$

It is an eigenvector of the unperturbed Schrödinger equation

$$\begin{aligned} \hat{H}^{AB0} \Phi_I^{AB} &= (\hat{H}^{A0} + \hat{H}^{B0}) \Phi_i^A \Phi_j^B = (E_i^{A0} + E_j^{B0}) \Phi_i^A \Phi_j^B \\ &= (E_i^{A0} + E_j^{B0}) \Phi_I^{AB} \,. \end{aligned} \tag{4.171}$$

The unperturbed supersystem wave function actually studied is, accordingly

$$\Phi^{AB}_{(00)} = \Phi^A_0 \Phi^B_0 \tag{4.172}$$

and the corresponding unperturbed energy is

$$E^{AB0}_{(00)} = E^{A0}_0 + E^{B0}_0 \ . \tag{4.173}$$

The perturbational series of the supersystem are

$$E^{AB} = E^{AB0}_{(00)} + \sum_{l=1}^{\infty} \lambda^l E^{AB(l)} \tag{4.174}$$

and

$$\Psi^{AB} = \Phi^{AB}_{(00)} + \sum_{l=1}^{\infty} \lambda^l \sum_{I \neq (00)} c^{(l)}_I \Phi^{AB}_I \ . \tag{4.175}$$

From the other side, the exact wave function Ψ^{AB} of the supersystem is, according to (4.162), the product of the exact solutions Ψ^A and Ψ^B of the non-interacting subsystems. If the perturbational series (4.169) converge absolutely, then this wave function is

$$\Psi^{AB} = \Psi^A \Psi^B = \left(\Phi^A_0 + \sum_{l=1}^{\infty} \lambda^l \sum_{i=1} c^{A(l)}_i \Phi^A_i \right) \left(\Phi^B_0 + \sum_{m=1}^{\infty} \lambda^m \sum_{j=1} c^{B(m)}_j \Phi^B_j \right) \ . \tag{4.176}$$

Performing the multiplication of the series and collecting terms with equal powers of λ we get

$$\begin{aligned} \Psi^{AB} &= \Phi^A_0 \Phi^B_0 + \sum_{l=1}^{\infty} \lambda^l \left(\sum_{i=1} c^{A(l)}_i \Phi^A_i \Phi^B_0 + \sum_{j=1} c^{B(l)}_j \Phi^A_0 \Phi^B_j \right) \\ &\quad + \sum_{l=2}^{\infty} \lambda^l \sum_{i,j=1} \sum_{k=1}^{l-1} c^{A(l-k)}_i c^{B(k)}_j \Phi^A_i \Phi^B_j \ . \end{aligned} \tag{4.177}$$

According to (4.177), the product of the wave functions in (4.169) leads to the following values of the coefficients $c^l_I = c^l_{ij}$ in (4.175):

$$c^{(1)}_{(ij)} = 0 \qquad (i, j \neq 0) \tag{4.178}$$

$$c^{(l)}_{(i0)} = c^{A(l)}_i \ ; \quad c^{(l)}_{(0j)} = c^{B(l)}_j \qquad (i, j \neq 0;\ l \geq 1) \tag{4.179}$$

and

$$c_{(ij)}^{(l)} = \sum_{k=1}^{l-1} c_i^{A(l-k)} c_j^{B(k)} \equiv \sum_{k=1}^{l-1} c_i^{A(k)} c_j^{B(l-k)} \quad (i,j \neq 0;\ l \geq 2)\ . \quad (4.180)$$

Size consistency means that the energy contributions satisfy (4.158), i.e., $E^{AB(l)} = E^{A(l)} + E^{B(l)}$. To prove size consistency, we are going to demonstrate that equalities (4.178)–(4.180) can be recovered directly from the perturbational formulae described in Section 1.2, independently whether the series (4.169) are convergent or not.

According to (4.173), size consistency is fulfilled for the zero-order energies. For the first-order energy we need the expectation value $\langle \Phi_{(00)}^{AB} | \hat{W}^{AB} | \Phi_{(00)}^{AB} \rangle$; for that reason we consider the matrix elements of $\hat{W}^{AB}$.

The general matrix element of $\hat{W}^{AB}$ between the unperturbed supersystem wave functions $\Phi_I^{AB} = \Phi_{(ij)}^{AB} = \Phi_i^A \Phi_j^B$ and $\Phi_J^{AB} = \Phi_{(kl)}^{AB} = \Phi_k^A \Phi_l^B$ is

$$\begin{aligned} W_{IJ}^{AB} &= \langle \Phi_I^{AB} | \hat{W}^{AB} | \Phi_J^{AB} \rangle = \langle \Phi_{(ij)}^{AB} | \hat{W}^{AB} | \Phi_{(kl)}^{AB} \rangle \qquad (4.181) \\ &= \langle \Phi_i^A \Phi_j^B | \hat{W}^A + \hat{W}^B | \Phi_k^A \Phi_l^B \rangle = W_{ik}^A \delta_{jl} + W_{jl}^B \delta_{ik} \end{aligned}$$

because one can interchange the order of $\hat{W}^A$ and Φ_l^B as well as that of $\hat{W}^B$ and Φ_k^A, and the zero-order subsystem wave functions are assumed orthonormalized. This means that $\hat{W}^{AB}$ has nonzero matrix elements only between those states that are excited with respect to each other in one of the subsystems but correspond to the same state of another. We may also note that (4.181) can be expressed by using matrix notations as

$$\mathbf{W}^{AB} = \mathbf{W}^A \otimes \mathbf{1}^B + \mathbf{1}^A \otimes \mathbf{W}^B\ , \qquad (4.182)$$

where $\mathbf{1}^A$ and $\mathbf{1}^B$ are the unit matrices corresponding to the Hilbert spaces of functions of subsystems A and B respectively (c.f. Appendix IX.)

By using (4.181) we have for the first-order energy correction

$$E^{AB(1)} = \langle \Phi_{(00)}^{AB} | \hat{W}^{AB} | \Phi_{(00)}^{AB} \rangle = W_{00}^A + W_{00}^B = E^{A(1)} + E^{B(1)}\ . \quad (4.183)$$

That means that the first-order energy is also size consistent.

The formula in (4.30) of the second-order energy reads in the present case as

$$\begin{aligned} E^{AB(2)} &= \sum_{I \neq (00)} \frac{-|W_{(00)I}^{AB}|^2}{E_I^{AB0} - E_{(00)}^{AB0}} \qquad (4.184) \\ &= \sum_{(ij) \neq (00)} \frac{-|\langle \Phi_{(00)}^{AB} | \hat{W}^{AB} | \Phi_{(ij)}^{AB} \rangle|^2}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}}\ . \end{aligned}$$

According to (4.181), the matrix element in the numerator is

$$\begin{aligned} W^{AB}_{(00)I} &= \langle \Phi^{AB}_{(00)} | \hat{W}^{AB} | \Phi^{AB}_{(ij)} \rangle = \langle \Phi^A_0 \Phi^B_0 | \hat{W}^A + \hat{W}^B | \Phi^A_i \Phi^B_j \rangle \\ &= W^A_{0i} \delta_{j0} + W^B_{0j} \delta_{i0} \, . \end{aligned} \tag{4.185}$$

In (4.184) the sum over $I = (ij)$ runs over all pairs of subscripts i and j, i.e., it leads to a double sum over i and j. However, the Kronecker symbols on the right-hand side of (4.185) indicate that only those excited configurations $\Phi^A_i \Phi^B_j$ will contribute to the second-order energy, in which either $i = 0$ or $j = 0$; i.e., we have to consider excitations either in the subsystem A or in the subsystem B, but not in both simultaneously. Therefore, the double sum over i and j will reduce to two single sums; a sum over $i \neq 0$, for which one has to assume $j = 0$ and a sum over $j \neq 0$, for which $i = 0$. In both cases two terms cancel in the energy denominator, so we get

$$E^{AB(2)} = \sum_{i \neq 0} \frac{-|W^A_{0i}|^2}{E^{A0}_i - E^{A0}_0} + \sum_{j \neq 0} \frac{-|W^B_{0j}|^2}{E^{B0}_j - E^{B0}_0} = E^{A(2)} + E^{B(2)} \, . \tag{4.186}$$

This means that the second-order energy is also size consistent.

For computing the higher order energy corrections, we shall use (4.24), which for the present case reads

$$E^{AB(l)} = \sum_{I \neq (00)} c^{(l-1)}_I W^{AB}_{(00)I} = \sum_{(ij) \neq (00)} c^{(l-1)}_{(ij)} W^{AB}_{(00)(ij)} \, , \tag{4.187}$$

where $c^{(l-1)}_I = c^{(l-1)}_{(ij)}$ is the expansion coefficient at $\Phi^{AB}_I = \Phi^{AB}_{(ij)} = \Phi^A_i \Phi^B_j$ in the $l-1$-th order supersystem wave function correction.

As a consequence of (4.185), the double sum in (4.187) again reduces to two single sums—one over i (for which $j = 0$ should be substituted) and another for j (with $i = 0$). We get

$$E^{AB(l)} = \sum_{i \neq 0} c^{(l-1)}_{(i0)} W^A_{0i} + \sum_{j \neq 0} c^{(l-1)}_{(0j)} W^B_{0j} \, . \tag{4.188}$$

Comparing with (4.154), we see that one has $E^{AB(l)} = E^{A(l)} + E^{B(l)}$ if

$$c^{(l-1)}_{(i0)} = c^{A(l-1)}_i \; ; \quad c^{(l-1)}_{(0j)} = c^{B(l-1)}_j \tag{4.189}$$

which represent the sufficient condition of size consistency of the lth order energy. This condition is nothing else than (4.179) written out for the order $l-1$ instead of l.

In order to prove the fulfillment of these conditions, we need to study systematically the wave function expansion coefficients. As will be seen, in order

to provide (4.179) to be satisfied, the fulfillment of (4.178) and (4.180) should also be required.

Equations (4.28) and (4.29), giving the expansion coefficients of the first-order wave function correction and those of the general, lth order one, read in the present case as

$$c_I^{(1)} = c_{(ij)}^{(1)} = -\frac{W_{I0}^{AB}}{E_I^{AB0} - E_{(00)}^{AB0}} \tag{4.190}$$

and

$$c_I^{(l)} = -\frac{1}{E_I^{AB0} - E_{(00)}^{AB0}} \left[\sum_{k=1}^{l-1} E^{AB(k)} c_I^{(l-k)} - \sum_{J \neq (00)} c_J^{(l-1)} W_{IJ}^{AB} \right] \tag{4.191}$$

respectively. Substituting into (4.190) $W_{(00)I}^{AB}$ as the complex conjugate of $W_{I(00)}^{AB}$ given in (4.185), we have

$$\begin{aligned} c_I^{(1)} &= c_{(ij)}^{(1)} = \frac{-W_{i0}^A \delta_{j0} - W_{j0}^B \delta_{i0}}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \\ &= -\frac{W_{i0}^A}{E_i^{A0} - E_0^{A0}} \delta_{j0} - \frac{W_{j0}^B}{E_j^{B0} - E_0^{B0}} \delta_{i0} = c_i^{A(1)} \delta_{j0} + c_j^{B(1)} \delta_{i0} \,. \end{aligned} \tag{4.192}$$

which is in agreement with (4.178)–(4.180). Denoting in (4.191) the compound index $J = (kl)$ and using (4.183), (4.192), and (4.181), we get the second-order wave function coefficient as

$$\begin{aligned} c_I^{(2)} = c_{(ij)}^{(2)} = \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \Big[&(W_{00}^A + W_{00}^B)(c_i^{A(1)} \delta_{j0} + c_j^{B(1)} \delta_{i0}) \\ &- \sum_{(kl) \neq (00)} (c_k^{A(1)} \delta_{l0} + c_l^{B(1)} \delta_{k0})(W_{ik}^A \delta_{jl} + W_{jl}^B \delta_{ik}) \Big] \,. \end{aligned} \tag{4.193}$$

It is worthwhile to consider in (4.193) three cases separately: *a)* $i \neq 0, j = 0$; *b)* $i = 0$; $j \neq 0$ and *c)* $i, j \neq 0$.

In case *a)*, $i \neq 0$ and $j = 0$, therefore $\delta_{j0} = 1$, $\delta_{i0} = 0$; $E_j^{B0} = E_0^{B0}$. We get

$$\begin{aligned} c_{(i0)}^{(2)} = \frac{1}{E_i^{A0} + E_0^{A0}} \Big[&(W_{00}^A + W_{00}^B) c_i^{A(1)} \\ &- \sum_{(kl) \neq (00)} (c_k^{A(1)} W_{ik}^A \delta_{l0} \delta_{0l} + c_l^{B(1)} W_{ik}^A \delta_{k0} \delta_{0l} \\ &+ c_k^{A(1)} W_{0l}^B \delta_{l0} \delta_{ik} + c_l^{B(1)} W_{0l}^B \delta_{k0} \delta_{ik}) \Big] \end{aligned} \tag{4.194}$$

The second term in the last sum vanishes because k and l should not be equal to 0 simultaneously; the fourth term also vanishes because only the term $k = i$ survives the summation over k but $i \neq 0$. In the first and third term only term $l = 0$ survives the summation over l:

$$\begin{aligned} c^{(2)}_{(i0)} &= \frac{1}{E_i^{A0} - E_0^{A0}} \Bigg[(W^A_{00} + W^B_{00}) c_i^{A(1)} \\ &\quad - \sum_{k\neq 0} c_k^{A(1)} W^A_{ik} - \sum_{k\neq 0} c_k^{A(1)} W^B_{00} \delta_{ik} \Bigg] \\ &= \frac{1}{E_i^{A0} - E_0^{A0}} \Bigg[W^A_{00} c_i^{A(1)} - \sum_{k\neq 0} c_k^{A(1)} W^A_{ik} \Bigg] = c_i^{A(2)} \,. \end{aligned} \tag{4.195}$$

Case *b)* is quite analogous to case *a)*, only the subsystems A and B (and subscripts i and j) are interchanged. One gets in the same manner

$$c^{(2)}_{(0j)} = c_j^{B(2)} \,. \tag{4.196}$$

Equations (4.195) and (4.196) indicate that the second-order wave function corrections satisfy conditions (4.189) for $l = 3$, therefore the third-order energy is also size consistent.

In order to be able to discuss size consistency of energy corrections of even higher orders, we need also to consider case *c)* for which neither i nor j are equal to zero. In this case $\delta_{j0} = \delta_{i0} = 0$ and we have only to consider the double sum in (4.193):

$$\begin{aligned} c_I^{(2)} &= c^{(2)}_{(ij)} = \frac{-1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \sum_{(kl)\neq(00)} (c_k^{A(1)} W^A_{ik} \delta_{l0} \delta_{jl} \\ &\quad + c_l^{B(1)} W^A_{ik} \delta_{k0} \delta_{jl} + c_k^{A(1)} W^B_{jl} \delta_{l0} \delta_{ik} + c_l^{B(1)} W^B_{jl} \delta_{k0} \delta_{ik}) \,. \end{aligned} \tag{4.197}$$

The first and fourth terms vanish, as the summation over l and k results in δ_{j0} and δ_{i0}, respectively, and we consider the case $i, j \neq 0$. In the second and third terms only the cases $k = 0$, $l = j$ and $k = i$, $l = 0$, respectively, survive the summations, yielding

$$c_I^{(2)} = c^{(2)}_{(ij)} = \frac{-1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \left(c_j^{B(1)} W^A_{i0} + c_i^{A(1)} W^B_{j0} \right). \tag{4.198}$$

Substituting here the standard expressions $c_j^{B(1)} = -\frac{W_{j0}^B}{E_j^{B0} - E_0^{B0}}$ and $c_i^{A(1)} = -\frac{W_{i0}^A}{E_i^{A0} - E_0^{A0}}$, we get, by bringing to a common denominator:

$$\begin{aligned} c_I^{(2)} &= c_{(ij)}^{(2)} = \frac{W_{i0}^A W_{j0}^B}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \left(\frac{1}{E_j^{B0} - E_0^{B0}} + \frac{1}{E_i^{A0} - E_0^{A0}} \right) \\ &= \frac{W_{i0}^A W_{j0}^B}{(E_j^{B0} - E_0^{B0})(E_i^{A0} - E_0^{A0})} = c_i^{A(1)} c_j^{B(1)} . \end{aligned} \tag{4.199}$$

which recovers (4.180) for the case of $l = 2$.

In order to prove that equalities (4.179) and (4.180) are fulfilled at every order, we shall proceed by induction.

Proof by induction

We have seen that the first- and second-order wave functions obey equalities (4.178)–(4.180). Now we shall demonstrate that if these equalities are fulfilled up to some order $t - 1$ of perturbation, then they will also hold at the order t; therefore, it follows by induction that they must be valid at any order, as well. Consequently, the Rayleigh–Schrödinger perturbational expansion is size consistent at every order.

We look for the coefficients $c_{(ij)}^{(t)}$ using the perturbational expansion formula (4.29) and assuming the fulfillment of (4.178)–(4.180)—and thus size consistency—for all orders up to $t - 1$ inclusively. The general formula (4.29) reads in the given case as

$$\begin{aligned} c_I^{(t)} &= c_{(ij)}^{(t)} = \frac{1}{E_I^{AB0} - E_0^{AB0}} \left[\sum_{s=1}^{t-1} E^{AB(s)} c_I^{(t-s)} - \sum_{J \neq (00)} c_J^{(t-1)} W_{IJ}^{AB} \right] \\ &= \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \left[\sum_{s=1}^{t-1} E^{AB(s)} c_{(ij)}^{(t-s)} \right. \\ &\qquad \left. - \sum_{(kl) \neq (00)} c_{(kl)}^{(t-1)} W_{(ij)(kl)}^{AB} \right] . \end{aligned} \tag{4.200}$$

We have the same three cases *a)*, *b)*, and *c)* as we have considered for $c_{(ij)}^{(2)}$.

Case *a)* $j = 0; i \neq 0$

$$c_{(i0)}^{(t)} = \frac{1}{E_i^{A0} - E_0^{A0}} \left[\sum_{s=1}^{t-1} \left(E^{A(s)} + E^{B(s)} \right) c_{(i0)}^{(t-s)} - \sum_{(kl) \neq (00)} c_{(kl)}^{(t-1)} W_{(i0)(kl)}^{AB} \right] . \tag{4.201}$$

The double sum should be decomposed into three sums: one for $k \neq 0$, $l = 0$; one for $k = 0, l \neq 0$; and one for $k, l \neq 0$:

$$c_{(i0)}^{(t)} = \frac{1}{E_i^{A0} - E_0^{A0}} \left[\sum_{s=1}^{t-1} E^{A(s)} c_{(i0)}^{(t-s)} + \sum_{s=1}^{t-1} E^{B(s)} c_{(i0)}^{(t-s)} \right.$$

$$- \sum_{k \neq 0} c_{(k0)}^{(t-1)} (W_{ik}^A + W_{00}^B \delta_{ik}) - \sum_{l \neq 0} c_{(0l)}^{(t-1)} (W_{i0}^A \delta_{0l} + W_{0l}^B \delta_{i0})$$

$$\left. - \sum_{k,l \neq 0} c_{(kl)}^{(t-1)} (W_{ik}^A \delta_{0l} + W_{0l}^B \delta_{ik}) \right] . \tag{4.202}$$

As we are considering the case with $j = 0$, we have substituted 0 for j everywhere in (4.201)–(4.202). Several terms of (4.202) vanish because of the Kronecker symbols, and two sums over k reduce to a single term each with $k = i$. The term containing $W_{00}^B = E^{B(1)}$ cancels the $s = 1$ term in the second sum. By substituting $c_{(i0)}^{(p)} = c_i^{A(p)}$, $c_{(0l)}^{(p)} = c_l^{B(p)}$, and $c_{(il)}^{(t-1)} = \sum_{q=1}^{t-2} c_i^{A(t-1-q)} c_l^{B(q)}$ according to (4.179) and (4.180), we have

$$c_{(i0)}^{(t)} = \frac{1}{E_i^{A0} - E_0^{A0}} \left[\sum_{s=1}^{t-1} E^{A(s)} c_i^{A(t-s)} - \sum_{k \neq 0} c_k^{A(t-1)} W_{ik}^A \right.$$

$$\left. + \sum_{s=2}^{t-1} E^{B(s)} c_i^{A(t-s)} - \sum_{l \neq 0} \sum_{q=1}^{t-2} c_i^{A(t-1-q)} c_l^{B(q)} W_{0l}^B \right] . \tag{4.203}$$

By using (4.29) and changing the summation index in the last sum to $s = q + 1$,

$$c_{(i0)}^{(t)} = c_i^{A(t)} + \frac{1}{E_i^{A0} - E_0^{A0}} \left[\sum_{s=2}^{t-1} E^{B(s)} c_i^{A(t-s)} \right.$$

$$\left. - \sum_{s=2}^{t-1} c_i^{A(t-s)} \sum_{l \neq 0} c_l^{B(s-1)} W_{0l}^B \right] = c_i^{A(t)} , \tag{4.204}$$

where we have taken into account that the sum over l gives $E^{B(s)}$ according to (4.24) and thus the two sums cancel.

Case *b)* $i = 0; j \neq 0$

Case *b)* is quite analogous to case *a)*, only the subsystems A and B (and the subscripts i and j) are interchanged. One gets

$$c_{(0j)}^{(t)} = c_j^{B(t)} \tag{4.205}$$

Case *c)* $i \neq 0; j \neq 0$

According to (4.200) we have

$$c_{(ij)}^{(t)} = \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \left[\sum_{s=1}^{t-2} \left(E^{A(s)} + E^{B(s)} \right) c_{(ij)}^{(t-s)} - \sum_{(kl)\neq(00)} c_{(kl)}^{(t-1)} W_{(ij)(kl)}^{AB} \right] \tag{4.206}$$

$$= \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \left[\sum_{s=1}^{t-2} (E^{A(s)} + E^{B(s)}) c_{(ij)}^{(t-s)} - c_i^{A(t-1)} W_{j0}^B - c_j^{B(t-1)} W_{i0}^A - \sum_{k\neq 0} c_{(kj)}^{(t-1)} W_{ik}^A - \sum_{l\neq 0} c_{(il)}^{(t-1)} W_{jl}^B \right],$$

where we have taken into account that the size consistency $E^{AB}(s) = E^{A(s)} + E^{B(s)}$ has been assumed for $s \leq t-1$ and that the term with $s = t-1$ in the first sum contains $c_{(ij)}^{(t-t+1)} = c_{(ij)}^{(1)} = 0$ and can be omitted ($i, j \neq 0$). Similarly to case *a)*, we have decomposed the double sum into three sums. Furthermore, we have substituted expression (4.179) for the coefficients $c_{(i0)}^{(t-1)}$ and $c_{(0j)}^{(t-1)}$, expression (4.181) for $W_{(ij)(kl)}^{AB}$, and performed the summation over the Kronecker symbols.

Now we substitute (4.180) for $c_{(ij)}^{(t-s)}$ and $c_{(ij)}^{(t-1)}$ choosing properly among the equivalent forms:

$$c_{(ij)}^{(t)} = \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \left[\sum_{s=1}^{t-2} E^{A(s)} \sum_{q=1}^{t-s-1} c_i^{A(t-s-q)} c_j^{B(q)} + \sum_{s=1}^{t-2} E^{B(s)} \sum_{p=1}^{t-s-1} c_j^{B(t-s-p)} c_i^{A(p)} - W_{j0}^B c_i^{A(t-1)} - W_{i0}^A c_j^{B(t-1)} - \sum_{k\neq 0} W_{ik}^A \sum_{q=1}^{t-2} c_k^{A(t-1-q)} c_j^{B(q)} - \sum_{l\neq 0} W_{jl}^B \sum_{p=1}^{t-2} c_l^{B(t-1-p)} c_i^{A(p)} \right]. \tag{4.207}$$

We separate out the term with $q = 1$ or $p = 1$ in the first two sums; then the upper limit of the remaining sums may be changed to $s = t-3$, because for $s = t-2$, q (p) assumes only the value $q = 1$ ($p = 1$), which is already separated out.

Also, we substitute the expansion of $c_i^{A(t-1)}$ and $c_j^{B(t-1)}$ according to (4.29):

$$
\begin{aligned}
c_{(ij)}^{(t)} = & \frac{1}{E_i^{A0}+E_j^{B0}-E_0^{A0}-E_0^{B0}}\Bigg\{\sum_{s=1}^{t-2} E^{A(s)} c_i^{A(t-1-s)} c_j^{B(1)} \\
& + \sum_{s=1}^{t-3} E^{A(s)} \sum_{q=2}^{t-s-1} c_i^{A(t-s-q)} c_j^{B(q)} \\
& + \sum_{s=1}^{t-2} E^{B(s)} c_j^{B(t-1-s)} c_i^{A(1)} + \sum_{s=1}^{t-3} E^{B(s)} \sum_{p=2}^{t-s-1} c_j^{B(t-s-p)} c_i^{A(p)} \\
& - \frac{W_{j0}^B}{E_i^{A0}-E_0^{A0}}\left[\sum_{s=1}^{t-2} E^{A(s)} c_i^{A(t-1-s)} - \sum_{k\neq 0} c_k^{A(t-2)} W_{ik}^A\right] \qquad (4.208) \\
& - \frac{W_{i0}^A}{E_j^{B0}-E_0^{B0}}\left[\sum_{s=1}^{t-2} E^{B(s)} c_j^{B(t-1-s)} - \sum_{l\neq 0} c_l^{B(t-2)} W_{jl}^B\right] \\
& - \sum_{k\neq 0} W_{ik}^A \sum_{q=1}^{t-2} c_k^{A(t-1-q)} c_j^{B(q)} - \sum_{l\neq 0} W_{jl}^B \sum_{p=1}^{t-2} c_l^{B(t-1-p)} c_i^{A(p)}\Bigg\}.
\end{aligned}
$$

The first and third terms contain the same sum as the first sums in the fifth and sixth term, respectively. After substituting the explicit expressions $c_i^{A(1)} = \frac{-W_{i0}^A}{E_i^{A0}-E_0^{A0}}$ and $c_j^{B(1)} = \frac{-W_{j0}^B}{E_j^{B0}-E_0^{B0}}$, we can bring them to a common denominator. Furthermore, we separate out the term $q=1$ and $p=1$ in the last two sums and substitute in them the expressions of $c_i^{A(1)}$ and $c_j^{B(1)}$:

$$
\begin{aligned}
c_{(ij)}^{(t)} = & \frac{-1}{(E_i^{A0}-E_0^{A0})(E_j^{B0}-E_0^{B0})}\left[\sum_{s=1}^{t-2} E^{A(s)} c_i^{A(t-1-s)} W_{j0}^B \right. \\
& \left. + \sum_{s=1}^{t-2} E^{B(s)} c_j^{B(t-1-s)} W_{i0}^A\right] \\
& + \frac{1}{E_i^{A0}+E_j^{B0}-E_0^{A0}-E_0^{B0}}\left[\sum_{s=1}^{t-3} E^{A(s)} \sum_{q=2}^{t-s-1} c_i^{A(t-s-q)} c_j^{B(q)}\right. \\
& \left. + \sum_{s=1}^{t-3} E^{B(s)} \sum_{p=2}^{t-s-1} c_j^{B(t-s-p)} c_i^{A(p)}\right] \qquad (4.209) \\
& + \frac{W_{i0}^A}{E_j^{B0}-E_0^{B0}} \sum_{l\neq 0} c_l^{B(t-2)} W_{jl}^B + \frac{W_{j0}^B}{E_i^{A0}-E_0^{A0}} \sum_{k\neq 0} c_k^{A(t-2)} W_{ik}^A
\end{aligned}
$$

$$
\begin{aligned}
&- \sum_{k\neq 0} W_{ik}^{A} c_k^{A(t-2)} \frac{-W_{j0}^{B}}{E_j^{B0} - E_0^{B0}} - \sum_{k\neq 0} W_{ik}^{A} \sum_{q=2}^{t-2} c_k^{A(t-1-q)} c_j^{B(q)} \\
&- \sum_{l\neq 0} W_{jl}^{B} c_l^{B(t-2)} \frac{-W_{i0}^{A}}{E_i^{A0} - E_0^{A0}} - \sum_{l\neq 0} W_{jl}^{B} \sum_{p=2}^{t-2} c_l^{B(t-1-p)} c_i^{A(p)} \Bigg] .
\end{aligned}
$$

Now we identify the definitions of $c_j^{B(1)}$ and $c_i^{A(1)}$ in the first and second terms, respectively. The terms containing the products $W_{i0}^{A} W_{jl}^{B}$ and $W_{j0}^{B} W_{ik}^{A}$ can be brought to a common denominator:

$$
\begin{aligned}
c_{(ij)}^{(t)} = & \sum_{s=1}^{t-2} E^{A(s)} c_i^{A(t-1-s)} \frac{c_j^{B(1)}}{E_i^{A0} - E_0^{A0}} + \sum_{s=1}^{t-2} E^{B(s)} c_j^{B(t-1-s)} \frac{c_i^{A(1)}}{E_j^{B0} - E_0^{B0}} \\
& + \frac{1}{(E_i^{A0} - E_0^{A0})(E_j^{B0} - E_0^{B0})} \Bigg[W_{i0}^{A} \sum_{l\neq 0} c_l^{B(t-2)} W_{jl}^{B} \\
& \qquad + W_{j0}^{B} \sum_{k\neq 0} c_k^{A(t-2)} W_{ik}^{A} \Bigg] \qquad (4.210) \\
& + \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \Bigg[\sum_{s=1}^{t-3} E^{A(s)} \sum_{q=2}^{t-s-1} c_i^{A(t-s-q)} c_j^{B(q)} \\
& + \sum_{s=1}^{t-3} E^{B(s)} \sum_{p=2}^{t-s-1} c_j^{B(t-s-p)} c_i^{A(p)} - \sum_{k\neq 0} W_{ik}^{A} \sum_{q=2}^{t-2} c_k^{A(t-1-q)} c_j^{B(q)} \\
& \qquad - \sum_{l\neq 0} W_{jl}^{B} \sum_{p=2}^{t-2} c_l^{B(t-1-p)} c_i^{A(p)} \Bigg] .
\end{aligned}
$$

In the third term we can again identify coefficients $c_i^{A(1)}$ and $c_j^{B(1)}$; they can be combined with the second and first term, respectively:

$$
\begin{aligned}
c_{(ij)}^{(t)} = & \frac{c_j^{B(1)}}{E_i^{A0} - E_0^{A0}} \Bigg[\sum_{s=1}^{t-2} E^{A(s)} c_i^{A(t-1-s)} - \sum_{k\neq 0} c_k^{A(t-2)} W_{ik}^{A} \Bigg] \\
& + \frac{c_i^{A(1)}}{E_j^{B0} - E_0^{B0}} \Bigg[\sum_{s=1}^{t-2} E^{B(s)} c_j^{B(t-1-s)} - \sum_{l\neq 0} c_l^{B(t-2)} W_{jl}^{B} \Bigg] \\
& + \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \Bigg[\sum_{s=1}^{t-3} E^{A(s)} \sum_{q=2}^{t-s-1} c_i^{A(t-s-q)} c_j^{B(q)}
\end{aligned}
$$

$$
\begin{aligned}
&- \sum_{k\neq 0} W^A_{ik} \sum_{q=2}^{t-2} c_k^{A(t-1-q)} c_j^{B(q)} \qquad (4.211)\\
&+ \sum_{s=1}^{t-3} E^{B(s)} \sum_{p=2}^{t-s-1} c_j^{B(t-s-p)} c_i^{A(p)} - \sum_{l\neq 0} W^B_{jl} \sum_{p=2}^{t-2} c_l^{B(t-1-p)} c_i^{A(p)} \Bigg] .
\end{aligned}
$$

The first two terms contain the definition of $c_i^{A(t-1)}$ and $c_j^{B(t-1)}$ according to (4.29). In the third term we have a summation over all values of $1 \leq s \leq t-3$ and $2 \leq q \leq t-2$ that satisfy $t-s-q \geq 1$; it can be rewritten as a summation $2 \leq q \leq t-2$ and $1 \leq s \leq t-q-1$. (The same holds for the fifth term by replacing p with q.) We get

$$
\begin{aligned}
c_{(ij)}^{(t)} &= c_j^{B(1)} c_i^{A(t-1)} + c_i^{A(1)} c_j^{B(t-1)} \qquad (4.212)\\
&+ \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \Bigg\{ \sum_{q=2}^{t-2} \Bigg[\sum_{s=1}^{t-q-1} E^{A(s)} c_i^{A(t-s-q)} \\
&\qquad - \sum_{k\neq 0} W^A_{ik} c_k^{A(t-1-q)} \Bigg] c_j^{B(q)} \\
&+ \sum_{q=2}^{t-2} \Bigg[\sum_{s=1}^{t-q-1} E^{B(s)} c_j^{B(t-s-q)} - \sum_{l\neq 0} W^B_{jl} c_l^{B(t-1-q)} \Bigg] c_i^{A(q)} \Bigg\} .
\end{aligned}
$$

Multiplying and dividing by $E_i^{A0} - E_0^{A0}$ and $E_j^{B0} - E_0^{B0}$, we can identify the formula (4.29) for $c_i^{A(t-q)}$ and $c_j^{B(t-q)}$:

$$
\begin{aligned}
c_{(ij)}^{(t)} &= c_j^{B(1)} c_i^{A(t-1)} + c_i^{A(1)} c_j^{B(t-1)} \qquad (4.213)\\
&+ \frac{1}{E_i^{A0} + E_j^{B0} - E_0^{A0} - E_0^{B0}} \Bigg[(E_i^{A0} - E_0^{A0}) \sum_{q=2}^{t-2} c_i^{A(t-q)} c_j^{B(q)} \\
&\qquad + (E_j^{B0} - E_0^{B0}) \sum_{q=2}^{t-2} c_j^{B(t-q)} c_i^{A(q)} \Bigg] .
\end{aligned}
$$

The two sums are equal – one of them is going "upward," another "downward"— and the denominator cancels:

$$
\begin{aligned}
c_{(ij)}^{(t)} &= c_j^{B(1)} c_i^{A(t-1)} + c_i^{A(1)} c_j^{B(t-1)} + \sum_{q=2}^{t-2} c_i^{A(t-q)} c_j^{B(q)} \\
&= \sum_{q=1}^{t-1} c_i^{A(t-q)} c_j^{B(q)} . \qquad (4.214)
\end{aligned}
$$

Q.E.D.

6. Löwdin's partitioning method

Let us consider the secular equation in the m-dimensional linear space of the orthonormalized basis functions $\{\varphi_i\}$:

$$\mathbf{Hc} = E\mathbf{c} \ . \tag{4.215}$$

Here $\mathbf{H}$ is a Hermitian $m \times m$ matrix, and $\mathbf{c}$ is an m-dimensional vector. Löwdin's partitioning permits one to reduce, at least formally, the dimensionality of the secular equation (4.215). This possibility is of interest from a conceptual point of view; in addition, the results obtained in the partitioning procedure can be used as a common starting point for developing most different perturbational schemes.

Let us divide the m basis functions φ_i into two groups. The first group—we may call it "primary"—consists of those n functions that we consider—from that or another reason—more interesting. (The subspace of these functions is often called the "model space.") The remaining $m - n$ functions may, accordingly, called "secondary" (or "external").

It is known from mathematics that matrices may be decomposed into blocks and that the usual matrix multiplication rules also apply to these block matrices. We shall assign subscript "1" to the set of functions in the primary subspace and subscript "2" to the functions in the secondary subspace. Then the matrix equation (4.215) becomes

$$\begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} = E \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} \ . \tag{4.216}$$

Here $\mathbf{c}_1$ is the n-dimensional vector formed of the first n elements of vector $\mathbf{c}$; the remaining $m - n$ elements of $\mathbf{c}$ form vector $\mathbf{c}_2$. Matrices $\mathbf{H}_{11}$ and $\mathbf{H}_{22}$ contain the matrix elements $\mathbf{H}_{ij}$ in the upper left and lower right corners of the original matrix $\mathbf{H}$, respectively, and describe the interactions between the functions in the same (primary or secondary) subspace. Obviously, $\mathbf{H}_{11}$ and $\mathbf{H}_{22}$ are Hermitian and have the dimensions $n \times n$ and $(m - n) \times (m - n)$, respectively. The matrix elements of $\mathbf{H}$ connecting the primary and secondary subspaces form blocks $\mathbf{H}_{12}$ and $\mathbf{H}_{21}$ of dimensions $n \times (m-n)$ and $(m-n) \times n$. As a consequence of the Hermiticity of the original matrix $\mathbf{H}$, these matrices are connected by the relationship

$$\mathbf{H}_{21} = \mathbf{H}_{12}^{\dagger} \ . \tag{4.217}$$

Now, we write out (4.216) in an expanded form

$$\begin{aligned} \mathbf{H}_{11}\mathbf{c}_1 + \mathbf{H}_{12}\mathbf{c}_2 &= E\mathbf{c}_1 \\ \mathbf{H}_{21}\mathbf{c}_1 + \mathbf{H}_{22}\mathbf{c}_2 &= E\mathbf{c}_2 \end{aligned} \tag{4.218}$$

and solve the second equation for $\mathbf{c}_2$. We have

$$(\mathbf{H}_{22} - E)\mathbf{c}_2 = -\mathbf{H}_{21}\mathbf{c}_1 \tag{4.219}$$

i.e.,

$$\mathbf{c}_2 = -(\mathbf{H}_{22} - E)^{-1}\mathbf{H}_{21}\mathbf{c}_1 \ . \tag{4.220}$$

We substitute this result into the first equation of the system in (4.218):

$$\mathbf{H}_{11}\mathbf{c}_1 - \mathbf{H}_{12}(\mathbf{H}_{22} - E)^{-1}\mathbf{H}_{21}\mathbf{c}_1 = E\mathbf{c}_1 \ , \tag{4.221}$$

that is

$$[\mathbf{H}_{11} - \mathbf{H}_{12}(\mathbf{H}_{22} - E)^{-1}\mathbf{H}_{21}]\mathbf{c}_1 = E\mathbf{c}_1 \ . \tag{4.222}$$

Equation (4.222) is the eigenvalue equation of the $n \times n$ *effective* Hamiltonian matrix

$$\mathbf{H}_{\text{eff}} = \mathbf{H}_{\text{eff}}(E) = \mathbf{H}_{11} - \mathbf{H}_{12}(\mathbf{H}_{22} - E)^{-1}\mathbf{H}_{21} \ . \tag{4.223}$$

It is important to stress that (4.223) is fully equivalent with the original secular equation (4.215) for all eigenvectors of the latter, which are not exactly orthogonal to the primary subspace—i.e., for which $\mathbf{c}_1 \neq 0$. (It is easy to see that $\mathbf{c}_1 = \mathbf{0}$ or $\mathbf{c}_2 = \mathbf{0}$ is possible only if matrix $\mathbf{H}$ is block-diagonal, i.e. $\mathbf{H}_{12} = \mathbf{H}_{21}^{\dagger} = \mathbf{0}$. Then the solutions can be obtained by considering the eigenvalue equation in the primary and secondary subspaces independently of each other.)

There are two difficulties connected with forming the effective Hamiltonian (4.223). The first is that it depends explicitly on the eigenvalue E we are looking for. Therefore, (4.222) is not a true eigenvalue equation but a pseudo-eigenvalue one, which can be solved by iterations only. An even bigger obstacle is that one has to compute the inverse matrix $(\mathbf{H}_{22} - E)^{-1}$—more precisely, the inverse of the matrix $\mathbf{H}_{22} - E\mathbf{1}_{22}$ where $\mathbf{1}_{22}$ is the $(m-n) \times (m-n)$ unit matrix corresponding to block 2—necessary to describe the effect of the secondary set of functions on the eigenvalue problem in the primary subspace; moreover, this inverse must be calculated as a function of E.

$\mathbf{H}_{\text{eff}}$ is an $n \times n$ matrix, therefore, for any value of E for which matrix $\mathbf{H}_{22} - E\mathbf{1}$ is not singular, one obtains n eigenvectors and eigenvalues. Obviously, only those solutions of (4.222) have a meaning, for which the eigenvalue obtained is equal to the value E used for constructing the matrix $\mathbf{H}_{\text{eff}}(E)$. If we consider a nondegenerate solution of (4.215), then only a single solution is meaningful. However, as stressed previously, by substituting different values of E into the pseudo-eigenvalue equation, all the solutions of (4.215) can be, at least in principle, recovered from (4.222) and (4.220).

Depending on the selection of the primary subspace and the different expansions (approximations) introduced for $(\mathbf{H}_{22} - E)^{-1}$, Löwdin's partitioning scheme can be used to derive most different perturbational expressions for the approximate treatment of the eigenvalue problem (4.215).

Example

As a simple example, let us consider the case when $n = 1$, i.e., the primary group consists of only one function φ_1. Then we may assume $c_1 = 1$ (this is equivalent to the use of the intermediate normalization in the Rayleigh–Schrödinger PT) and we get from (4.222), by expanding the matrix product,

$$E = H_{11} - \sum_{\substack{j,k \\ (j,k\neq 1)}} H_{1j}[(\mathbf{H}_{22} - E\mathbf{1})^{-1}]_{jk} H_{k1} \,. \tag{4.224}$$

Now, let us approximate the inverse in (4.224) by taking into account only the diagonal part of $\mathbf{H}_{22}$ and approximating E by H_{11}. Then

$$[(\mathbf{H}_{22} - E\mathbf{1})^{-1}]_{jk} \approx (H_{jj} - H_{11})^{-1}\delta_{jk} \tag{4.225}$$

and we get

$$E \approx H_{11} - \sum_{\substack{j,k \\ (j,k\neq 1)}} \frac{H_{1j}\delta_{jk}H_{k1}}{H_{jj} - H_{11}} = H_{11} - \sum_{\substack{j \\ (j\neq 1)}} \frac{|H_{1j}|^2}{H_{jj} - H_{11}} \tag{4.226}$$

which is nothing else than the second-order Rayleigh–Schrödinger formula for the case when the unperturbed Hamiltonian consists of the diagonal part of $\mathbf{H}$, whereas its off-diagonal part is considered the perturbation. (This is the so-called Epstein–Nesbet partitioning of the Hamiltonian into unperturbed and perturbed parts.)

As another example, let us utilize the identity

$$(\mathbf{A} + \mathbf{B})^{-1} = (\mathbf{1} + \mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1} \tag{4.227}$$

that one can check trivially by multiplying the right-hand side with $(\mathbf{A} + \mathbf{B})$ from the right:

$$(\mathbf{1} + \mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1}(\mathbf{A} + \mathbf{B}) = (\mathbf{1} + \mathbf{A}^{-1}\mathbf{B})^{-1}(\mathbf{1} + \mathbf{A}^{-1}\mathbf{B}) = \mathbf{1} \tag{4.228}$$

We denote $\Delta E = E - E_0$, E_0 being (any) zero-order approximation to the energy E, define $\mathbf{A}$ and $\mathbf{B}$ as

$$\begin{aligned} \mathbf{A} &= \mathbf{H}_{22} - E_0\mathbf{1}_{22} \\ \mathbf{B} &= -\Delta E\mathbf{1}_{22} \,, \end{aligned} \tag{4.229}$$

and apply (4.227) to express $(\mathbf{H}_{22} - E)^{-1} = (\mathbf{A} + \mathbf{B})^{-1}$. We get, by taking into account that $\mathbf{B}$ is proportional to the unit matrix,

$$(\mathbf{H}_{22} - E)^{-1} = [\mathbf{1}_{22} - \Delta E(\mathbf{H}_{22} - E_0\mathbf{1}_{22})^{-1}](\mathbf{H}_{22} - E_0\mathbf{1}_{22})^{-1} \,. \quad (4.230)$$

By denoting

$$\mathbf{R} = (\mathbf{H}_{22} - E_0\mathbf{1}_{22})^{-1} \quad (4.231)$$

the effective Hamiltonian matrix (4.223) becomes

$$\mathbf{H}_{\text{eff}} = \mathbf{H}_{11} - \mathbf{H}_{12}(\mathbf{1}_{22} - \Delta E\mathbf{R})^{-1}\mathbf{R}\mathbf{H}_{21} \,. \quad (4.232)$$

It can be seen that the second term in the parentheses is a minor correction, assuming that the zero-order approximation E_0 is not very bad. This form is well suited for iterations and can be used also as a starting point of different series expansions.

In the particular case of a one-dimensional model space and of the choice $E_0 = H_{11}$, we get an equation for the energy-shift ΔE as

$$\Delta E = -\mathbf{H}_{12}(\mathbf{1}_{22} - \Delta E\mathbf{R})^{-1}\mathbf{R}\mathbf{H}_{21} \,. \quad (4.233)$$

Notes

Section 1.

The classical perturbational technique of Lord Rayleigh was adapted by Schrödinger to quantum mechanics [1]. (For another early —and apparently independent—application of the perturbational expansion we refer to the paper of Born and Oppenheimer [2].)

The treatment presented in this chapter is mainly a straightforward generalization of the approach in which the author was taught perturbation theory by using the classical textbook of Blokhincev [3]. This, in particular, pertains to the way of introducing the perturbational parameter λ in Eq. (4.3).

Section 1.2.

The recursive character of the Rayleigh–Schrödinger perturbation theory was utilized by Knowles *et al* [4] to perform high-order perturbational calculations for small model systems. They used concise notations resembling our "compact matrix notations". We will need the explicit expressions for the coefficients $c_{ij}^{(l)}$ for studying the size–consistency problem in Section 6.2. (These expressions were included e.g., in [5], too.)

Section 1.3.

Standard treatment, see e.g., [6].

Section 1.4.

As noted in Section 1.2, the theorem about the accuracy of the expectation value also follows from the stationary properties of the exact wave functions. The proof given here follows the treatment in [7].

The possibility of expressing the exact energy corrections up to order $2n+1$ by using the wave function corrections up to order n has been known since the fifties [8, 9] (also see [6, 10]). The

proof presented here essentially follows that in [11]. (The author is indebted to Dr. Emili Besalú for a copy of his unpublished manuscript.)

Section 2.
Originally introduced by Hylleraas [12]. Here we give a standard treatment, see e.g., [13]. For a generalization for the case of a non-Hermitian unperturbed Hamiltonian see [14].

Section 3.
Standard material; e.g., [15]

Section 4.
Originally introduced in [16,17]. Here we give a standard treatment following e.g., [18].

Section 5.
Originally introduced in [19]. May be considered as the theoretical background for any effective Hamiltonian formalism, including Heisenberg's theory of magnetism.

Section 6.
The size consistency of the Rayleigh–Schrödinger perturbational theory is usually proved by involved diagrammatic methods [20]. Our results indicate this to be unnecessary.

The considerations based on the uniqueness of the perturbation series (Section 6.1.) may be (but are not claimed to be) original. The proof by induction given in section 6.2. is probably original.

Suggested readings

The exhaustive series of Löwdin *"Studies in Perturbation Theory"*, parts I to XIV [21–33] is recommended, as well as the excellent survey of Hirschfelder *et al* [6].

References

1 E. Schrödinger, Ann. Phys. **80**, 437 (1926)
2 M. Born and J.R. Oppenheimer, Ann. Phys. **84**, 457 (1927)
3 D. I. Blokhincev, *Osnovy Kvantovoi Mekhaniki*, Vycshaia Shkola, Moskow 1963.
4 P.J. Knowles, S. Somasundram, N.C. Handy and K. Hirao, Chem. Phys. Letters **113**, 8 (1985)
5 I. Mayer, *Fejezetek a Kvantumkémiából*, BME Mérnöktovábbképző Int., Budapest 1987.
6 J.O. Hirschfelder, W. Byers Brown and S.T. Epstein, Adv. Quantum Chem. **1**, 255 (1964)
7 L. Zülicke, *Quantenchemie*, Deutcher Verlag der Wissenschaften, Berlin 1973.
8 A. Dalgarno and A.L. Stewart, Proc. Roy. Soc (London) **A 238**, 269 (1956)
9 F. Dupont-Bourdelet, J. Tillieu and J. Guy, J. Phys. Radium **21**, 776 (1960)
10 E. Brändas and O. Goscinski, Phys. Rev. **A 1** 552 (1970)
11 E. Besalú, Institute of Computational Chemistry, University of Girona, Girona, Spain (unpublished work).
12 E.A. Hylleraas, Z. Physik **65**, 209 (1930)
13 S.T. Epstein, *The Variation Method in Quantum Chemistry*, Academic Press, New York 1974.
14 I. Mayer, Mol. Phys. **89**, 515 (1996)
15 L.D. Landau and I. M. Lifshitz, *Kvantovaya Mekhanika*, Nauka, Moscow 1974.
16 l. Brillouin, J. Phys. Radium (7) **3**, 373 (1932)
17 E. Wigner, Math. u. naturw. Anz. ungar Akad. Wiss. **53**, 477 (1935)
18 J.M. Ziman, *Elements of Advanced Quantum Theory*, University Press, Cambridge 1969.

19 P.-O. Löwdin, J. Chem. Phys. **19**, 1396 (1951)

20 P. Jørgensen and J. Simons, *Second Quantization-Based Methods in Quantum Chemistry*, Academic Press, Mew York 1981.

21 P.-O. Löwdin, J. Mol. Spectry, **10**, 12 (1963)

22 P.-O. Löwdin, J. Mol. Spectry, **13**, 326 (1964) (Parts II and III of the series.)

23 P.-O. Löwdin, J. Math. Phys. **3**, 969 (1962)

24 P.-O. Löwdin, J. Math. Phys. **3**, 1171 (1962)

25 P.-O. Löwdin, J. Mol. Spectry, **14**, 112 (1964)

26 P.-O. Löwdin, J. Mol. Spectry, **14**, 119 (1964)

27 P.-O. Löwdin, J. Mol. Spectry, **14**, 131 (1964)

28 P.-O. Löwdin, J. Math. Phys. **6** 1341 (1965)

29 P.-O. Löwdin, Phys. Rev. **139** A357 (1965)

30 P.-O. Löwdin, J. Chem. Phys. **43** S175 (1965)

31 P.-O. Löwdin, in *Perturbation Theory and its Application in Quantum Mechanics* ed. C. H. Wilcox, John Wiley, 1966.

32 P.-O. Löwdin, Int. J. Quantum Chem. **2**, 867 (1968)

33 P.-O. Löwdin and O. Goscinski, Int. J. Quantum Chem. Symp. **5**, 665 (1971)

Chapter 5

DETERMINANT WAVE FUNCTIONS

1. Spin orbitals

In quantum chemistry one usually deals with many-electron systems, and the wave functions depend on the coordinates of all the electrons. Except for very special cases (like He, H_2), the wave functions applied are constructed of one-electron functions as building blocks, requesting the many-electron wave functions to satisfy the Pauli principle (and trying, of course to get an approximation to the exact solution that is as good as possible).

As is known, the electron has its own internal angular momentum, the spin, which can have two different projections $+\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$ on a selected spatial axis z. For one-electron problems—provided that the spin-free Hamiltonian usually applied in quantum chemistry is used—there is no need to consider this degree of freedom explicitly. However, we cannot avoid dealing with the spin variables when many-electron wave functions are considered.

The one-electron wave function that depends on both the spatial coordinates x, y, z (radius vector $\vec{r}$) and the spin coordinate σ is usually called the "spin orbital," in order to be distinguished from the (spatial) orbital depending only on $\vec{r}$:

$$\psi = \psi(x, y, z, \sigma) = \psi(\vec{r}, \sigma) \ . \tag{5.1}$$

As the spin of the electron can have only two values $s_z = \pm\frac{1}{2}$ (in units of $\hbar$, i.e., atomic units), the independent variable σ can assume only two values $\pm\frac{1}{2}$, and we get two functions—one for $\sigma = \frac{1}{2}$, another for $\sigma = -\frac{1}{2}$:

$$\psi_1(\vec{r}) = \psi(\vec{r}, -\tfrac{1}{2}) \ ; \qquad \psi_2(\vec{r}) = \psi(\vec{r}, -\tfrac{1}{2}) \ . \tag{5.2}$$

In this case the one-electron wave function can be considered as a two-row vector

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

$$\psi = \begin{pmatrix} \psi_1(\vec{r}) \\ \psi_2(\vec{r}) \end{pmatrix} , \tag{5.3}$$

the first row of which corresponds to $\sigma = \frac{1}{2}$ and the second to $\sigma = -\frac{1}{2}$.

One usually separates out the spin dependence by introducing the "spin up" ($s_z = +\frac{1}{2}$) and "spin down" ($s_z = -\frac{1}{2}$) states ("spin functions") α and β, respectively:

$$\alpha = \alpha(\sigma) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} ; \quad \beta = \beta(\sigma) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5.4}$$

(This means that, as a function of σ, $\alpha(\frac{1}{2}) = 1$, $\alpha(-\frac{1}{2}) = 0$, whereas $\beta(\frac{1}{2}) = 0$, $\beta(-\frac{1}{2}) = 1$.) Then the spin orbital ψ can be presented as

$$\psi(\vec{r}, \sigma) = \psi_1(\vec{r})\alpha(\sigma) + \psi_2(\vec{r})\beta(\sigma) . \tag{5.5}$$

However, the use of such "general spin orbitals" is cumbersome and one usually applies spin orbitals corresponding to pure spin up and spin down one-electron spin states: $\psi(\vec{r})\alpha(\sigma)$ or $\psi(\vec{r})\beta(\sigma)$.

The spin of the electron is described by operators acting in the two-dimensional space of the spin functions α and β; accordingly, they are represented by 2×2 matrices:

$$\hat{s}_z = \frac{1}{2}\boldsymbol{\sigma}_z ; \quad \hat{s}_x = \frac{1}{2}\boldsymbol{\sigma}_x ; \quad \hat{s}_y = \frac{1}{2}\boldsymbol{\sigma}_y \tag{5.6}$$

(in atomic units; otherwise $\hat{s}_x = \frac{\hbar}{2}\,\boldsymbol{\sigma}_x$, etc.) where

$$\boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} ; \quad \boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} ; \quad \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{5.7}$$

are the so-called Pauli matrices.

It is easy to check directly the following basic properties of the operators in (5.6):

a)

$$\hat{s}^2 = \hat{s}_x^2 + \hat{s}_y^2 + \hat{s}_z^2 = \frac{3}{4}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{5.8}$$

b) The spin operators in (5.6) satisfy the commutation relationships that any operators of angular momentum type should obey in quantum mechanics:

$$\hat{s}_x\hat{s}_y - \hat{s}_y\hat{s}_x = i\hat{s}_z \tag{5.9}$$

and one can perform circular permutations $x \to y \to z \to x$.

c) The spin functions α and β are eigenfunctions of operator $\hat{s}_z$ with eigenvalues $+\frac{1}{2}$ and $-\frac{1}{2}$, respectively; in addition both are eigenfunctions of the operator

$\hat{s}^2$ with the expected eigenvalue $S(S+1) = \frac{1}{2}\left(\frac{1}{2}+1\right) = \frac{3}{4}$. ($S$ is the spin quantum number of the system; for a single electron $S = \frac{1}{2}$.)

For computing expectation values or matrix elements of different operators, one has to integrate over all the variables on which the wave function depends. As the spin variable σ assumes only the discrete values $+\frac{1}{2}$ and $-\frac{1}{2}$, practically this means that one has to perform a *summation* over the spins. This is often included in the integration, and one writes

$$\int d\tau \equiv \sum_{\sigma} \int dv \; ; \qquad dv = dxdydz \tag{5.10}$$

to denote together the summation over the spin variables and integration over the spatial ones.

The spin functions α and β are orthonormalized:

$$\begin{aligned} \sum_{\sigma} \alpha^*(\sigma)\alpha(\sigma) &= \sum_{\sigma} \beta^*(\sigma)\beta(\sigma) = 1 \\ \sum_{\sigma} \alpha^*(\sigma)\beta(\sigma) &= \sum_{\sigma} \beta^*(\sigma)\alpha(\sigma) = 0 \end{aligned} \tag{5.11}$$

(Owing to the definition in (5.4) of the spin functions α and β as two-dimensional vectors, these equalities can also be written compactly by using matrix notations as $\alpha^\dagger\alpha = \beta^\dagger\beta = 1$; $\alpha^\dagger\beta = \beta^\dagger\alpha = 0$.)

Owing to the orthonormalization property in (5.11), the use of spin orbitals built up of the pure spin functions α and β simplifies to a great extent the summations over the spins: we get a nonzero value only if we have the same spin function in the "bra" and in the "ket" for each electron separately—and then the spins give simply a factor 1.

Assuming $\psi_i = \varphi_i(\vec{r})\gamma_i(\sigma)$ with γ_i being either α or β, the previous considerations give for the matrix element of a spin-free operator $\hat{L}$:

$$\begin{aligned} \langle\psi_1(\vec{r},\sigma)|\hat{L}|\psi_2(\vec{r},\sigma)\rangle &= \langle\varphi_1(\vec{r})\gamma_1(\sigma)|\hat{L}|\varphi_2(\vec{r})\gamma_2(\sigma)\rangle \\ &= \int d\tau\varphi_1^*(\vec{r})\gamma_1^*(\sigma)\hat{L}\varphi_2(\vec{r})\gamma_2(\sigma) \\ &= \sum_{\sigma}\gamma_1^*\gamma_2 \int dv\varphi_1^*(\vec{r})\hat{L}\varphi_2(\vec{r}) \\ &= \delta_{\gamma_1\gamma_2}\int dv\varphi_1^*(\vec{r})\hat{L}\varphi_2(\vec{r}) \\ &= \langle\varphi_1(\vec{r})|\hat{L}|\varphi_2(\vec{r})\rangle\delta_{\gamma_1\gamma_2} \; . \end{aligned} \tag{5.12}$$

Similarly, for a two-electron matrix element (integral) of a spin-independent two-particle operator $\hat{g}(1,2)$ we have

$$\begin{aligned} &\langle\psi_1(1)\psi_2(2)|\hat{g}(1,2)|\psi_3(1)\psi_4(2)\rangle \\ &\qquad = \delta_{\gamma_1\gamma_3}\delta_{\gamma_2\gamma_4}\langle\varphi_1(\vec{r}_1)\varphi_2(\vec{r}_2)|\hat{g}(1,2)|\varphi_3(\vec{r}_1)\varphi_4(\vec{r}_2)\rangle \; , \end{aligned} \tag{5.13}$$

where we have used the common shorthands "1" and "2" for denoting all the spatial and spin variables of electron 1 and 2, respectively: $\psi_i(1) \equiv \psi_i(\vec{r}_1, \sigma_1)$, $\psi_j(2) \equiv \psi_j(\vec{r}_2, \sigma_2)$.

2. Many-electron spin states

The Born–Oppenheimer Hamiltonian does not act on spin variables, so it commutes with the spin operators. Therefore the eigenfunctions of the Hamiltonian can be classified according to the eigenvalues of the resulting spin-square of the system ($\hat{S}^2$) and of the projection of the overall spin on the selected — usually the z—axis ($\hat{S}_z$). Let us denote by S the largest projection possible for the given eigenstate of $\hat{S}^2$: $|S_z|_{\max} = S$. (Quantum number S is a half-integer if the number of electrons is odd and it is an integer if the number of electrons is even.) Then, in accord with the quantum mechanical rules for the angular momentum, the eigenvalue of $\hat{S}^2$ is $S(S+1)$ and the state has a spin multiplicity $2S+1$, because for a given S there are $2S+1$ possible projections S_z, namely $-S, -S+1, \ldots, S-1, S$. Accordingly, the different spin states are usually referred as

$$S = 0 \quad \text{singlet}$$

$$S = \frac{1}{2} \quad \text{doublet}$$

$$S = 1 \quad \text{triplet}$$

$$S = \frac{3}{2} \quad \text{quadruplet}$$

$$S = 2 \quad \text{quintuplet}$$

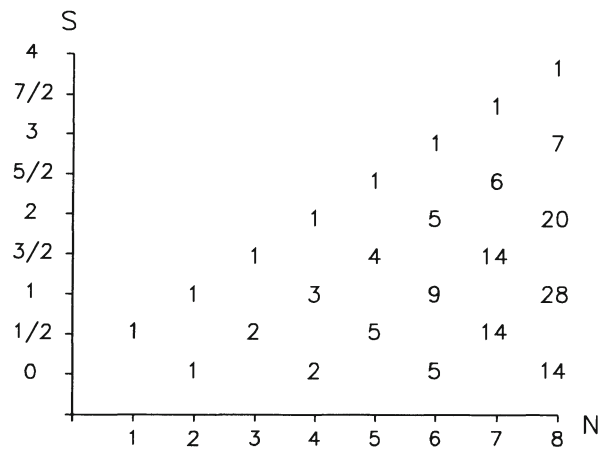
and so on.

The treatment of spin functions is complicated by the fact that in the case when the number of the electrons assigned to singly occupied spatial orbitals exceeds two, then usually one can construct a number of different many-electron spin functions corresponding to the same eigenvalues of operators $\hat{S}^2$ and $\hat{S}_z$. (The doubly occupied orbitals represent singlet subsystems and may be omitted from consideration.) The existence of several independent spin functions is sometimes called "spin degeneracy," but this is only a traditional nomenclature, as there is no true degeneracy—the different spin functions usually correspond to different energies.

The possible values of S for the case of N electrons and N singly occupied spatial orbitals vary from 0 (if N is even) or $\frac{1}{2}$ (if N is odd) to a maximal value of $S = \frac{N}{2}$. For a given value of S_z the numbers n_α and n_β of electrons with spins α and β, respectively, are, obviously, $n_\alpha = \frac{N}{2} + S_z$ and $n_\beta = \frac{N}{2} - S_z$. The number of different *elementary spin functions* (simple products of αs and βs) in which

n_α electrons have spin α and n_β electrons have spin β is $\binom{N}{n_\alpha} = \binom{N}{n_\beta}$—the number of possible selection of n_α positions among the N ones to which spin α is to be attributed—or, which is the same, the number of possible selections of n_β positions to which spin β is attributed. (The equality follows from the fact that $n_\alpha + n_\beta = N$). The individual elementary spin functions are, in general, not "pure spin function," i.e., are not eigenfunctions of operator $\hat{S}^2$; forming appropriate linear combinations of them one can construct pure spin functions with all values of S for which the given $S_z = \frac{1}{2}(n_\alpha - n_\beta)$ is possible (i.e., for which $|S_z| \leq S$).

The number of linearly independent spin eigenfunctions corresponding to different numbers N of electrons on singly occupied spatial orbitals and the different spin quantum numbers S is given by the so-called "branching diagram," shown on the figure, resembling to some extent Pascal's triangle known in mathematics. In the branching diagram each number gives the number of independent spin eigenfunctions for the given values of N and S indicated on the axes. If one considers a system of N electrons in the spin eigenstate with the spin quantum number S and couples it with an additional electron, then the spin of the latter can be either parallel or antiparallel with the resulting spin of the first N electrons; thus one can form two different $N+1$-electron spin functions: one with $S' = S + \frac{1}{2}$ and another with $S'' = S - \frac{1}{2}$. From this one obtains the rule by which the branching diagram is formed: each number is the sum of the two closest numbers that can be found to the left from it: an N-electron state with quantum number S can be formed by adding one electron to each of the $(N-1)$-electron states with quantum numbers $S \pm \frac{1}{2}$.



The total number of the different spin states for a given N can be determined as follows. The elementary spin functions containing n_α αs and $n_\beta = N - n_\alpha$ βs form a basis of the N-electron spin functions having $S_z = \frac{1}{2}(n_\alpha - n_\beta)$. The

spin eigenfunctions are linear combinations of them, thus one can form as many different spin eigenfunctions with the given N and S_z values as the number of the respective elementary spin functions. Every spin state has a component with the lowest possible projection $S_z = 0$ (if N is even) or $S_z = \frac{1}{2}$ (if N is odd), therefore the number of different N-electron spin states with all possible values of S is the same as the number $\binom{N}{n_\alpha}$ of the elementary spin functions with the lowest S_z value. (In this case $n_\alpha = \frac{N}{2}$ if N is even and $n_\alpha = \frac{N+1}{2}$ if N is odd.) This number gives the *sum* of the numbers in the column of the branching diagram corresponding to the given N.

The previous considerations hold for the case if each of the N spatial orbitals is singly occupied. In practice one often has to consider cases in which some orbitals can also be doubly occupied whereas others are left empty. Then the number of configurations (determinants—see the next section) can be determined by distributing n_α spins α and n_β spins β between the N spatial orbitals independently from each other. (It is no more necessary to assume $N = n_\alpha + n_\beta$). This leads to $\binom{N}{n_\alpha}\binom{N}{n_\beta}$—a number that increases extremely sharply with N, n_α, and n_β.

There are several methods for constructing the different spin eigenfunctions. Their systematic study represents a topic for independent monographs and requires heavy use of group theory. We shall not consider this problem in any detail but mention only one way of constructing a linearly independent (but nonorthogonal) set of spin eigenfunctions. One has to select as many elementary spin functions as the number of different spin functions for the given value of S and N according to the branching diagram and subject them to spin projection by using Löwdin's spin projection operator discussed in Section 4.3 of Chapter 6.

3. Slater determinants

The many-electron wave function should be antisymmetric with respect to all the electrons, i.e., it should change its sign under the simultaneous interchange of the spatial and spin coordinates of any pairs of electrons:

$$\Psi(1, 2, ..., i, ..., j, ..., N) = -\Psi(1, 2, ..., j, ..., i, ..., N) \ . \tag{5.14}$$

In principle, it is always possible to present, at least formally, the antisymmetric wave function as a product of purely spatial and purely spin components, but usually these are rather complicated and require the use of elaborated group theoretical methods. The situation is simple only in the two-electron case in which either one requires the spatial part of the wave function to be symmetric and the spin part to be antisymmetric with respect of the two electrons, and gets a singlet state, or, vice versa, one requires the spin part to be symmetric and

the spatial part to be antisymmetric, and gets a triplet. Such a construction is, in general, impossible for a larger number of electrons; in their case the spatial part is obtained by symmetrizing with respect of some electron pairs and antisymmetrizing with respect the others, whereas the roles are interchanged for the spin part, providing the product of the spatial and spin parts to be fully antisymmetric. (We shall not go into details of such a theory.)

In general practice it is more suitable to treat together the spatial and spin coordinates of the individual electrons, and build up the wave function in the form of one or more "Slater determinants," providing the required antisymmetry by way of their construction.

Considering N electrons and N spin orbitals ψ_k, one can build the determinant wave function as

$$\Psi = \Psi(1,2,...,N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(1) & \psi_2(1) & \psi_3(1) & \dots & \psi_N(1) \\ \psi_1(2) & \psi_2(2) & \psi_3(2) & \dots & \psi_N(2) \\ \vdots & \vdots & \vdots & & \vdots \\ \psi_1(N) & \psi_2(N) & \psi_3(N) & \dots & \psi_N(N) \end{vmatrix} . \tag{5.15}$$

Here $\psi_k(i) \equiv \psi_k(\vec{r}_i, \sigma_i)$ is the kth spin orbital considered as a function of the spatial and spin coordinates of the ith electron.

By virtue of the known properties of determinants, the wave function Ψ defined in (5.15) changes its sign by interchanging any pairs of electrons—because that means interchanging two rows of the determinant. As will be discussed below in Section 6, the factor $\frac{1}{\sqrt{N!}}$ provides that the many-electron determinant wave function Ψ in (5.15) is normalized to unity if it is built up of orthonormalized one-electron spin orbitals.

The Slater determinant wave function Ψ is antisymmetric not only under the interchange of the electrons (rows of the determinant) but also under the interchange of the spin orbitals (columns of the determinant). Therefore Ψ cannot contain the same spin orbital twice—it would be $\Psi \equiv 0$—in agreement with the more pictorial formulation of the Pauli principle ("there can be only one electron in a given state"). One can produce two independent spin orbitals from a given spatial orbital (one with spin α, another with spin β), so each spatial orbital can be filled at most twice. More generally, it follows from the properties of the determinants that all the spin orbitals must be linearly independent, otherwise $\Psi \equiv 0$.

3.1. Two-electron examples

In the specific case of two electrons, the properly constructed Slater determinant can be expressed as a product of a spatial part and a spin part, one of

them being symmetric, the other antisymmetric. It is simultaneously an eigenfunction of the overall spin $\hat{S}^2$ as well, so it can be characterized by its spin multiplicity.

Singlet: a determinant constructed by using a doubly filled spatial orbital φ. The two spin orbitals accordingly are

$$\psi_1(\vec{r},\sigma) = \varphi(\vec{r})\alpha(\sigma) \; ; \quad \psi_2(\vec{r},\sigma) = \varphi(\vec{r})\beta(\sigma) \; . \tag{5.16}$$

By using the shorthand notations $\varphi(i) = \varphi(\vec{r}_i)$, $\alpha(i) = \alpha(\sigma_i)$, and $\beta(i) = \beta(\sigma_i)$, the determinant wave function can be written as

$$\begin{aligned} {}^1\Psi(1,2) &= \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi(1)\alpha(1) & \varphi(1)\beta(1) \\ \varphi(2)\alpha(2) & \varphi(2)\beta(2) \end{vmatrix} \\ &= \frac{1}{\sqrt{2}}[\varphi(1)\alpha(1)\varphi(2)\beta(2) - \varphi(1)\beta(1)\varphi(2)\alpha(2)] \\ &= \underbrace{\varphi(1)\varphi(2)}_{\text{symmetric}} \underbrace{\frac{1}{\sqrt{2}}[\alpha(1)\beta(2) - \beta(1)\alpha(2)]}_{\text{antisymmetric}} \; . \end{aligned} \tag{5.17}$$

Triplet: One should observe that two different spatial orbitals are necessary to construct a triplet wave function of two electrons. In the simplest case they are orthogonal, but it is not imperative. Accordingly, we have two spin orbitals of α spin $\psi_1(\vec{r},\varphi) = \varphi_1(\vec{r})\alpha(\sigma)$ and $\psi_2(\vec{r},\sigma) = \varphi_2(\vec{r})\alpha(\sigma)$, and the $S_z = 1$ triplet determinant wave function is

$$\begin{aligned} {}^3\Psi(1,2) &= \frac{1}{\sqrt{2}} \begin{vmatrix} \varphi_1(1)\alpha(1) & \varphi_2(1)\alpha(1) \\ \varphi_1(2)\alpha(2) & \varphi_2(2)\alpha(2) \end{vmatrix} \\ &= \frac{1}{\sqrt{2}}[\varphi_1(1)\alpha(1)\varphi_2(2)\alpha(2) - \varphi_2(1)\alpha(1)\varphi_1(2)\alpha(2)] \\ &= \underbrace{\frac{1}{\sqrt{2}}[\varphi_1(1)\varphi_2(2) - \varphi_2(1)\varphi_1(2)]}_{\text{antisymmetric}} \underbrace{\alpha(1)\alpha(2)}_{\text{symmetric}} \; . \end{aligned} \tag{5.18}$$

Replacing both spins by β, one gets the $S_z = -1$ component of the same triplet configuration; however, the $S_z = 0$ component of the triplet cannot be given as a single determinant, only as a sum of two determinants. Similarly, one needs two determinants to write out a singlet wave function if it is constructed by using two different spatial orbitals.

We note that these two-determinant two-electron wave functions still can be presented as a product of a symmetric spatial part and an antisymmetric spin part in the singlet case, and vice versa in the triplet one. Assuming φ_1 and φ_2

to be orthogonal spatial orbitals, the normalized functions are

$$^{1,3}\Psi = \frac{1}{2}\left\{ \begin{vmatrix} \varphi_1(1)\alpha(1) & \varphi_2(1)\beta(1) \\ \varphi_1(2)\alpha(2) & \varphi_2(2)\beta(2) \end{vmatrix} \mp \begin{vmatrix} \varphi_1(1)\beta(1) & \varphi_2(1)\alpha(1) \\ \varphi_1(2)\beta(2) & \varphi_2(2)\alpha(2) \end{vmatrix} \right\} , \tag{5.19}$$

where the upper sign corresponds to the singlet and the lower one to the triplet.

Expanding the determinants in (5.19) one readily gets

$$\begin{aligned} ^{1,3}\Psi &= \frac{1}{2}\{\varphi_1(1)\alpha(1)\varphi_2(2)\beta(2) - \varphi_2(1)\beta(1)\varphi_1(2)\alpha(2) \\ &\quad \mp\varphi_1(1)\beta(1)\varphi_2(2)\alpha(2) \pm \varphi_2(1)\alpha(1)\varphi_1(2)\beta(2)\} \end{aligned} \tag{5.20}$$

i.e.,

$$^{1}\Psi = \frac{1}{2}\underbrace{[\varphi_1(1)\varphi_2(2) + \varphi_1(2)\varphi_2(1)]}_{\text{symmetric}}\underbrace{[\alpha(1)\beta(2) - \beta(1)\alpha(2)]}_{\text{antisymmetric}} \tag{5.21}$$

and

$$^{3}\Psi = \frac{1}{2}\underbrace{[\varphi_1(1)\varphi_2(2) - \varphi_2(1)\varphi_1(2)]}_{\text{antisymmetric}}\underbrace{[\alpha(1)\beta(2) + \beta(1)\alpha(2)]}_{\text{symmetric}} \tag{5.22}$$

4. The antisymmetrizing operator

The determinant wave function Ψ defined in (5.15) can be compactly written out by using the antisymmetrizing operator ("antisymmetrizer") $\hat{\mathcal{A}}$:

$$\Psi = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)] , \tag{5.23}$$

where $\psi_1(1)\psi_2(2)\ldots\psi_N(N)$ is a so-called Hartree product, and operator $\hat{\mathcal{A}}$ is defined as

$$\hat{\mathcal{A}} = \frac{1}{\sqrt{N!}}\sum_{\hat{P}\in S_N}(-1)^p\hat{P} . \tag{5.24}$$

The summation in the right-hand side of (5.24) is performed for all possible permutations $\hat{P}$ of the N numbers (electrons) $1, 2, ..., N$—these permutations form the so called "symmetric group" S_N. The factor $(-1)^p$ represents the *parity* of permutation $\hat{P}$. (Some of the basic properties of permutations are summarized in Appendix X.) It is to be stressed that definitions (5.23)-(5.24) represent only another way of writing down the determinant in the right-hand side of (5.15). In fact, applying operator $\hat{\mathcal{A}}$ to the Hartree product $\psi_1(1)\psi_2(2)\ldots\psi_N(N)$ one gets

$$\Psi = \frac{1}{\sqrt{N!}}\sum_{\hat{P}\in S_N}(-1)^p\psi_1(P_1)\psi_2(P_2)\ldots\psi_N(P_N)] , \tag{5.25}$$

which is the explicit expansion of the determinant in (5.15) with the aid of the different permutations.

4.1. The projection character of the antisymmetrizing operator

The antisymmetrizing operator $\hat{\mathcal{A}}$ is, up to a constant coefficient, a projection operator that extracts from any wave function that component that is antisymmetric with respect to the interchange of any pair of electrons.

It is known that an operator is a projection operator (Appendix VII) if and only if it is a) *idempotent*, i.e., the square of the operator equals the operator itself, and b) *Hermitian*. We are going here to demonstrate these properties for the antisymmetrizing operator $\hat{\mathcal{A}}$. (As idempotency holds up to a constant factor only, operator $\hat{\mathcal{A}}$ is only proportional to the respective projection operator. As already noted, the coefficient $\frac{1}{\sqrt{N!}}$ in the definition of $\hat{\mathcal{A}}$ is chosen as to provide the N-electron determinant wave function built up of orthonormalized spin orbitals to be normalized to unity.)

Idempotency up to a constant factor $\sqrt{N!}$

Let us consider the operator product $\hat{\mathcal{A}}\hat{\mathcal{A}}$

$$\begin{aligned}\hat{\mathcal{A}}\hat{\mathcal{A}} &= \frac{1}{\sqrt{N!}}\sum_{\hat{P}\in S_N}(-1)^p\hat{P}\frac{1}{\sqrt{N!}}\sum_{\hat{Q}\in S_N}(-1)^q\hat{Q} \\ &= \frac{1}{N!}\sum_{\hat{P}\in S_N}\sum_{\hat{Q}\in S_N}(-1)^{p+q}\hat{P}\hat{Q}\,. \end{aligned} \tag{5.26}$$

If permutation $\hat{Q}$ runs over all the permutations, then—for a given $\hat{P}$—the permutation $\hat{R} = \hat{P}\hat{Q}$ also runs over all the permutations, only in a different order. Therefore, the second sum on the right-hand side of (5.26), which is over $\hat{Q} \in S_N$, may be replaced by a summation over $\hat{R} \in S_N$. The parity of permutation $\hat{R}$ is the product of the parities of the two permutations, so $(-1)^{p+q} = (-1)^r$. Accordingly, we may write

$$\hat{\mathcal{A}}\hat{\mathcal{A}} = \frac{1}{N!}\sum_{\hat{P}\in S_N}\left[\sum_{\hat{R}\in S_N}(-1)^r\hat{R}\right]\,. \tag{5.27}$$

The outer sum over $\hat{P} \in S_N$ indicates that one has to add $N!$ identical operators (the number of different permutations $\hat{P}$ is $N!$ and the operator in the braces is independent of $\hat{P}$), so we get

$$\hat{\mathcal{A}}\hat{\mathcal{A}} = \sum_{\hat{R}\in S_N}(-1)^r\hat{R} = \sqrt{N!}\hat{\mathcal{A}}\,. \tag{5.28}$$

Q.E.D.

It follows from this result that the strictly idempotent projector is $\frac{1}{\sqrt{N!}}\hat{\mathcal{A}}$, i.e., the operator which one obtains by replacing factor $\frac{1}{\sqrt{N!}}$ with $\frac{1}{N!}$ in definition (5.24).

Hermiticity

An operator $\hat{L}$ is Hermitian (self-adjoint) if $\hat{L}^\dagger = L$, which means $\langle \hat{L}\psi|\varphi\rangle = \langle\psi|\hat{L}\varphi\rangle$ for every ψ and φ. (The Hermitian adjoint $\hat{L}^\dagger$ of the operator $\hat{L}$ is defined as the operator for which $\langle L^\dagger\psi|\varphi\rangle \equiv \langle\psi|\hat{L}\varphi\rangle$.) Accordingly, we have to show that

$$\langle\hat{\mathcal{A}}\Phi(1,2,\ldots,N)|\Psi(1,2,\ldots,N)\rangle = \langle\Phi(1,2,\ldots,N)|\hat{\mathcal{A}}\Psi(1,2,\ldots,N)\rangle \tag{5.29}$$

for any N-electron functions Φ and Ψ.

It is obvious that

$$\langle\Phi(1,2,\ldots,N)|\Psi(1,2,\ldots,N)\rangle = \langle\hat{Q}\Phi(1,2,\ldots,N)|\hat{Q}\Psi(1,2,\ldots,N)\rangle \tag{5.30}$$

for every permutation operator $\hat{Q} \in S_N$, because the right-hand side differs from the left-hand side only in the change of the notations used for the variables of integration. Using this observation, let us investigate

$$\begin{aligned} &\langle\hat{\mathcal{A}}\Phi(1,2,\ldots,N)|\Psi(1,2,\ldots,N)\rangle \qquad (5.31)\\ = \;& \frac{1}{\sqrt{N!}}\sum_{\hat{P}\in S_N}(-1)^p\langle\hat{P}\Phi(1,2,\ldots,N)|\Psi(1,2,\ldots,N)\rangle\,. \end{aligned}$$

We insert on both sides the permutation $\hat{Q} = \hat{P}^{-1}$ (the inverse of $\hat{P}$) and get

$$\begin{aligned} &\langle\hat{\mathcal{A}}\Phi(1,2,\ldots,N)|\Psi(1,2,\ldots,N)\rangle\\ = \;& \frac{1}{\sqrt{N!}}\sum_{\hat{P}\in S_N}(-1)^p\langle\hat{P}^{-1}\hat{P}\Phi(1,2,\ldots,N)|\hat{P}^{-1}\Psi(1,2,\ldots,N)\rangle\\ = \;& \frac{1}{\sqrt{N!}}\sum_{\hat{P}\in S_N}(-1)^p\langle\Phi(1,2,\ldots,N)|\hat{P}^{-1}\Psi(1,2,\ldots,N)\rangle\,. \qquad (5.32) \end{aligned}$$

If $\hat{P}$ runs over all the possible permutations, then the same holds for its inverse $\hat{Q} = \hat{P}^{-1}$, as well; the parities of $\hat{P}$ and $\hat{P}^{-1} = \hat{Q}$ are also the same, so we get

$$\begin{aligned} &\langle\hat{\mathcal{A}}\Phi(1,2,\ldots,N)|\Psi(1,2,\ldots,N)\rangle\\ = \;& \frac{1}{\sqrt{N!}}\sum_{\hat{Q}\in S_N}(-1)^q\langle\Phi(1,2,\ldots,N)|\hat{Q}\Psi(1,2,\ldots,N)\rangle \qquad (5.33)\\ = \;& \langle\Phi(1,2,\ldots,N)|\hat{\mathcal{A}}\Psi(1,2,\ldots,N)\rangle\,. \end{aligned}$$

Q.E.D.

Two-electron example for the projection character of the antisymmetrizing operator.

The Hartree product $\Psi_H(1,2) = \varphi(1)\psi(2)$ of the (different) spin orbitals φ and ψ can be considered as a sum of an antisymmetric component

$$\Phi_A(1,2) = \frac{1}{2}[\varphi(1)\psi(2) - \varphi(2)\psi(1)] \tag{5.34}$$

and of a symmetric component

$$\Phi_S(1,2) = \frac{1}{2}[\varphi(1)\psi(2) + \varphi(2)\psi(1)] \ . \tag{5.35}$$

As it is easy to see, $\Psi_H = \Phi_A + \Phi_B$, indeed. $\Phi_A(1,2)$ differs from the determinant wave function $\hat{\mathcal{A}}[\varphi(1)\psi(2)] = \frac{1}{\sqrt{2}}[\varphi(1)\psi(2) - \varphi(2)\psi(1)]$ only by a coefficient $\sqrt{2}$. This means that in the case of two electrons the projector on the antisymmetric subspace is—in accord with (5.28)—the operator

$$\hat{P}_A = \frac{1}{\sqrt{2}}\hat{\mathcal{A}} = \frac{1}{2}[\hat{I} - \hat{P}_{12}] \ , \tag{5.36}$$

where $\hat{I}$ is the identity operator (identical permutation) and $\hat{P}_{12}$ is the permutation operator interchanging the coordinates of the two electrons.

If one applies $\hat{P}_A$ to the antisymmetric function $\Phi_A(1,2)$, the latter should remain unchanged, as it is already in the antisymmetric subspace and is therefore an eigenvector of $\hat{P}_A$ with an eigenvalue 1: $\hat{P}_A\Phi_A(1,2) = \Phi_A(1,2)$. Indeed:

$$\begin{aligned}
\hat{P}_A\Phi_A(1,2) &= \frac{1}{2}(\hat{I} - \hat{P}_{12})\frac{1}{2}[\varphi(1)\psi(2) - \varphi(2)\psi(1)] \\
&= \frac{1}{4}\left\{\hat{I}[\varphi(1)\psi(2) - \varphi(2)\psi(1)] - \hat{P}_{12}[\varphi(1)\psi(2) - \varphi(2)\psi(1)]\right\} \\
&= \frac{1}{4}\{\varphi(1)\psi(2) - \varphi(2)\psi(1) - \varphi(2)\psi(1) + \varphi(1)\psi(2)\} \\
&= \frac{1}{2}[\varphi(1)\psi(2) - \varphi(2)\psi(1)] = \Phi_A(1,2) \ .
\end{aligned} \tag{5.37}$$

In the case of more than two electrons, the Hartree product $\varphi_1(1)\varphi_2(2)\ldots\varphi_N(N)$ cannot be presented as a sum of a fully symmetric and a fully antisymmetric component—it has also components that are symmetric for some electrons and antisymmetric for others. However, *if the spin orbitals φ_i are linearly independent*, the Hartree product has a nonzero component that is antisymmetric with respect to all the electrons. This component is selected (up to a constant factor) when forming the determinant wave function $\hat{\mathcal{A}}[\varphi_1(1)\varphi_2(2)\ldots\varphi_N(N)]$. As electrons are fermions, we are interested only in such fully antisymmetric wave functions.

According to the known properties of determinants, one can interchange the rows and columns without changing the determinant. As applied to definition (5.25), that means that its right-hand side can be rewritten also as

$$\hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)] = \frac{1}{\sqrt{N!}} \sum_{\hat{Q}\in S_N} (-1)^q \psi_{Q_1}(1)\psi_{Q_2}(2)\ldots\psi_{Q_N}(N) \,. \tag{5.38}$$

It may be instructive to consider how the equivalence of (5.25) and (5.38) can be established in the framework of the present formalism.

Let us define operator $\hat{Q}'$ that performs permutation $\hat{Q}$ of the indices (subscripts) of the spin orbitals:

$$\hat{Q}'\psi_1(1)\psi_2(2)\ldots\psi_N(N) = \psi_{Q_1}(1)\psi_{Q_2}(2)\ldots\psi_{Q_N}(N) \,. \tag{5.39}$$

We have obviously the following equality

$$\begin{aligned} \hat{Q}'\hat{Q}\psi_1(1)\psi_2(2)\ldots\psi_N(N) &= \hat{Q}'\psi_1(Q_1)\psi_2(Q_2)\ldots\psi_N(Q_N) \\ &= \psi_{Q_1}(Q_1)\psi_{Q_2}(Q_2)\ldots\psi_{Q_N}(Q_N) \\ &= \psi_1(1)\psi_2(2)\ldots\psi_N(N) \end{aligned} \tag{5.40}$$

for every $\hat{Q} \in S_N$, because the simultaneous permutation of the electrons and of the spin orbitals changes only the order of the factors in the product. Therefore—as applied to Hartree products—we have $\hat{Q}'\hat{Q} = \hat{I}$ or $\hat{Q}' = \hat{Q}^{-1}$.

Let us now consider

$$\Psi = \frac{1}{\sqrt{N!}} \sum_{\hat{P}\in S_N} (-1)^p \hat{P}\psi_1(1)\psi_2(2)\ldots\psi_N(N) \tag{5.41}$$

and insert operator $\hat{I} = \hat{Q}\hat{Q}'$ after $\hat{P}$:

$$\Psi = \frac{1}{\sqrt{N!}} \sum_{\hat{P}\in S_N} (-1)^p \hat{P}\hat{Q}\hat{Q}'\psi_1(1)\psi_2(2)\ldots\psi_N(N) \,. \tag{5.42}$$

This equality is valid for every $\hat{Q}$, so we may select $\hat{Q} = \hat{P}^{-1}$, and get

$$\Psi = \frac{1}{\sqrt{N!}} \sum_{\hat{P}\in S_N} (-1)^p \hat{Q}'\psi_1(1)\psi_2(2)\ldots\psi_N(N) \,. \tag{5.43}$$

If $\hat{P}$ runs over all the permutations, the same holds for $\hat{Q} = \hat{P}^{-1}$, as well; their parity is also equal, $(-1)^p = (-1)^q$; we have, therefore

$$\begin{aligned} \Psi &= \frac{1}{\sqrt{N!}} \sum_{\hat{Q}\in S_N} (-1)^q \hat{Q}'\psi_1(1)\psi_2(2)\ldots\psi_N(N) \\ &= \frac{1}{\sqrt{N!}} \sum_{\hat{Q}\in S_N} (-1)^q \psi_{Q_1}(1)\psi_{Q_2}(2)\ldots\psi_{Q_N}(N) \,. \end{aligned} \tag{5.44}$$

Q.E.D.

4.2. Commutation properties of the antisymmetrizing operator

Operator $\hat{\mathcal{A}}$ commutes with any operator that is symmetric with respect to the coordinates of all the different electrons. The most important examples are the Born–Oppenheimer Hamiltonian and the total spin operators $\hat{S}_z = \sum_i \hat{s}_z(i)$ and $\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2$, where $\hat{S}_x = \sum_i \hat{s}_x(i)$; $\hat{S}_y = \sum_i \hat{s}_y(i)$.

To see this, let us first consider an operator $\hat{L}$ representing the symmetric sum of one-electron operators $\hat{l}(i)$:

$$\hat{L} = \sum_{i=1}^{N} \hat{l}(i) \ . \tag{5.45}$$

We have to compare $\hat{\mathcal{A}}\hat{L}\Psi(1,2,\ldots,N)$ with $\hat{L}\hat{\mathcal{A}}\Psi(1,2,\ldots,N)$. To see that they are equal, it is sufficient to prove that $\sum_{i=1}^{N} \hat{l}(i)$ commutes with *every* permutation operator $\hat{P}$. For that reason we consider

$$\hat{P}\sum_{i=1}^{N} \hat{l}(i)\Psi(1,2,\ldots,N) = \sum_{i=1}^{N} \hat{l}(P_i)\Psi(P_1,P_2,\ldots,P_N) \tag{5.46}$$

and

$$\sum_{i=1}^{N} \hat{l}(i)\hat{P}\Psi(1,2,\ldots,N) = \sum_{i=1}^{N} \hat{l}(i)\Psi(P_1,P_2,\ldots,P_N) \ . \tag{5.47}$$

But every P_i in the argument of $\hat{l}$ on the right-hand side of (5.46) is an element of the set $\{1,2,\ldots,N\}$—that which is on the ith place in the sequence $P_1, P_2, \ldots, P_N$ —and each element of this set occurs once and only once among the set $\{P_1, P_2, \ldots, P_N\}$. Therefore $\sum_{i=1}^{N} \hat{l}(i) = \sum_{i=1}^{N} \hat{l}(P_i)$, and the right-hand sides of (5.46) and (5.47) are equal.
Q.E.D.

The proof is quite similar for the two-electron operators. Such operators are usually written in the form

$$\hat{G} = \sum_{i<j} \hat{g}(i,j) = \frac{1}{2} \sum_{\substack{i,j \\ (i\neq j)}} \hat{g}(i,j) \ . \tag{5.48}$$

The two forms indicated are equal because operator $\hat{G}$ is assumed to be symmetric with respect of electrons: $\hat{g}(i,j) = \hat{g}(j,i)$. Because of this symmetry,

one may rewrite (5.48) also in the form

$$\hat{G} = \sum_{\{i,j\}} \hat{g}(i,j) \ , \tag{5.49}$$

where the summation is over the distinct *pairs* $\{i,j\}$ of electrons, without specifying whether index i or j is the larger. Then applying permutation to $\hat{G}$ it will change to $\sum_{\{i,j\}} \hat{g}(P_i, P_j)$. But both P_i and P_j are elements of the set $\{1, 2, \ldots, N\}$, and they are different for $i \neq j$. In addition, we have different pairs $\{P_i, P_j\}$ for different pairs $\{i,j\}$. Therefore the sum will run over all the $\binom{N}{2}$ different combinations of the elements $\{1, 2, \ldots, N\}$, which again means that the *sums* are the same before and after the permutation is performed:

$$\sum_{\{i,j\}} \hat{g}(P_i, P_j) = \sum_{\{i,j\}} \hat{g}(i,j) \tag{5.50}$$

and, therefore,

$$\hat{P} \sum_{\{i,j\}} \hat{g}(i,j)\Psi(1,2,\ldots,N) = \sum_{\{i,j\}} \hat{g}(i,j)\hat{P}\Psi(1,2,\ldots,N) \ . \tag{5.51}$$

Q.E.D.

4.3. Antisymmetrization between distant subsystems

According to the Pauli principle, the wave function should be antisymmetric with respect to the interchange between any pairs of electrons. Let us consider two systems A and B, containing N_A and N_B electrons, respectively, and their properly antisymmetrized wave functions Ψ_A and Ψ_B, representing the eigenfunctions of the respective Hamiltonians $\hat{H}_A$ and $\hat{H}_B$, with eigenvalues E_A and E_B, respectively. We assume A and B to be at such a large distance from each other that any interaction between them can be completely neglected. Nevertheless, we may formally consider A and B as parts of a supersystem AB, containing $N_A + N_B$ electrons.

If one considers A and B separately, then the arguments of Ψ_A are the spatial and spin coordinates of the electrons with numbers $1, 2, \ldots, N_A$ and the argument of Ψ_B are those of electrons numbered $1, 2, \ldots, N_B$. In the case of the supersystem AB we have to number the electrons consecutively from 1 to $N_A + N_B$. When considering the Schrödinger equation of the supersystem, at first we shall use the scheme mentioned in Section 1.1 of Chapter 1, without requesting the wave function to be antisymmetric with respect to interchange of electrons between the two subsystems. Then we may assign the first N_A electrons to subsystem A and the remaining N_B ones to subsystem B. Thus

the arguments of Ψ_A are the coordinates of electrons $1, 2, \ldots, N_A$ and the arguments of Ψ_B are those of electrons $N_A+1, N_A+2, \ldots, N_A+N_B$: $\Psi_A = \Psi_A(1, 2, \ldots, N_A)$; $\Psi_B = \Psi_B(N_A+1, N_A+2, \ldots, N_A+N_B)$.

For this case the Hamiltonians $\hat{H}_A$ and $\hat{H}_B$ can be written, using atomic units (Appendix VI) as

$$\hat{H}_A = \sum_{i=1}^{N_A} \left[-\frac{1}{2}\Delta_i + V_A^N(\vec{r}_i) \right] + \frac{1}{2} \sum_{\substack{i,j=1 \\ (i \neq j)}}^{N_A} \frac{1}{r_{ij}} \tag{5.52}$$

and

$$\hat{H}_B = \sum_{i=N_A+1}^{N_A+N_B} \left[-\frac{1}{2}\Delta_i + V_B^N(\vec{r}_i) \right] + \frac{1}{2} \sum_{\substack{i,j=N_A+1 \\ (i \neq j)}}^{N_A+N_B} \frac{1}{r_{ij}} \tag{5.53}$$

with $V_A^N(\vec{r}_i), V_B^N(\vec{r}_i)$ being the potential energy caused by the nuclei in A and B, respectively, at the position of the ith electron. One has obviously (c.f. Section 1.1 of Chapter 1)

$$(\hat{H}_A + \hat{H}_B)\Psi_A\Psi_B = (E_A + E_B)\Psi_A\Psi_B \,. \tag{5.54}$$

The wave function $\Psi_A\Psi_B$ is not antisymmetric with respect to the electrons assigned to different subsystems. One could introduce trivially the full antisymmetrizer Eq.(5.24) for $N = N_A + N_B$ electrons, but the wave function would no longer be an eigenfunction of $\hat{H}_A + \hat{H}_B$.

In such a situation one may proceed in the following manner. If the subsystems are strictly noninteracting, then $V_A^N(\vec{r}_i)$, $V_B^N(\vec{r}_j)$, and $\frac{1}{r_{ij}}$ should vanish if electrons i and j belong to the subsystems B and A, respectively. Therefore, combining all such "cross-terms" into

$$\hat{H}_{\text{cross}} = \sum_{i=N_A+1}^{N_A+N_B} V_A^N(\vec{r}_i) + \sum_{j=1}^{N_A} V_B^N(\vec{r}_j) + \sum_{j=1}^{N_A} \sum_{i=N_A+1}^{N_B} \frac{1}{r_{ij}} \tag{5.55}$$

we have, for noninteracting subsystems,

$$\begin{aligned} &\hat{H}_{\text{cross}}\Psi_A\Psi_B \\ &\quad = \hat{H}_{\text{cross}}\Psi_A(1, 2, \ldots, N_A)\Psi_B(N_A+1, N_A+2, \ldots, N_A+N_B) = 0 \end{aligned} \tag{5.56}$$

We may add (5.56) to (5.54) and get

$$(\hat{H}_A + \hat{H}_B + \hat{H}_{\text{cross}})\Psi_A\Psi_B = (E_A + E_B)\Psi_A\Psi_B \tag{5.57}$$

Now, it is easy to see by trivial algebra that the sum $\hat{H}_A + \hat{H}_B + \hat{H}_{\text{cross}}$ has a form that is symmetric with respect to all the electrons—it is none other than the

usual Born–Oppenheimer Hamiltonian written out for the $N_A + N_B$ electrons of the supersystem AB:

$$\hat{H}_{AB} = \hat{H}_A + \hat{H}_B + \hat{H}_{\text{cross}} = \sum_{i=1}^{N_A+N_B} \left[-\frac{1}{2}\Delta_i + V_A^N(\vec{r}_i) + V_B^N(\vec{r}_i) \right] + \frac{1}{2} \sum_{\substack{i,j=1 \\ (i \neq j)}}^{N_A+N_B} \frac{1}{r_{ij}} \,. \tag{5.58}$$

Therefore, we have

$$\hat{H}_{AB}\Psi_A\Psi_B = (E_A + E_B)\Psi_A\Psi_B \,. \tag{5.59}$$

Now, we shall utilize the fact that the Hamiltonian $\hat{H}_{AB}$ is symmetric with respect to all the electrons, therefore it commutes with all their permutations (Section 4.2)—including those that interchange the electrons between the two subsystems. As the antisymmetrizer $\hat{\mathcal{A}}$ is a linear combination of permutations, we arrive at

$$\hat{\mathcal{A}}\hat{H}_{AB}\Psi_A\Psi_B = \hat{H}_{AB}\hat{\mathcal{A}}\Psi_A\Psi_B = (E_A + E_B)\hat{\mathcal{A}}\Psi_A\Psi_B \,, \tag{5.60}$$

i.e., the antisymmetrization of the electrons between noninteracting subsystems has no substantial effect on the separability of their Schrödinger equations: the overall *antisymmetrized* wave function $\Psi_{AB} = \hat{\mathcal{A}}\Psi_A\Psi_B$ is an eigenfunction of the supersystem Hamiltonian $\hat{H}_{AB}$ with an eigenvalue $E = E_A + E_B$, in full agreement with the notion of noninteracting subsystems.

We may add the following remark at this point. If the subsystems are at a finite distance, then $\hat{H}_{\text{cross}}\Psi_A\Psi_A$ does not vanish and one cannot apply the considerations discussed previously. This explains the origin of the dilemma, one encounters when wishes to study intermolecular interactions by using a perturbation theory in which the noninteracting molecules serve for defining the unperturbed problem: the wave function that is an eigenfunction of $\hat{H}^0 = \hat{H}_A + \hat{H}_B$ is not fully antisymmetric; if one performs its antisymmetrization, it ceases to be an eigenfunction of $\hat{H}^0$. We are not going to consider intermolecular perturbation theory in this book, but only note very briefly that there are two approaches for treating this problem: one can either incorporate the antisymmetrization into the the perturbational procedure as it is done in different variants of the so-called "symmetry adapted perturbation theory" (SAPT) or reformulate the whole problem by using second quantization technique, in which the full antisymmetry of every wave function is guaranteed by the formalism itself.

5. Invariance of the determinant wave function with respect to "mixing" the occupied orbitals

The following property of the determinant wave functions is of utmost importance. If one performs a "mixing" of the spin orbitals that are occupied

in a Slater determinant wave function $\Psi = \hat{\mathcal{A}}[\varphi_1(1)\varphi_2(2)\ldots\varphi_N(N)]$, i.e., subjects them to any *nonsingular* (having nonzero determinant) linear transformation between themselves, by replacing spin orbitals φ_i by the spin orbitals

$$\psi_j = \sum_{k=1}^{N} T_{kj}\varphi_k \tag{5.61}$$

then the determinant wave function remains invariant or changes only with a (physically unimportant) constant coefficient; the latter is equal to the determinant of the linear transformation:

$$\Psi' = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)] = \mathrm{Det}(\mathbf{T})\Psi\ . \tag{5.62}$$

If the transformation of the orbitals is unitary, then only the phase of the wave function can change.

Proof

We substitute explicitly the expansion (5.61) into the expression of the wave function Ψ':

$$\begin{aligned}\Psi' &= \hat{\mathcal{A}}\left[\sum_{k_1=1}^{N} T_{k_1 1}\varphi_{k_1}(1)\sum_{k_2=1}^{N} T_{k_2 2}\varphi_{k_2}(2)\ldots\sum_{k_N=1}^{N} T_{k_N N}\varphi_{k_N}(N)\right] \\ &= \sum_{k_1,k_2,\ldots,k_N=1}^{N} T_{k_1 1}T_{k_2 2}\ldots T_{k_N N}\hat{\mathcal{A}}[\varphi_{k_1}(1)\varphi_{k_2 2}(2)\ldots\varphi_{k_N}(N)]\end{aligned} \tag{5.63}$$

It is known that a determinant vanishes if two (or more) of its columns are the same. Therefore, only those terms survive in the last sum in (5.63) in which each orbital φ_{k_j} is different. Then the set $\{k_1, k_2, \ldots k_N\}$ is the same as the set $\{1, 2, \ldots, N\}$; i.e., each nonvanishing term is a determinant built up from the original spin orbitals φ_i, only the order of the latter is varying. (All the possible permutations $\hat{K}$ will occur in the sum and each will occur only once.) Such a determinant with a permuted order of orbitals is equal to the starting Ψ if the parity of the permutation $\hat{K}^{-1}$ by which the original order can be restored is even, and it is equal to $-\Psi$ if that parity is odd. Therefore the right-hand side of (5.63) is reduced to a sum over the permutations $\hat{K}$ with the respective parity factor. (The parity of $\hat{K}$ and $\hat{K}^{-1}$ is the same.) Thus we get

$$\Psi' = \sum_{\hat{K}\in S_N} (-1)^k T_{K_1 1}T_{K_2 2}\ldots T_{K_N N}\Psi\ . \tag{5.64}$$

Now, one can see that the factor on the right-hand side of (5.64) is none other than the determinant of matrix $\mathbf{T}$

$$\sum_{\hat{K}}(-1)^k T_{K_1 1}T_{K_2 2}\ldots T_{K_N N} \equiv \mathrm{Det}(\mathbf{T})\ , \tag{5.65}$$

therefore

$$\Psi' = \mathrm{Det}(\mathbf{T})\Psi \ . \tag{5.66}$$

Q.E.D.

The same result can also be obtained by using the known theorem about the determinant of a matrix product: if **A** and **B** are $m \times m$ matrices, then

$$\mathrm{Det}(\mathbf{AB}) = \mathrm{Det}(\mathbf{A})\mathrm{Det}(\mathbf{B}) \ . \tag{5.67}$$

Wave function Ψ is a Slater determinant; it can be written as a determinant of matrix **A**, the matrix elements of which are defined as $A_{ij} = \varphi_j(i)$:

$$\Psi = \frac{1}{\sqrt{N!}}\mathrm{Det}\{A_{ij}\} = \frac{1}{\sqrt{N!}}\mathrm{Det}(\mathbf{A}) \ . \tag{5.68}$$

Similarly, the determinant Ψ' built up of the spin orbitals ψ_j can be written as the determinant of matrix **B** having the elements $B_{ij} = \psi_j(i)$:

$$\Psi' = \frac{1}{\sqrt{N!}}\mathrm{Det}\{B_{ij}\} = \frac{1}{\sqrt{N!}}\mathrm{Det}(\mathbf{B}) \ . \tag{5.69}$$

But

$$B_{ij} = \psi_j(i) = \sum_{k=1}^{N} T_{kj}\varphi_k(i) = \sum_{k=1}^{N} A_{ik}T_{kj} = (\mathbf{AT})_{ij} \tag{5.70}$$

i.e.,

$$\mathbf{B} = \mathbf{AT} \ , \tag{5.71}$$

therefore

$$\Psi' = \frac{1}{\sqrt{N!}}\mathrm{Det}(\mathbf{B}) = \frac{1}{\sqrt{N!}}\mathrm{Det}(\mathbf{AT}) = \mathrm{Det}(\mathbf{T})\Psi \ . \tag{5.72}$$

Q.E.D.

If the transformation is unitary, $\mathbf{T} = \mathbf{U}$, $\mathbf{U}^\dagger = \mathbf{U}^{-1}$, then $|\mathrm{Det}(\mathbf{T})| = |\mathrm{Det}(\mathbf{U})| = 1$—the determinant of any unitary matrix has the absolute value 1. Therefore, the determinant wave function is *invariant* under the unitary transformation of the occupied orbitals—except, possibly, a physically irrelevant phase factor having absolute value 1.

We may note that the invariance properties discussed can be considered as generalizations of the well-known property according to which the value of a determinant does not change if one adds to each element of a column the elements of another column multiplied by an arbitrary constant factor.

An important consequence of the previous results is that every determinant wave function can be presented (disregarding a constant factor) by using orthonormalized spin orbitals. The latter can be obtained by subjecting the occupied orbitals to an orthonormalization procedure. For instance, Löwdin's symmetric orthogonalization (Section 2 of Chapter 3) is performed if one uses in (5.61) the transformation matrix $\mathbf{T} = \mathbf{S}^{-1/2}$. (Here $S_{ij} = \langle\varphi_i|\varphi_j\rangle$ is the overlap integral of the spin orbitals, of which the determinant Ψ is built up.) In general, one obtains orthonormalized orbitals by using any transformation matrix that can be written down as $\mathbf{T} = \mathbf{S}^{-1/2}\mathbf{V}$, where $\mathbf{V}$ is an arbitrary *unitary* matrix. In fact:

$$\begin{aligned}
\langle\psi_i|\psi_j\rangle &= \langle\sum_{l=1}^{N} T_{li}\varphi_l|\sum_{k=1}^{N} T_{kj}\varphi_k\rangle = \sum_{k,l=1}^{N} T_{li}^{*}\langle\varphi_l|\varphi_k\rangle T_{kj} \\
&= \sum_{k,l}(\mathbf{T}^{\dagger})_{il}S_{lk}T_{kj} = (\mathbf{T}^{\dagger}\mathbf{S}\mathbf{T})_{ij} \qquad (5.73) \\
&= (\mathbf{V}^{\dagger}\underbrace{\mathbf{S}^{-1/2}\mathbf{S}\mathbf{S}^{-1/2}}_{\mathbf{1}}\mathbf{V})_{ij} = (\mathbf{V}^{\dagger}\mathbf{V})_{ij} = \delta_{ij}\ ,
\end{aligned}$$

where we have taken into account the rule $(\mathbf{AB})^{\dagger} = \mathbf{B}^{\dagger}\mathbf{A}^{\dagger}$, as well as that $\mathbf{V}$ is unitary ($\mathbf{V}^{\dagger} = \mathbf{V}^{-1}$) and $\mathbf{S}^{-1/2}$ is Hermitian ($\mathbf{S}^{-1/2\dagger} = \mathbf{S}^{-1/2}$).

Based on the previous results, the following important conclusion can also be drawn: the determinant wave function is (up to a physically unimportant constant factor) uniquely determined by the *subspace* spanned by the occupied spin orbitals. One may select *any* set of N linearly independent vectors (one-electron functions) completely lying in that subspace and build up the determinant wave function with their use. Any such sets of N linearly independent functions are connected with each other by a linear transformation; consequently all of these sets give—disregarding a constant factor—*the same* N-electron determinant wave function.

6. Matrix elements between determinant wave functions: Löwdin's general formulae for nonorthogonal orbitals

Slater derived the matrix elements of different operators between Slater determinant wave functions for the case of orthonormalized spin orbitals. His results were generalized by Löwdin who admitted nonorthogonal orbitals, as well.

6.1. Overlap

In quantum chemistry one usually calls "overlap" or "overlap integral" the scalar product $\langle\varphi|\psi\rangle$ of any pair of functions φ and ψ.

6.1.1. General case

Let U and V be two single determinant wave functions built up of the spin orbitals u_i and v_j, respectively:

$$U = \hat{\mathcal{A}}[u_1(1)u_2(2)\dots u_N(N)] \tag{5.74}$$

and

$$V = \hat{\mathcal{A}}[v_1(1)v_2(2)\dots v_N(N)] \ . \tag{5.75}$$

The overlap of U and V is, accordingly,

$$\langle U|V\rangle = \langle \hat{\mathcal{A}}[u_1(1)u_2(2)\dots u_N(N)]|\hat{\mathcal{A}}[v_1(1)v_2(2)\dots v_N(N)]\rangle \ . \tag{5.76}$$

We move the antisymmetrizing operator $\hat{\mathcal{A}}$ from the "bra" to the "ket," utilizing its Hermiticity and idempotency properties $\hat{\mathcal{A}}^\dagger = \hat{\mathcal{A}}$ and $\hat{\mathcal{A}}^2 = \sqrt{N!}\hat{\mathcal{A}}$, discussed in Section 4. Equation (5.76) is then rewritten as

$$\langle U|V\rangle = \sqrt{N!}\langle u_1(1)u_2(2)\dots u_N(N)|\hat{\mathcal{A}}[v_1(1)v_2(2)\dots v_N(N)]\rangle \ . \tag{5.77}$$

In order to expand (5.77), we shall use the presentation (5.38) of the antisymmetrizer, in which it is acting on the spin orbital indices:

$$\begin{aligned}\langle U|V\rangle &= \sqrt{N!}\langle u_1(1)u_2(2)\dots u_N(N)| \\ &\qquad \times \frac{1}{\sqrt{N!}} \sum_{\hat{Q}\in S_N} (-1)^q v_{Q_1}(1)v_{Q_2}(2)\dots v_{Q_N}(N)\rangle \\ &= \sum_{\hat{Q}\in S_N} (-1)^q \langle u_1(1)u_2(2)\dots u_N(N)|v_{Q_1}(1)v_{Q_2}(2)\dots v_{Q_N}(N)\rangle \ .\end{aligned} \tag{5.78}$$

We have to integrate over the coordinates of all the electrons, including summations over the spins. As it is easy to see, each integral on the right-hand side of (5.78) factorizes into a product of integrals for the individual electrons; the variables of integrations then need not be explicitly indicated:

$$\langle U|V\rangle = \sum_{\hat{Q}\in S_N} (-1)^q \langle u_1|v_{Q_1}\rangle\langle u_2|v_{Q_2}\rangle\dots\langle u_N|v_{Q_N}\rangle \ . \tag{5.79}$$

The right-hand side of (5.79) is none other than the definition of the determinant D formed of the overlaps $\langle u_i|v_j\rangle$ of the individual spin orbitals. Thus we get Löwdin's general formula for the overlap of the determinant wave functions U and V:

$$\langle U|V\rangle = D = \mathrm{Det}\{\langle u_i|v_j\rangle\} = \mathrm{Det}(\mathbf{S}^{uv}) \ . \tag{5.80}$$

This result means that the overlap integral for two many-electron determinant wave functions (Slater determinants) is equal to the determinant of the mutual overlap integral matrix $\mathbf{S}^{uv}$ of the spin orbitals used to build up the determinant wave functions in question ($S^{uv}_{ij} = \langle u_i|v_j\rangle$).

6.1.2. Factorization

In the next sections we will need such presentations of the overlap integral $\langle U|V\rangle$, in which the integrations over some variables are separated out.

a) Let us consider (5.79), choose a subscript i ($1 \leq i \leq N$), and separate the term originating from the integration over the coordinates of the ith electron as

$$\begin{aligned}\langle U|V\rangle &= \sum_{\hat{Q}\in S_N} (-1)^q \langle u_i|v_{Q_i}\rangle\langle u_1|v_{Q_1}\rangle\ldots \\ &\qquad \times\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots\langle u_n|v_{Q_n}\rangle \\ &= \sum_{j=1}^{N}\langle u_i|v_j\rangle \sum_{\hat{Q}\in S_N} \delta_{jQ_i}(-1)^q\langle u_1|v_{Q_1}\rangle\ldots \\ &\qquad \times\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots\langle u_n|v_{Q_n}\rangle\ .\end{aligned} \tag{5.81}$$

In the first line only the order of the factors is changed as compared with (5.79), whereas in the second one we have utilized the trivial fact that the subscript Q_i is one of the integers between 1 and N, therefore $\sum_{j=1}^{N}\langle u_i|v_j\rangle\delta_{jQ_i} = \langle u_i|v_{Q_i}\rangle$. Comparing with (5.80), one can easily recognize in (5.81) the expansion of the determinant D in terms of its ith row, according to Laplace theorem known from mathematics:

$$\langle U|V\rangle = D = \mathrm{Det}\{\langle u_i|v_j\rangle\} = \sum_{j=1}^{N}\langle u_i|v_j\rangle D(i|j)\ , \tag{5.82}$$

where $D(i|j) = (-1)^{i+j}M(i|j)$ is the i,jth "cofactor" (or "algebraic complement") of determinant D; it is the minor $M(i|j)$ (determinant obtained by deleting the ith row and jth column from the original one) multiplied with the factor $(-1)^{i+j}$. Therefore the coefficient at $\langle u_i|v_j\rangle$ in (5.81) is

$$\sum_{\hat{Q}\in S_N} \delta_{jQ_i}(-1)^q\langle u_1|v_{Q_1}\rangle\ldots\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots\langle u_n|v_{Q_n}\rangle = D(i|j)\ . \tag{5.83}$$

b) Now, let us choose two subscripts i and j ($1 \leq i < j \leq N$) and separate out the integrations over the coordinates of the ith and jth electrons. At first we perform analogous manipulations to those in the previous case, but now we introduce additional summations and Kronecker deltas for two subscripts:

$$\begin{aligned}\langle U|V\rangle &= \sum_{\hat{Q}\in S_N} (-1)^q \langle u_i|v_{Q_i}\rangle\langle u_j|v_{Q_j}\rangle\langle u_1|v_{Q_1}\rangle\ldots \\ &\qquad \times\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots \\ &\qquad \times\langle u_{j-1}|v_{Q_{j-1}}\rangle\langle u_{j+1}|v_{Q_{j+1}}\rangle\ldots\langle u_N|v_{Q_N}\rangle\end{aligned}$$

$$= \sum_{k,l=1}^{N} \langle u_i|v_k\rangle\langle u_j|v_l\rangle \sum_{\hat{Q}\in S_N} \delta_{kQ_i}\delta_{lQ_j}(-1)^q\langle u_1|v_{Q_1}\rangle\ldots \tag{5.84}$$
$$\times\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots$$
$$\times\langle u_{j-1}|v_{Q_{j-1}}\rangle\langle u_{j+1}|v_{Q_{j+1}}\rangle\ldots\langle u_N|v_{Q_N}\rangle\ .$$

On the right-hand side of (5.84), the summation indices k and l run independently of each other. (Only terms with $k \neq l$ give a contribution, because for every permutation $\hat{Q}$ one has $Q_i \neq Q_j$; so $\delta_{kQ_i}\delta_{lQ_j} = 0$ if $k = l$.) For a given pair of numbers k and l there are two terms in the sum that differ by k and l being interchanged. This interchange means that the respective permutations $\hat{Q}$ differ by one transposition, therefore their parities are opposite: $(-1)^q$ for one of the permutations and $-(-1)^q$ for the other. Otherwise the coefficients of these terms are the same. Thus we get

$$\begin{aligned}\langle U|V\rangle &= \sum_{k<l}(\langle u_i|v_k\rangle\langle u_j|v_l\rangle - \langle u_i|v_l\rangle\langle u_j|v_k\rangle)\\ &\times \sum_{\hat{Q}\in S_N} \delta_{kQ_i}\delta_{lQ_j}(-1)^q\langle u_1|v_{Q_1}\rangle\ldots\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots\\ &\times\langle u_{j-1}|v_{Q_{j-1}}\rangle\langle u_{j+1}|v_{Q_{j+1}}\rangle\ldots\langle u_N|v_{Q_N}\rangle\ .\end{aligned} \tag{5.85}$$

The term in parentheses is the second-order determinant

$$\begin{vmatrix} \langle u_i|v_k\rangle & \langle u_i|v_l\rangle \\ \langle u_j|v_k\rangle & \langle u_j|v_l\rangle \end{vmatrix}$$

(second-order minor of the determinant D) and the overlap integral $\langle U|V\rangle$ is written as the sum of such second-order determinants and some coefficients originating from the integration over the coordinates of electrons other than the ith and jth. Equation (5.85) should be compared with the following generalized Laplace theorem. For any pair of rows $i < j$, any determinant can be expanded as a sum of all its second-order minors involving rows i and j and all the possible columns $k < l$, multiplied by their algebraic complements, i.e., minors of order $N - 2$, obtained by deleting rows i, j and columns k, l and multiplied with the factor $(-1)^{i+j+k+l}$, $D(ij|kl) = (-1)^{i+j+k+l}M(ij|kl)$:

$$\begin{aligned}\langle U|V\rangle &= D = \mathrm{Det}\{\langle u_i|v_j\rangle\} \\ &= \sum_{k<l}(\langle u_i|v_k\rangle\langle u_j|v_l\rangle - \langle u_i|v_l\rangle\langle u_j|v_k\rangle)D(ij|kl)\ .\end{aligned} \tag{5.86}$$

(Laplace theorem can be generalized to minors of any order and their corresponding algebraic complements.)

Comparison with (5.86) shows that the coefficient at $(\langle u_i|v_k\rangle\langle u_j|v_l\rangle - \langle u_i|v_l\rangle\langle u_j|v_k\rangle)$ in (5.85) is

$$\sum_{\hat{Q}\in S_N} \delta_{kQ_i}\delta_{lQ_j}(-1)^q\langle u_1|v_{Q_1}\rangle\ldots\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots$$
$$\times\langle u_{j-1}|v_{Q_{j-1}}\rangle\langle u_{j+1}|v_{Q_{j+1}}\rangle\ldots\langle u_N|v_{Q_N}\rangle = D(ij|kl)\ . \qquad (5.87)$$

6.1.3. Special cases for orthonormalized orbitals (Slater rules)

In all cases of calculating matrix elements it is useful to bring the orbitals of the two determinants into maximal correspondence by changing their order if necessary—and introducing a factor ± 1. Here and further on we shall consider special cases assuming that these interchanges of orbitals have already been completed.

1. Let us consider the wave function $\Psi = \hat{\mathcal{A}}[\varphi_1(1)\varphi_2(2)\ldots\varphi_N(N)]$ and determine the normalization integral $\langle\Psi|\Psi\rangle$. In this case we have to apply (5.80) for $U = V = \Psi$ and $u_i = v_i = \varphi_i$. Then $S_{ij}^{uv} = S_{ij} = \langle\varphi_i|\varphi_j\rangle$ and

$$\langle\Psi|\Psi\rangle = \mathrm{Det}\{S_{ij}\} = \mathrm{Det}(\mathbf{S})\ . \qquad (5.88)$$

If the orbitals φ_i are orthonormalized, then $S_{ij} = \langle\varphi_i|\varphi_j\rangle = \delta_{ij}$, $\mathbf{S} = \mathbf{1}$, and the determinant in the right-hand side of (5.88) equals 1: *a Slater determinant wave function built up of orthonormalized spin orbitals is normalized to unity.* (This normalization condition motivated the selection of the factor $\frac{1}{\sqrt{N!}}$ in definitions (5.15) and (5.24) of the Slater determinant and of the antisymmetrizer, respectively.)

2. If either of the determinant wave functions U and/or V contains one or more spin orbitals that are orthogonal to all the spin orbitals of the other wave function, then $\mathbf{S}^{uv}$ will have one or more rows and/or columns, all elements of which are zero. Then $\mathrm{Det}(\mathbf{S}^{uv}) = 0$ according to the known properties of determinants; accordingly, the many-electron determinant wave functions U and V are orthogonal, too.

Therefore, if one uses an orthonormalized set of spin orbitals, then only those determinant wave functions have nonzero overlap integrals that contain the same spin orbitals. This overlap integral $\langle U|V\rangle$ equals 1 if the order of the orbitals is the same in the two determinants U and V, respectively, or if the orders of orbitals are related to each other by a permutation of even parity (i.e., if the order of the orbitals becomes the same after an even number of transpositions); the overlap $\langle U|V\rangle$ equals -1 if the order of orbitals in the two determinant is connected by a permutation of odd parity.

Illustration.

Equation (5.80) can also be used to demonstrate the invariance of determinant wave functions with respect of unitary transformations of the spin orbitals.

Let Ψ be a Slater determinant built of the orthonormalized spin orbitals φ_i ($\langle\varphi_i|\varphi_j\rangle = \delta_{ij}$):

$$\Psi = \hat{\mathcal{A}}[\varphi_1(1)\varphi_2(2)\ldots\varphi_N(N)] \tag{5.89}$$

and let

$$\Psi' = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)] \tag{5.90}$$

be a Slater determinant built up of the spin orbitals ψ_j obtained as a unitary transformation of the orbitals φ_i:

$$\psi_j = \sum_{k=1}^{N} U_{kj}\varphi_k \ . \tag{5.91}$$

As $\mathbf{U}$ is unitary ($\mathbf{U}^{-1} = \mathbf{U}^{\dagger}$), the spin orbitals ψ_j are also orthonormalized. Because Ψ and Ψ' both are built up of orthonormalized spin orbitals, both are normalized to 1:

$$\langle\Psi|\Psi\rangle = \langle\Psi'|\Psi'\rangle = 1 \ . \tag{5.92}$$

Let us now compute the overlap $\langle\Psi|\Psi'\rangle$. According to (5.80)

$$\begin{aligned}\langle\Psi|\Psi'\rangle &= \mathrm{Det}\{\langle\varphi_i|\psi_j\rangle\} = \mathrm{Det}\{\langle\varphi_i|\sum_{k=1}^{N} U_{kj}\varphi_k\rangle\} \\ &= \mathrm{Det}\{\sum_k U_{kj}\underbrace{\langle\varphi_i|\varphi_k\rangle}_{\delta_{ik}}\} = \mathrm{Det}\{U_{ij}\} \ .\end{aligned} \tag{5.93}$$

The determinant of any unitary matrix has the absolute value 1, therefore $|\langle\Psi|\Psi'\rangle| = 1$. If the overlap of two normalized wave functions has the absolute value 1, then these wave functions are identical—disregarding a possible, physically unimportant, phase factor.

6.2. One-electron operators

6.2.1. General case

Let $\hat{H}_1$ be the sum of the one-electron operators $\hat{h}(i)$, symmetric with respect of the individual electrons i:

$$\hat{H}_1 = \sum_{i=1}^{N} \hat{h}(i) \ . \tag{5.94}$$

Each $\hat{h}(i)$ acts on the functions depending on the coordinates of the ith electron. For instance, $\hat{H}_1$ may be the one-electron part of the Born–Oppenheimer Hamiltonian; in that case (assuming that atomic units—Appendix VI—are used) $\hat{h}(i) = -\frac{1}{2}\Delta_i - \sum_{\alpha}\frac{Z_\alpha}{r_{\alpha i}}$.

We have to determine the matrix element $\langle U|\hat{H}_1|V\rangle$, where U and V are again the Slater determinants built up of the spin orbitals u_i, v_j, defined in (5.74) and (5.75). The derivation is similar to the approach used for calculating the overlap integral $\langle U|V\rangle$. We again move the antisymmetrizer from the "bra" to the "ket," then utilize the fact that operator $\hat{H}_1$, being symmetric with respect to the electrons commutes with the antisymmetrizer $\hat{\mathcal{A}}$ (Section 4.2) and again make use of the relationships $\hat{\mathcal{A}}^\dagger = \hat{\mathcal{A}}$; $\hat{\mathcal{A}}^2 = \sqrt{N!}\hat{\mathcal{A}}$:

$$\begin{aligned}\langle U|\hat{H}_1|V\rangle &= \langle \hat{\mathcal{A}}u_1(1)u_2(2)\dots u_N(N)|\hat{H}_1|\hat{\mathcal{A}}[v_1(1)v_2(2)\dots v_N(N)]\rangle \\ &= \sqrt{N!}\langle u_1(1)u_2(2)\dots u_N(N)|\hat{H}_1|\hat{\mathcal{A}}[v_1(1)v_2(2)\dots v_N(N)]\rangle\ .\end{aligned} \tag{5.95}$$

We shall again use that form of the antisymmetrizer, in which it acts on the spin orbital indices, and get

$$\begin{aligned}\langle U|\hat{H}_1|V\rangle &= \sqrt{N!}\langle u_1(1)u_2(2)\dots u_N(N)|\sum_{i=1}^{N} h(i)|\frac{1}{\sqrt{N!}}\sum_{\hat{Q}\in S_N}(-1)^q v_{Q_1}(1) \\ &\qquad \times v_{Q_2}(2)\dots v_{Q_N}(N)\rangle \\ &= \sum_{i=1}^{N}\sum_{\hat{Q}\in S_N}(-1)^q\langle u_1(1)u_2(2)\dots u_i(i)\dots u_N(N)|\hat{h}(i)|v_{Q_1}(1) \\ &\qquad \times v_{Q_2}(2)\dots v_{Q_i}(i)\dots v_{Q_N}(N)\rangle\ .\end{aligned} \tag{5.96}$$

Each integral on the right-hand side of (5.96) again factorizes into a product of integrals; all of them, except that for the ith electron, is the same overlap integral between the spin orbitals as we have seen in the case of calculating overlap $\langle U|V\rangle$. We move the integral $\langle u_i(i)|\hat{h}(i)|V_{Q_i}(i)\rangle$ to the first place (there is no need to indicate the variables of integration any more):

$$\begin{aligned}\langle U|\hat{H}_1|V\rangle &= \sum_{i=1}^{N}\sum_{\hat{Q}\in S_N}(-1)^q\langle u_i|\hat{h}|v_{Q_i}\rangle\langle u_1|v_{Q_1}\rangle\langle u_2|v_{Q_2}\rangle\dots \\ &\qquad \times\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\dots\langle u_N|v_{Q_N}\rangle\ .\end{aligned} \tag{5.97}$$

Similarly to the case of (5.81), we introduce the equality $\sum_{j=1}^{N}\langle u_i|\hat{h}|v_j\rangle\delta_{jQ_i} = \langle u_i|\hat{h}|v_{Q_i}\rangle$, and get

$$\begin{aligned}\langle U|\hat{H}_1|V\rangle &= \sum_{i=1}^{N}\sum_{\hat{Q}\in S_N}(-1)^q\sum_{j=1}^{N}\langle u_i|\hat{h}|v_j\rangle\delta_{jQ_i}\langle u_1|v_{Q_1}\rangle\dots \\ &\qquad \times\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\dots\langle u_N|v_{Q_N}\rangle\end{aligned}$$

$$= \sum_{i,j=1}^{N} \langle u_i|\hat{h}|v_j\rangle \sum_{\hat{Q}\in S_N} (-1)^q \delta_{jQ_i} \langle u_1|v_{Q_1}\rangle \ldots \tag{5.98}$$
$$\times \langle u_{i-1}|v_{Q_{i-1}}\rangle \langle u_{i+1}|v_{Q_{i+1}}\rangle \ldots \langle u_N|v_{Q_N}\rangle \ .$$

Comparison with (5.83) shows that the coefficient at the integral $\langle u_i|\hat{h}|v_j\rangle$ is the algebraic complement $D(i|j)$, i.e., the minor $M(i|j)$ of the overlap matrix $\mathbf{S}^{uv}$, multiplied with the factor $(-1)^{i+j}$. Accordingly, we get the result

$$\langle U|\hat{H}_1|V\rangle = \sum_{i,j=1}^{N} \langle u_i|\hat{h}|v_j\rangle D(i|j) \ . \tag{5.99}$$

6.2.2. Special cases for orthonormalized orbitals (Slater rules)

1. Expectation value

Let $\Psi = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)]$ be a Slater determinant wave function built up of the orthonormalized spin orbitals $\psi_i = \psi_i(\vec{r},\sigma) = \varphi_i(\vec{r})\gamma_i(\sigma)$, where $\varphi_i(\vec{r})$ and $\gamma_i(\sigma)$ denote the spatial and spin parts of ψ_i, respectively ($\gamma_i = \alpha$ or β). Then we have to use (5.99) with $U = V = \Psi$; $u_i = v_i = \psi_i$. The mutual overlap matrix $\mathbf{S}^{uv} = \mathbf{S} = \mathbf{1}$; therefore $D(i|j) = \delta_{ij}$—only the "principal minors" of a unit matrix differ from 0, and they are equal to 1.

We have, therefore

$$\begin{aligned}\langle\Psi|\hat{H}_1|\Psi\rangle &= \langle\Psi|\sum_{i=1}^{N}\hat{h}(i)|\Psi\rangle = \sum_{i,j=1}^{N}\langle\psi_i|\hat{h}|\psi_j\rangle \underbrace{D(i|j)}_{\delta_{ij}} \\ &= \sum_{i=1}^{N}\langle\psi_i|\hat{h}|\psi_i\rangle = \sum_{i=1}^{N}\langle\varphi_i|\hat{h}|\varphi_i\rangle \ ,\end{aligned} \tag{5.100}$$

where in the last term we have carried out the summation over the spins, so the remaining integral is over the spatial variables only. Therefore the result is

$$\langle\Psi|\sum_{i=1}^{N}\hat{h}(i)|\Psi\rangle = \sum_{i=1}^{N}\langle\varphi_i|\hat{h}|\varphi_i\rangle \ . \tag{5.101}$$

We note that the wave function Ψ built up of orthonormalized spin orbitals is normalized to 1, therefore the expectation value of $\hat{H}_1$ for Ψ equals the matrix element $\langle\Psi|\hat{H}_1|\Psi\rangle$ given in (5.101).

Owing to the orthonormality of the spin functions α and β, the requirement that the spin orbitals ψ_i are orthonormalized can be formulated as $\langle\psi_i|\psi_j\rangle = \langle\varphi_i|\varphi_j\rangle\delta_{\gamma_i\gamma_j} = \delta_{ij}$. Therefore, it is sufficient to require that the spatial parts of the spin orbitals having *the same spin* be orthogonal, but there

are no restrictions for the spatial parts filled with different spins—they are often taken to be identical. In the case of using such *doubly filled orbitals* one has $\psi_{2i-1} = \varphi_i(\vec{r})\alpha(\sigma)$, $\psi_{2i} = \varphi_i(\vec{r})\beta(\sigma)$, and (5.101) reduces to

$$\langle \Psi_R | \sum_{i=1}^{N} \hat{h}(i) | \Psi_R \rangle = 2 \sum_{i=1}^{N/2} \langle \varphi_i | \hat{h}(i) | \varphi_i \rangle \ , \tag{5.102}$$

where subscript "R" refers to the use of doubly filled orbitals, as it is usual in the "restricted" Hartree–Fock method. In (5.102) the sum is over the different *spatial* orbitals only.

2. *Matrix element between determinants differing in one spin orbital.*

Let us consider

$$V = \Psi = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)] \tag{5.103}$$

—the same as in the previous case—and

$$\begin{aligned} U &= \Psi_1(\psi_k \to \psi_k') \qquad (5.104) \\ &= \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_{k-1}(k-1)\psi_k'(k)\psi_{k+1}(k+1)\ldots\psi_N(N)] \end{aligned}$$

—a determinant that differs from Ψ by replacing spin orbital ψ_k by the spin orbital ψ_k' that is orthogonal to all the spin orbitals ψ_i, $i = 1, 2, \ldots, N$. (Of course, it is meaningful to consider only spin orbitals ψ_k' having the same spin as ψ_k, otherwise U and V would correspond to different overall spin projections S_z and any matrix elements of spin-free operators would vanish between them. This means that the spatial part of ψ_k' must be orthogonal to the spatial parts of all spin orbitals of the same spin: $\langle \varphi_k' | \varphi_i \rangle = 0$ if $\gamma_k = \gamma_i$.)

In this case the the mutual overlap matrix $\mathbf{S}^{uv}$ differs from a unit matrix in one place: the kth diagonal element is 0, because $\langle \psi_k' | \psi_k \rangle = 0$. Therefore, all the minors $M(i|j)$ of $\mathbf{S}^{uv}$ are zero, except the one that is obtained by deleting the k-th row and k-th column; this minor equals 1. Accordingly $D(i|j) = (-1)^{i+j} M(i|j) = \delta_{ik}\delta_{jk}$. Substituting this into the general formula in (5.99) we get, performing also summation over the spins,

$$\langle \Psi_1(\psi_k \to \psi_k') | \sum_{i=1}^{N} \hat{h}(i) | \Psi \rangle = \langle \psi_k' | \hat{h} | \psi_k \rangle = \langle \varphi_k' | \hat{h} | \varphi_k \rangle \ . \tag{5.105}$$

3. If orthonormalized spin orbitals are used, *the matrix element* $\langle U | \hat{H}_1 | V \rangle$ *vanishes if the determinants U and V differ in two or more spin orbitals:* all the minors $M(i|j)$ of $\mathbf{S}^{uv}$ will be zero because they contain one or more rows and columns all elements of which are zero.

6.3. Two-electron operators

6.3.1. General case

Let $\hat{H}_2$ be the sum of two-electron operators $\hat{g}(i,j) = \hat{g}(j,i)$, symmetric with respect to all the electrons:

$$\hat{H}_2 = \sum_{i<j}^{N} \hat{g}(i,j) \ . \tag{5.106}$$

Each $\hat{g}(i,j)$ acts on the functions of coordinates of electrons i and j. The most important operator of this type is the two-electron part of the Born–Oppenheimer Hamiltonian; in that case (assuming atomic units) $\hat{g}(i,j)= \frac{1}{r_{ij}}$.

We shall proceed similarly to the previous two cases. We again use the Hermiticity and idempotency properties of the antisymmetrizing operator $\hat{\mathcal{A}}$, as well as the fact that it commutes with the operator $\hat{H}_2$ which is symmetric with respect to the electrons:

$$\begin{aligned}
\langle U|\hat{H}_2|V\rangle &= \langle \hat{\mathcal{A}}[u_1(1)u_2(2)\dots u_N(N)]|\hat{H}_2|\hat{\mathcal{A}}[v_1(1)v_2(2)\dots v_N(N)]\rangle \\
&= \sqrt{N!}\langle [u_1(1)u_2(2)\dots u_N(N)]|\hat{H}_2|\hat{\mathcal{A}}[v_1(1)v_2(2)\dots v_N(N)]\rangle \\
&= \sqrt{N!}\langle u_1(1)u_2(2)\dots u_N(N)|\sum_{i<j}^{N}\hat{g}(i,j)|\frac{1}{\sqrt{N!}}\sum_{\hat{Q}\in S_N}(-1)^q v_{Q_1}(1) \\
&\qquad \times v_{Q_2}(2)\dots v_{Q_N}(N)\rangle \qquad (5.107) \\
&= \sum_{i<j}^{N}\sum_{\hat{Q}\in S_N}(-1)^q\langle u_1(1)\dots u_i(i)\dots u_j(j)\dots u_N(N)|\hat{g}(i,j)|v_{Q_1}(1)\dots \\
&\qquad \times v_{Q_i}(i)\dots v_{Q_j}(j)\dots v_{Q_N}(N)\rangle \ .
\end{aligned}$$

The integral on the right-hand side again factorizes into the product of overlap integrals that resulted from the integration over the coordinates of the individual electrons, except the integration over the coordinates of ith and jth electrons; moving again the respective integral $\langle u_i(i)u_j(j)|\hat{g}(i,j)|v_{Q_i}(i)v_{Q_j}(j)\rangle$ to the first place, we get

$$\begin{aligned}
\langle U|\hat{H}_2|V\rangle &= \sum_{i<j}^{N}\sum_{\hat{Q}\in S_N}(-1)^q\langle u_i(i)u_j(j)|\hat{g}(i,j)|v_{Q_i}(i)v_{Q_j}(j)\rangle \\
&\quad \times\langle u_1|v_{Q_1}\rangle\dots\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\dots \qquad (5.108) \\
&\quad \times\langle u_{j-1}|v_{Q_{j-1}}\rangle\langle u_{j+1}|v_{Q_{j+1}}\rangle\dots\langle u_N|v_{Q_N}\rangle \ .
\end{aligned}$$

Similarly to the derivation of (5.84), we again introduce two Kronecker summations, giving

$$\begin{aligned}\langle U|\hat{H}_2|V\rangle &= \sum_{i<j}^{N}\sum_{k,l=1}^{N}\sum_{\hat{Q}\in S_N}\delta_{kQ_i}\delta_{lQ_j}(-1)^q\langle u_i(i)u_j(j)|\hat{g}(i,j)|v_k(i)v_l(j)\rangle \\ &\quad\times\langle u_1|v_{Q_1}\rangle\ldots\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots \\ &\quad\times\langle u_{j-1}|v_{Q_{j-1}}\rangle\langle u_{j+1}|v_{Q_{j+1}}\rangle\ldots\langle u_N|v_{Q_N}\rangle\,.\end{aligned} \tag{5.109}$$

In the following equations we shall use "1" and "2," instead of i and j, as variables of integration. (Here "1" and "2" stand for the combinations $\{\vec{r}_1,\sigma_1\}$ and $\{\vec{r}_2,\sigma_2\}$ of the spatial and spin coordinates of the two electrons.) Similarly to the considerations leading from (5.84) to (5.85), we may again combine the two cases in which k and l are interchanged but all the other subscripts Q_p are the same; these again correspond to permutations of opposite parity, but give otherwise the same coefficient. Accordingly:

$$\begin{aligned}\langle U|\hat{H}_2|V\rangle &= \sum_{i<j}^{N}\sum_{k<l}^{N}[\langle u_i(1)u_j(2)|\hat{g}(1,2)|v_k(1)v_l(2)\rangle \\ &\qquad -\langle u_i(1)u_j(2)|\hat{g}(1,2)|v_l(1)v_k(2)\rangle] \\ &\quad\times\sum_{\hat{Q}\in S_N}\delta_{kQ_i}\delta_{lQ_j}(-1)^q\langle u_1|v_{Q_1}\rangle\ldots\langle u_{i-1}|v_{Q_{i-1}}\rangle\langle u_{i+1}|v_{Q_{i+1}}\rangle\ldots \\ &\qquad\times\langle u_{j-1}|v_{Q_{j-1}}\rangle\langle u_{j+1}|v_{Q_{j+1}}\rangle\ldots\langle u_N|v_{Q_N}\rangle\,.\end{aligned} \tag{5.110}$$

Comparison with (5.87) shows that the coefficient at the parentheses containing the difference of two two-electron integrals is the algebraic complement $D(ij|kl)$, i.e., the minor $M(ij|kl)$ of matrix $\mathbf{S}^{uv}$ of the mutual overlap, multiplied by the factor $(-1)^{i+j+k+l}$. Accordingly, we get the result

$$\begin{aligned}\langle U|\hat{H}_2|V\rangle &= \sum_{\substack{i<j\\k<l}}^{N}[\langle u_i(1)u_j(2)|\hat{g}(1,2)|v_k(1)v_l(2)\rangle \\ &\qquad -\langle u_i(1)u_j(2)|\hat{g}(1,2)|v_l(1)v_k(2)\rangle]\,D(ij|kl)\,.\end{aligned} \tag{5.111}$$

Another, equivalent, form of this formula can be obtained if one utilizes the symmetry of the integrals with respect to the interchange of the integration variables

$$\langle u_i(1)u_j(2)|\hat{g}(1,2)|v_k(1)v_l(2)\rangle = \langle u_j(1)u_i(2)|\hat{g}(1,2)|v_l(1)v_k(2)\rangle \tag{5.112}$$

as well as generalizes the definition of the quantities $D(ij|kl)$ for the case when $i>j$ and/or $k>l$, by requiring them to be antisymmetric in both sets of indices

i, j and k, l. Then, as is trivial to check, (5.111) can also be rewritten in a more compact form:

$$\langle U|\hat{H}_2|V\rangle = \frac{1}{2}\sum_{i,j,k,l=1}^{N} \langle u_i(1)u_j(2)|\hat{g}(1,2)|v_k(1)v_l(2)\rangle D(ij|kl) \ . \quad (5.113)$$

This is the form originally derived by Löwdin. At the same time, the form (5.111) is more convenient in most practical applications.

6.3.2. Special cases for orthonormalized orbitals (Slater rules)

1. Expectation value

This is the same case as has been discussed previously when the expectation value of the one-electron operator $\hat{H}_1$ has been considered: $U = V = \Psi$; $u_i = v_i = \psi_i = \varphi_i\gamma_i$. Similarly to that case, only the principal minors of the matrix $\mathbf{S}^{uv} = \mathbf{1}$ differ from zero, and they are equal to 1. This means that if we delete row i then we have also to delete column i and vice versa, i.e., $D(ij|kl) = \delta_{ik}\delta_{jl}$ ($i < j$, $k < l$ are assumed). Equation (5.111) reduces to

$$\begin{aligned}\langle\Psi|\hat{H}_2|\Psi\rangle &= \sum_{i<j}^{N}[\langle\psi_i(1)\psi_j(2)|\hat{g}(1,2)|\psi_i(1)\psi_j(2)\rangle \\ &\quad - \langle\psi_i(1)\psi_j(2)|\hat{g}(1,2)|\psi_j(1)\psi_i(2)\rangle] \ .\end{aligned} \quad (5.114)$$

It is easy to perform the summation over the spins explicitly. In the first integral it simply gives a factor of 1, as for both electrons one has the same spin orbital in the "bra" and in the "ket." The summation over the spins in the second integral results in a factor $\delta_{\gamma_i\gamma_j}$, i.e., it gives 1 if the spin orbitals ψ_i and ψ_j are of the same spin and 0 otherwise. Thus we get in terms of the spatial orbitals

$$\begin{aligned}\langle\Psi|\hat{H}_2|\Psi\rangle &= \sum_{i<j}^{N}[\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_i(1)\varphi_j(2)\rangle \\ &\quad -\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_j(1)\varphi_i(2)\rangle\delta_{\gamma_i\gamma_j}] \ .\end{aligned} \quad (5.115)$$

Sometimes it is worthwhile to rewrite (5.115) symmetrically in the indices i and j as

$$\begin{aligned}\langle\Psi|\hat{H}_2|\Psi\rangle &= \frac{1}{2}\sum_{i,j=1}^{N}[\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_i(1)\varphi_j(2)\rangle \\ &\quad -\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_j(1)\varphi_i(2)\rangle\delta_{\gamma_i\gamma_j}] \ .\end{aligned} \quad (5.116)$$

We shall note that the term with $i = j$ cancels automatically.

In the case of *doubly filled* orbitals (subscript R), (5.115) and (5.116) become, in terms of sums over the different *spatial* orbitals,

$$\begin{aligned}\langle\Psi_R|\hat{H}_2|\Psi_R\rangle &= \sum_{i<j}^{N/2}[4\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_i(1)\varphi_j(2)\rangle \\ &\qquad -2\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_j(1)\varphi_i(2)\rangle] \\ &\quad + \sum_{i=1}^{N/2}\langle\varphi_i(1)\varphi_i(2)|\hat{g}(1,2)|\varphi_i(1)\varphi_i(2)\rangle \qquad (5.117) \\ &= \sum_{i,j=1}^{N/2}[2\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_i(1)\varphi_j(2)\rangle \\ &\qquad -\langle\varphi_i(1)\varphi_j(2)|\hat{g}(1,2)|\varphi_j(1)\varphi_i(2)\rangle]\,.\end{aligned}$$

2. *Matrix elements between determinants differing in one spin orbital*

Again we shall refer to the notations and discussion for the similar case considered in connection of the one-electron operator $\hat{H}_1$. When considering (5.111) we see that one of the rows and one of the columns of matrix $\mathbf{S}^{uv}$, deleted when forming the algebraic complement $D(ij|kl)$, should coincide with that pertinent to the orbital that is different in the two determinants—otherwise there remains a row and/or column containing only zeros. (We shall assume that orbital to be the kth.) For the same reason, the second row and the second column deleted when forming the minor must have the same index. Thus we get the nonzero algebraic complements $D(ik|ik)$ for $i < k$ and $D(ki|ki)$ for $i > k$; both are obviously equal to one. Owing to the symmetry of the two-electron integrals with respect to interchanging the variables of integration, there is no need to distinguish further these cases. One gets

$$\begin{aligned}\langle\Psi_1(\psi_k\to\psi_k')|\hat{H}_2|\Psi\rangle &= \sum_{\substack{i\\(i\neq k)}}^{N}[\langle\psi_k'(1)\psi_i(2)|\hat{g}(1,2)|\psi_k(1)\psi_i(2)\rangle \\ &\qquad -\langle\psi_k'(1)\psi_i(2)|\hat{g}(1,2)|\psi_i(1)\psi_k(2)\rangle]\,. \qquad (5.118)\end{aligned}$$

The restriction $i \neq k$ on the summation can obviously by omitted because the respective term cancels automatically. One can rewrite (5.118) to spatial orbitals as

$$\begin{aligned}\langle\Psi_1(\psi_k\to\psi_k')|\hat{H}_2|\Psi\rangle &= \sum_{\substack{i=1\\(i\neq k)}}^{N}[\langle\varphi_k'(1)\varphi_i(2)|\hat{g}(1,2)|\varphi_k(1)\varphi_i(2)\rangle \\ &\qquad -\langle\varphi_k'(1)\varphi_i(2)|\hat{g}(1,2)|\varphi_i(1)\varphi_k(2)\rangle\delta_{\gamma_i\gamma_k}]\,. \qquad (5.119)\end{aligned}$$

In the case of using doubly filled orbitals (5.119) can be written as

$$\langle \Psi_{R1}(\psi_k \to \psi_k')|\hat{H}_2|\Psi_R\rangle = \sum_{i=1}^{N/2}[2\langle \varphi_k'(1)\varphi_i(2)|\hat{g}(1,2)|\varphi_k(1)\varphi_i(2)\rangle - \langle \varphi_k'(1)\varphi_i(2)|\hat{g}(1,2)|\varphi_i(1)\varphi_k(2)\rangle] , \quad (5.120)$$

where the sum is over the different spatial orbitals and for the sake of simplicity we have utilized the possibility of omitting the restriction $i \neq k$. (Otherwise one would have to treat separately the term $\langle \varphi_k'(1)\varphi_k(2)|\hat{g}(1,2)|\varphi_k(1)\varphi_k(2)\rangle$ originating from the case when ψ_k and ψ_i in (5.118) correspond to the same spatial orbital filled with different spins.)

3. Matrix elements between determinants differing in two spin orbitals

In this case $U = \Psi_2(\psi_k \to \psi_k', \psi_l \to \psi_l')$; the only nonzero algebraic complement is $D(kl|kl)$; it is equal to 1. Thus we get, performing also the summations over the spins,

$$\langle \Psi_2(\psi_k \to \psi_k', \psi_l \to \psi_l')|\hat{H}_2|\Psi\rangle \quad (5.121)$$

$$= \langle \psi_k'(1)\psi_l'(2)|\hat{g}(1,2)|\psi_k(1)\psi_l(2)\rangle - \langle \psi_k'(1)\psi_l'(2)|\hat{g}(1,2)|\psi_l(1)\psi_k(2)\rangle$$

$$= \langle \varphi_k'(1)\varphi_l'(2)|\hat{g}(1,2)|\varphi_k(1)\varphi_l(2)\rangle - \langle \varphi_k'(1)\varphi_l'(2)|\hat{g}(1,2)|\varphi_l(1)\varphi_k(2)\rangle\delta_{\gamma_k\gamma_l} .$$

4. If orthonormalized spin orbitals are used, then the *matrix element* $\langle U|\hat{H}_2|V\rangle$ *vanishes if the determinants U and V differ in three or more spin orbitals*: there are no nonzero minors $M(ij|kl)$.

7. Löwdin's pairing theorem and its generalization

Löwdin's pairing theorem utilizes the invariance of the determinant wave functions under unitary transformations of the orbitals (Section 5). It is applicable in those cases in which one can divide the spin orbitals into two distinct sets (or several pairs of distinct sets) based on some criterion. Originally the two sets were formed by the spin orbitals of spin α and β, respectively, and the pairing theorem was utilized in the methods using spin projection (see Section 4.3 of Chapter 6). In these cases no doubly filled orbitals were used and the aim of using the pairing theorem was to ensure the maximum degree of orthogonality between the *spatial parts* of the spin orbitals filled with spins α and β, respectively, which can be obtained without changing the many-electron determinant wave function. Another application of this theorem is in the theory of intermolecular interactions. There the use of the pairing theorem permitted one to derive and analyze an explicit formula for the energy of two weakly interacting molecules in the framework of the following simplest model: each molecule is described by using doubly filled orbitals and one neglects, as a first approximation, the change of the molecular orbitals during the complex formation—but

accounts explicitly for the more important effects caused by the intermolecular overlap. In that case the two sets are the (doubly filled) molecular orbitals of the two weakly interacting molecules, and the use of the pairing theorem ensures—without changing the many electron wave function—that only a few intermolecular overlap integrals will differ from zero.

7.1. Löwdin's pairing theorem

Let us consider two sets of orbitals $\{|a_i\rangle; i = 1, 2, \ldots n_a\}$ and $\{|b_j\rangle; j = 1, 2, \ldots n_b\}$ that are assumed to be orthonormalized:

$$\langle a_i|a_j\rangle = \langle b_i|b_j\rangle = \delta_{ij} \ . \tag{5.122}$$

(The "bra-ket" formalism is used in this section.) The pairing theorem states that it is always possible to find *unitary* transformations mixing the orbitals only *within* the individual sets $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$, respectively, such as to achieve the two sets of the transformed orbitals $\{|a_i'\rangle\}$ and $\{|b_j'\rangle\}$ to be also mutually orthogonal, except pairs of "corresponding orbitals":

$$\langle a_i'|b_j'\rangle = \lambda_i \delta_{ij} \ . \tag{5.123}$$

Note that the unitary transformations conserve the orthonormalization property (5.122), as well.

Proof[1]

Let us assume $n_a \geq n_b$ (otherwise the two sets should be interchanged) and form the operator $\hat{P}^b$ of projection on the subspace spanned by orbitals $\{|b_j\rangle\}$

$$\hat{P}^b = \sum_{l=1}^{n_b} |b_l\rangle\langle b_l| \ . \tag{5.124}$$

Then we can form the matrix $\mathbf{Q}^b$ with the elements

$$Q^b_{ij} = \langle a_i|\hat{P}^b|a_j\rangle \ . \tag{5.125}$$

Matrix $\mathbf{Q}^b$ is Hermitian, so it can be diagonalized by a unitary matrix $\mathbf{U}$:

$$\mathbf{U}^\dagger \mathbf{Q}^b \mathbf{U} = diag\{\eta_i\} \ . \tag{5.126}$$

We form the transformed set of orbitals $\{|a_j'\rangle\}$ as

$$|a_j'\rangle = \sum_{l=1}^{n_a} U_{lj}|a_l\rangle \ , \tag{5.127}$$

[1]The following proof is reproduced from my paper "Simple Proof of the Pairing Theorem," I. Mayer, Int. J. Quantum Chem. **63**, 31–33 (1997); Copyright 1997 Wiley Periodicals.

and observe that—owing to (5.126)—they satisfy

$$\langle a_i'|\hat{P}^b|a_j'\rangle = \eta_i\delta_{ij} \ . \tag{5.128}$$

Operator $\hat{P}^b$ is idempotent and Hermitian, so Eq. (5.128) can be trivially transformed as

$$\langle a_i'|\hat{P}^b\hat{P}^b|a_j'\rangle = \langle \hat{P}^b a_i'|\hat{P}^b a_j'\rangle = \eta_i\delta_{ij} \ . \tag{5.129}$$

It follows from the last equality that $\eta_i \geq 0$, so one can define it as

$$\eta_i = \lambda_i^2 \ , \tag{5.130}$$

where $\lambda_i \geq 0$, as well.

Functions $|\hat{P}^b a_i'\rangle \equiv \hat{P}^b|a_i'\rangle$ for which $\lambda_i \neq 0$ entirely lie in the subspace spanned by the set of original orbitals $\{|b_j\rangle\}$, whereas those with $\lambda_i = 0$ vanish according to (5.129). The functions with $\lambda_i \neq 0$ are also orthogonal (but not yet normalized), so they are linearly independent. As it is impossible to form more than n_b linearly independent orbitals entirely lying in the subspace of orbitals $\{|b_j\rangle\}$, at least $n_a - n_b$ eigenvalues in (5.126) should be equal to zero. (If $n_a = n_b$, usually no zero eigenvalues appear—except for some very special symmetry reasons.) Normalizing the orbitals just discussed we get the transformed set of orbitals $\{|b_i'\rangle\}$ as

$$|b_i'\rangle = \frac{1}{\lambda_i}\hat{P}^b|a_i'\rangle \ . \tag{5.131}$$

It follows from (5.129) and (5.130) that these orbitals are indeed orthonormalized:

$$\langle b_i'|b_j'\rangle = \frac{1}{\lambda_i\lambda_j}\langle \hat{P}^b a_i'|\hat{P}^b a_j'\rangle = \frac{1}{\lambda_i\lambda_j}\lambda_i^2\delta_{ij} = \delta_{ij} \ . \tag{5.132}$$

Owing to (5.128) and (5.130), the two sets of orbitals $\{|a_i'\rangle\}$ and $\{|b_j'\rangle\}$ obtained really satisfy the pairing theorem in (5.123):

$$\langle a_i'|b_j'\rangle = \langle a_i'|\frac{1}{\lambda_j}\hat{P}^b a_j'\rangle = \frac{1}{\lambda_j}\langle a_i'|\hat{P}^b|a_j'\rangle = \frac{1}{\lambda_j}\lambda_i^2\delta_{ij} = \lambda_i\delta_{ij}, \tag{5.133}$$

which completes the proof for the case of all orbital pairs for which the λ_i values differ from zero.

The orbitals $\{|a_i'\rangle\}$ obtained by the unitary transformation span the whole subspace of the starting orbitals $\{|a_i\rangle\}$ and satisfy the conditions of the theorem. However, we cannot use (5.131) to obtain those transformed orbitals $|b_i'\rangle$ that correspond to zero λ_i values (if there are any), as λ_i occurs in the denominator. In such cases we have to complete the set $\{|b_j'\rangle\}$ with some number of orbitals satisfying the conditions of the theorem. It is easy to see, however,

that one may simply choose arbitrary orthonormal vectors in the orthogonal complement to the orbitals $\{|b_j'\rangle\}$ already determined with the use of (5.131), as defined with respect to the original subspace $\{|b_j\rangle\}$. They can be most easily determined as eigenvectors (with eigenvalues one) of the projection operators on this orthogonal complement $\hat{P}^b_{ort}$

$$\hat{P}^b_{ort} = \hat{P}^b - \sum_{l=1}^{n_b'} |b_l'\rangle\langle b_l'| \ ; \tag{5.134}$$

here n_b' is the number of nonzero λ_i values.

To complete the proof for this case as well, we have to show that the orbitals $\{|b_k'\rangle; k = n_b', n_b'+1, \ldots, n_b\}$ obtained in this manner are orthogonal to all the orbitals $\{|a_i'\rangle\}$. In light of the properties of the projection operators, the operator $\hat{P}^b_{ort}$ can be expressed as its spectral resolution:

$$\hat{P}^b_{ort} = \sum_{l=n_b'+1}^{n_b} |b_l'\rangle\langle b_l'| \ . \tag{5.135}$$

Forming the matrix element $\langle a_i'|\hat{P}^b_{ort}|a_i'\rangle$ we have

$$\langle a_i'|\hat{P}^b_{ort}|a_i'\rangle = \sum_{l=n_b'+1}^{n_b} \langle a_i'|b_l'\rangle\langle b_l'|a_i'\rangle = \sum_{l=n_b'+1}^{n_b} |\langle a_i'|b_l'\rangle|^2 \ . \tag{5.136}$$

At the same time, according to (5.134),

$$\langle a_i'|\hat{P}^b_{ort}|a_i'\rangle = \langle a_i'|\hat{P}^b|a_i'\rangle - \sum_{l=1}^{n_b'} \langle a_i'|b_l'\rangle\langle b_l'|a_i'\rangle = 0, \tag{5.137}$$

where the last equality follows from (5.128), (5.130), and (5.133). Substituting this result in (5.136) we get

$$\sum_{l=n_b'+1}^{n_b} |\langle a_i'|b_l'\rangle|^2 = 0 \ , \tag{5.138}$$

The sum of positive numbers can be zero only if each of these numbers is zero. Q. E. D.

In practice one uses orbitals expanded in a finite, usually nonorthogonal set of basis orbitals. Whereas forming the linear combinations prescribed by (5.127) does not represent a problem, special attention should be paid to the consequences of the overlap on the proper matrix representation of the projection operators (c.f. Appendix VII).

Assuming that the orbital coefficients form column vectors $\mathbf{a}^i$, $\mathbf{a}^{i\prime}$, $\mathbf{b}^i$, and $\mathbf{b}^{i\prime}$, the calculation of $\mathbf{b}^{i\prime}$ corresponding to (5.131) will have the form

$$\mathbf{b}^{i\prime} = \frac{1}{\lambda_i}\mathbf{P}^b\mathbf{S}\mathbf{a}^{i\prime} \ , \tag{5.139}$$

where $\mathbf{S}$ is the overlap matrix of the basis orbitals and

$$\mathbf{P}^b = \sum_{l=1}^{n_b} \mathbf{b}_l\mathbf{b}_l^\dagger \ . \tag{5.140}$$

If there are orbitals that have to be determined from the orthogonal complement discussed, then one has to solve the non-Hermitian eigenvalue problem

$$(\mathbf{P}^b - \mathbf{P}^{b\prime})\mathbf{S}\mathbf{b}_k' = \varepsilon_k\mathbf{b}_k' \ , \tag{5.141}$$

with matrix $\mathbf{P}^b$ defined in (5.140),

$$\mathbf{P}^{b\prime} = \sum_{l=1}^{n_b'} \mathbf{b}_l'\mathbf{b}_l'^\dagger \ , \tag{5.142}$$

and $n_b' < k \leq n_b$. By multiplying (5.141) with matrix $\mathbf{S} = \mathbf{S}^\dagger$ from the left, we arrive at the generalized Hermitian eigenvalue problem

$$\mathbf{S}(\mathbf{P}^b - \mathbf{P}^{b\prime})\mathbf{S}\mathbf{b}_k' = \varepsilon_k\mathbf{S}\mathbf{b}_k' \ . \tag{5.143}$$

Special numerical problems may arise if some λ_i values differ from zero but are very small. In this case iterations based on the two-by-two pairing of orbitals with small or zero λ_i values can be used.

7.2. Karadakov's extended pairing theorem

One usually calls the one-electron (spin) orbitals used to build up the given determinant wave function "occupied orbitals" ("orbitals occupied in the given determinant"), whereas the functions in the orthogonal complement to the "occupied subspace" (to the subspace spanned by the occupied orbitals) are called "virtual orbitals," and their subspace is the "virtual subspace." (The latter is, in principle, of infinite dimensions, but often we consider only that part of it that can be expanded in a given finite basis set.)

When considering the two sets of orbitals discussed in Löwdin's pairing theorem, one can define two virtual subspaces—one representing the orthogonal complement to subspace of the orbitals $\{|a_i\rangle\}$, the other representing that of the orbitals $\{|b_i\rangle\}$. In the following equations we shall use superscripts "o" and "v" to distinguish occupied and virtual sets of orbitals. Obviously, the sets of orbitals $\{|a_i^o\rangle\}$ and $\{|a_i^v\rangle\}$ together span the same—finite or infinite—linear

space of one-electron functions as do the sets $\{|b_i^o\rangle\}$ and $\{|b_i^v\rangle\}$ together. All four sets are assumed orthonormalized separately, and each virtual orbital is orthogonal to orbitals of the respective occupied set, in accord with the notion of the orthogonal complement:

$$\begin{aligned} \langle a_i^o|a_j^o\rangle &= \langle a_i^v|a_j^v\rangle = \delta_{ij} \\ \langle b_i^o|b_j^o\rangle &= \langle b_i^v|b_j^v\rangle = \delta_{ij} \\ \langle a_i^o|a_j^v\rangle &= \langle b_i^o|b_j^v\rangle = 0 \end{aligned} \tag{5.144}$$

—but not, in general, to the orbitals in the other occupied and virtual sets.

Karadakov had generalized Löwdin's pairing theorem to these two virtual subspaces. Of course, we can perform their pairing independently of that for the occupied orbitals by using the method described in the previous section. Karadakov's generalized pairing theorem, however, states more: one can *simultaneously* perform the pairing of each virtual subspace (say $\{|a_i^v\rangle\}$) with both the occupied and the virtual subspaces of another set ($\{|b_i^o\rangle\}$ and $\{|b_i^v\rangle\}$ in the given case).

The theorem can easily be proved, starting from the assumption that the pairing of the occupied subspaces has already been performed, as described in the previous section. Then we can very simply construct the basis sets in the two virtual subspaces, satisfying the conditions of the theorem.

The occupied orbitals $\{|a_i^o\rangle\}$ and $\{|b_i^o\rangle\}$ together span a subspace of dimension not exceeding $n_a + n_b$. (If all of these orbitals are linearly independent, then this dimension is n_a+n_b; if there are k doubly filled orbitals corresponding to $\lambda_i = 1$ values, then the dimension of this subspace is $n_a + n_b - k$.) All the orbitals orthogonal to this combined subspace form the common part of the two virtual subspaces. Selecting any set of orthonormalized vectors in this subspace, one gets orbitals satisfying the conditions of the theorem: Each of these orbitals has a non-zero overlap integral (of value 1) with only one orbital of another virtual subspace and is orthogonal to both occupied subspaces. There is no need to consider these orbitals in any further detail.

Now we shall form the projection of each occupied orbital $|a_i^o\rangle$ and $|b_j^o\rangle$ on the virtual subspace corresponding to the other set. These projections represent vectors in the respective virtual subspaces. The virtual subspaces $\{|a_i^v\rangle\}$ and $\{|b_i^v\rangle\}$ are the orthogonal complements to the occupied subspaces $\{|a_i^o\rangle\}$ and $\{|b_i^o\rangle\}$, respectively; the corresponding projectors are, therefore, $\hat{P}_a^{\text{virt}} = 1 - \sum_{l=1}^{n_a}|a_l^o\rangle\langle a_l^o|$ and $\hat{P}_b^{\text{virt}} = 1 - \sum_{l=1}^{n_b}|b_l^o\rangle\langle b_l^o|$. Thus we get the vectors

$$\begin{aligned} \hat{P}_b^{\text{virt}}|a_i^o\rangle &= |a_i^o\rangle - \sum_{l=1}^{n_b}|b_l^o\rangle\langle b_l^o|a_i^o\rangle \\ &= |a_i^o\rangle - \lambda_i^*|b_i^o\rangle\ ; \end{aligned}$$

$$\begin{aligned}\hat{P}_a^{\text{virt}}|b_j^o\rangle &= |b_j^o\rangle - \sum_{l=1}^{n_a}|a_l^o\rangle\langle a_l^o|b_j^o\rangle \\ &= |b_j^o\rangle - \lambda_j|a_j^o\rangle\ ,\end{aligned} \tag{5.145}$$

where we have taken into account the fact that the occupied orbitals are assumed paired: $\langle b_l^o|a_i^o\rangle = \lambda_i^*\delta_{li}$; $\langle a_l^o|b_j^o\rangle = \lambda_j\delta_{lj}$. We normalize the orbitals obtained and get

$$\begin{aligned}|b_i^v\rangle &= \frac{1}{\sqrt{1-|\lambda_i|^2}}(|a_i^o\rangle - \lambda_i^*|b_i^o\rangle)\ ; \\ |a_j^v\rangle &= \frac{1}{\sqrt{1-|\lambda_j|^2}}(|b_j^o\rangle - \lambda_j|a_j^o\rangle)\ .\end{aligned} \tag{5.146}$$

It is easy to see that the orbitals in (5.146) satisfy the conditions of the theorem. Thus $|b_i^v\rangle$ and $|a_j^v\rangle$ are by construction orthogonal to all the occupied orbitals $|b_i^o\rangle$ and $|a_i^o\rangle$, respectively. The overlap integral $\langle a_j^o|b_i^v\rangle$ is nonzero only for one orbital $|a_i^o\rangle$:

$$\begin{aligned}\langle a_j^o|b_i^v\rangle &= \frac{1}{\sqrt{1-|\lambda_i|^2}}(\langle a_j^o|a_i^o\rangle - \lambda_i^*\langle a_j^o|b_i^o\rangle) \\ &= \frac{1}{\sqrt{1-|\lambda_i|^2}}(\delta_{ji} - |\lambda_i|^2\delta_{ij}) = \sqrt{1-|\lambda_i|^2}\delta_{ji}\ .\end{aligned} \tag{5.147}$$

In the same manner we obtain

$$\langle b_j^o|a_i^v\rangle = \sqrt{1-|\lambda_i|^2}\ \delta_{ji}\ . \tag{5.148}$$

(If $n_a > n_b$, then there are $n_a - n_b$ orbitals $|b_i^v\rangle$ in which the term $\lambda_i^*|b_i^o\rangle$ is absent; these occupied orbitals $|a_i^o\rangle$ $(i > n_b)$ coincide with the first $n_a - n_b$ virtual orbitals $|b_{i-n_b}^v\rangle$.)

To complete the proof, we have to consider the overlap between the virtual orbitals formed according to (5.146):

$$\begin{aligned}\langle a_i^v|a_j^v\rangle &= \frac{1}{\sqrt{(1-|\lambda_i|^2)(1-|\lambda_j|^2)}}\langle b_i^o - \lambda_i a_i^o|b_j^o - \lambda_j a_j^o\rangle \\ &= \frac{1}{\sqrt{(1-|\lambda_i|^2)(1-|\lambda_j|^2)}}(\delta_{ij} - \lambda_i^*\lambda_i\delta_{ij} - \lambda_j\lambda_i^*\delta_{ij} + \lambda_i^*\lambda_j\delta_{ij}) = \delta_{ij}\ ,\end{aligned} \tag{5.149}$$

where we have utilized the fact that the occupied orbitals are orthonormalized and paired. Similarly one gets $\langle b_i^v|b_j^v\rangle = \delta_{ij}$, as well; whereas

$$\begin{aligned}
\langle a_i^v | b_j^v \rangle &= \frac{1}{\sqrt{(1-|\lambda_i|^2)(1-|\lambda_j|^2)}} \langle b_i^o - \lambda_i a_i^o | a_j^o - \lambda_j^* b_j^o \rangle \\
&= \frac{1}{\sqrt{(1-|\lambda_i|^2)(1-|\lambda_j|^2)}} (\lambda_i^* \delta_{ij} - \lambda_i^* \delta_{ij} - \lambda_j^* \delta_{ij} + \lambda_i^* \lambda_j^* \lambda_i \delta_{ij}) \\
&= -\lambda_i^* \delta_{ij} \ . \qquad (5.150)
\end{aligned}$$

Q.E.D.

8. Theorem about the existence of orbitals of special structure

In this section we shall discuss a theorem that originally emerged in connection with localized orbitals, but it can be applicable in studying most different problems related to single determinant wave functions.

In qualitative discussions of molecular structure one often uses concepts like the core and lone pair orbitals of individual atoms, diatomic bonding and antibonding orbitals, etc. The localized orbitals one can get by *a posteriori* analysis of single determinant wave functions obtained in the actual calculations usually support these notions to a great extent, but most of the localized orbitals have also (minor) components delocalized over larger parts of the molecule. Conventional localized orbitals are obtained by a unitary transformation of the "canonical" ones by requiring some "localization criterion" to be fulfilled. Accordingly, these localized orbitals are orthonormalized, similarly to the canonical orbitals.

We have seen in Section 5 that every determinant wave function can be presented (up to a physically irrelevant constant factor) by using orthonormalized orbitals, but this is not necessary: one may use any set of linearly independent one-electron orbitals which span the "occupied subspace." It was natural to ask whether by relaxing the orthogonality requirement one can describe the molecule by using even more localized orbitals without changing the many-electron single determinant wave function. The positive answer to this question was given by the theorem discussed next.

Let us consider the single-determinant wave function for a system treated by using a finite basis set, and let us assume that we have a set of zero-order approximations to the orbitals—both the occupied and the virtual ones. (The origin of this zero-order approximation is irrelevant from the point of view of the theorem.) The zero-order orbitals need not be orthonormalized; it is sufficient to require them to be linearly independent.

We shall denote $\varphi_i^0, i = 1, 2, \ldots, n$ the zero-order occupied orbitals and ψ_j^0, $j = 1, 2, \ldots, m-n$ the zero-order virtual ones, m and n being the dimension of the whole basis and the number of the occupied orbitals, respectively. As the zero-order orbitals are assumed to be linearly independent, they can be used as

a basis in which one can expand the "true" occupied orbitals from which the determinant wave function considered is built up.

We shall call the zero-order approximation $\{\varphi_i^0\}$, $\{\psi_j^0\}$ as being not "singularly wrong" if there is no orbital in the true occupied subspace that could be expanded *exactly* as a linear combination of the approximate *virtual* orbitals.

8.1. Existence theorem

This theorem states that if the zero-order approximation $\{\varphi_i^0\}$, $\{\psi_j^0\}$ is not singularly wrong, then the true determinant wave function can always be built up (up to an unimportant constant factor) by using n (generally nonorthogonal) orbitals φ_i having the special structure

$$\varphi_i = \varphi_i^0 + \sum_{j=1}^{m-n} \kappa_{ij}\psi_j^0 \tag{5.151}$$

(κ_{ij} being some coefficients), i.e., each orbital contains *only one* from the approximate (zero-order) occupied orbitals. (Therefore any mixing between the zero-order occupied orbitals can be considered as being merely an orthogonalization effect.)

Proof

Let us consider for a given i ($i = 1, 2, \ldots, n$) the functions of the structure

$$\varphi_i' = \alpha_i\varphi_i^0 + \sum_{j=1}^{m-n} \lambda_{ij}\psi_j^0 \,. \tag{5.152}$$

The orbitals in (5.152) with arbitrary coefficients α_i and λ_{ij} form a subspace of dimension $m - n + 1$, which means that we can choose $m - n + 1$ linearly independent vectors in that subspace. The true occupied subspace is of dimension n, i.e., we can choose n linearly independent vectors in it. If we join these subspaces, then we get $m - n + 1 + n = m + 1$ vectors—but they cannot be all linearly independent, as the overall dimension of the basis is only m. Consequently, these subspaces must have a common vector: it has the structure in (5.152) and completely lies in the true occupied subspace. For this vector coefficient α_i cannot be 0 because then the zero order approximation would be singularly wrong, which contradicts the assumption of theorem. Dividing (5.152) by $\alpha_i \neq 0$, and denoting $\varphi_i = \frac{\varphi_i'}{\alpha_i}$ and $\kappa_{ij} = \frac{\lambda_{ij}}{\alpha_i}$, we arrive at (5.151). Such orbitals can be constructed for each $i = 1, 2, \ldots, n$ value; each of them lies in the true occupied subspace and they are linearly independent, owing to the linear independence of the orbitals φ_i^0; therefore the determinant wave function considered can be built up by their use.

Q.E.D.

8.2. *A posteriori* determination

Now, we shall demonstrate that the orbitals φ_i of structure (5.151) not only exist in the mathematical sense, but one can also provide an algorithm for their practical calculation.

Let us introduce the common notation ϑ_i for the basis functions φ_i^0 and ψ_j^0 as

$$\vartheta_i = \begin{cases} \varphi_i^0 & (1 \le i \le n) \\ \psi_{i-n}^0 & (n < i \le m) \end{cases} . \tag{5.153}$$

The functions ϑ_i are assumed to be linearly independent, therefore the matrix of their overlap integrals $\langle \vartheta_i | \vartheta_j \rangle$ is not singular. We shall use the notation $\mathbf{Q}$ for the *inverse* of this overlap matrix.

Let us consider some (e.g., canonical) set of the true occupied orbitals, and denote them φ_i^{SCF}, and introduce the $m \times n$ rectangular matrix $\mathbf{D}$ containing the overlap integrals

$$D_{kj} = \langle \vartheta_k | \varphi_j^{SCF} \rangle . \tag{5.154}$$

We expand φ_j^{SCF} in the basis $\{\vartheta_i\}$:

$$\varphi_j^{SCF} = \sum_{l=1}^{m} f_l^j \vartheta_l . \tag{5.155}$$

Multiplying (5.155) with ϑ_k^* and integrating, we have

$$D_{kj} = \sum_{l=1}^{m} \langle \vartheta_k | \vartheta_l \rangle f_l^j \tag{5.156}$$

from which, after multiplying with the element Q_{qk} of the inverse overlap matrix $\mathbf{Q}$ and summing over k, we get

$$f_q^j = \sum_{k=1}^{m} Q_{qk} D_{kj} = (\mathbf{QD})_{qj} \tag{5.157}$$

i.e.,

$$\varphi_j^{SCF} = \sum_{l=1}^{m} (\mathbf{QD})_{lj} \vartheta_l . \tag{5.158}$$

On the other hand, function φ_i that we are seeking is in the true occupied subspace; therefore it can be presented as a linear combination of the functions φ_j^{SCF} spanning that subspace:

$$\varphi_i = \sum_{j=1}^{m} A_{ji} \varphi_j^{SCF} . \tag{5.159}$$

Substituting here the expansion (5.158), we can write

$$\varphi_i = \sum_{l=1}^{m} \sum_{j=1}^{n} (\mathbf{QD})_{lj} A_{ji} \vartheta_l \ . \tag{5.160}$$

Now, we return to the original notations φ_i^0 and ψ_j^0; expansion (5.159) then becomes

$$\varphi_i = \sum_{j,l=1}^{n} (\mathbf{QD})_{lj} A_{ji} \varphi_l^0 + \sum_{l=1}^{m-n} \sum_{j=1}^{n} (\mathbf{QD})_{l+n,j} A_{ji} \psi_l^0 \ . \tag{5.161}$$

We require φ_i to have the form of (5.151), which means that only one φ_l^0 in the first sum should have a nonzero coefficient:

$$\sum_{j=1}^{n} (\mathbf{QD})_{lj} A_{ji} = \delta_{li} \ . \tag{5.162}$$

This result means that we get orbitals satisfying the theorem, if matrix **A** is determined as the inverse of the upper $n \times n$ block of matrix **QD**.

After matrix **A** has been determined, the requested orbitals satisfying (5.151) can be determined by using (5.159). Alternately, one can determine the explicit value of the coefficients from (5.161) as

$$\kappa_{il} = \sum_{j=1}^{n} (\mathbf{QD})_{l+n,j} A_{ji} \ . \tag{5.163}$$

Notes

Sections 1–5.

Standard material (e.g., [1,2]); the presentation largely follows that in [3]. (The Slater determinant was introduced in [4] in which its most important properties were already established.)

Section 6.

The Slater rules were introduced in [4,5]. (Also see [6] – for that reason the terminology "Slater–Condon rules" is also sometimes used.) The matrix elements for non-orthogonal orbitals were first derived by Löwdin [7] by using a density matrix formalism. Also see, e.g., [2]. The present "factorization" technique may be (but is not claimed to be) original.

Section 7.1

The pairing theorem was first formulated by Löwdin as a conjecture [8] and was proved by Amos and Hall [9]. Also see [10–12]. The present proof has been published in [13]. The pairing theorem was mostly used in the spin-projected "extended" Hartree-Fock theory (for a survey see [12]); for a recent application in the theory of intermolecular interactions see [14].

Section 7.2

The extended pairing theorem was proposed and proved by Karadakov [15], by using the so

called "singular value decomposition" method (a generalization of diagonalization for the case of rectangular matrices). The present proof is a simplified version of that in [16].

Section 8

Present author's results [17]. For use in the theory of localized orbitals and delocalization interactions see [18, 19], for that in the theory of the "basis set superposition error" in calculating intermolecular interactions see [20].

Suggested readings

For the construction of spin eigenfunctions we refer e.g., to [2, 10, 21]. For different properties of determinant wave functions and the concept of density matrices not considered in the present book, a unique source is the series of Löwdin's papers [7, 22].

This is the proper place to stress that many results of this chapter can be obtained more easily by using the formalism of *second quantization* which permits one to "build in" the antisymmetry requirement in any wave functions. The books [23] and, especially, [24] represent good introductions. (Also see [2].)

References

1 L. Zülicke, *Quantenchemie*, Deutcher Verlag der Wissenschaften, Berlin 1973.

2 R. McWeeny, *Methods of Molecular Quantum Mechanics*, Academic Press, London 1992.

3 I. Mayer, *Fejezetek a Kvantumkémiából*, BME Mérnöktovábbképző Int., Budapest 1987.

4 J.C. Slater, Phys. Rev. **34**, 1293 (1929)

5 J.C. Slater, Phys. Rev. **38**, 1109 (1931)

6 E.U. Condon, Phys. Rev. **36**, 1121 (1930)

7 P.-O. Löwdin, Phys. Rev. **97**, 1470 (1955)

8 P.-O. Löwdin, J. Appl. Phys. Suppl. **33**, 251 (1962)

9 A. T. Amos and G. G. Hall, Proc. Roy. Soc. (London) **A263**, 483 (1961)

10 R. Pauncz, *Alternant Molecular Orbital Method* Saunders, Philadelphia 1967.

11 Y. I. Gorlov and I. I. Ukrainsky, Inst. for Theoretical Physics, Ukrainian Academy of Sciences, Preprints ITP-73-138 (1973) and ITP-74-69E (1974)

12 I. Mayer, Adv. Quantum Chem. **12**, 189 (1980).

13 I. Mayer, Int. J. Quantum Chem. **63**, 31 (1997)

14 A. Hamza and I. Mayer, Int. J. Quantum Chem. **82**, 53, 105 (2001)

15 P. Karadakov, Int. J. Quantum Chem. **27**, 699 (1985)

16 I. Mayer, Int. J. Quantum Chem. **29**, 31 (1986)

17 I. Mayer, Chem. Phys. Letters, **89**, 390 (1982)

18 P.R. Surján, I. Mayer and M. Kertész, J. Chem. Phys. **77**, 2454 (1982)

19 I. Mayer and P.R. Surján, J. Chem. Phys. **80**, 5649 (1984)

20 I. Mayer and L. Turi, J. Molec. Struct. (Theochem), **227**, 43 (1991)

21 R. Pauncz, *Spin Eigenfunctions*, Plenum Press, New York 1979.

22 P.-O. Löwdin, Phys. Rev. **97**, 1490, 1510 (1955)

23 P. Jørgensen and J. Simons, *Second Quantization-Based Methods in Quantum Chemistry*, Academic Press, New York 1981.

24 P.R. Surján, *Second Quantized Approach to Quantum Chemistry*, Springer, Berlin 1989.

Chapter 6

THE HARTREE–FOCK METHOD

The Hartree–Fock (HF) method has a central importance in quantum chemistry—both conceptually and computationally. In its case one approximates the wave function as a single Slater determinant and optimizes it according to the variation principle. Besides the importance of the HF method in itself, most of the more refined methods taking into account "electron correlation" also use the HF method as their starting point (Chapter 8).

The HF method is often called the "independent particle"—or "independent electron"—approximation (model). This does not mean that one neglects the interelectronic interactions, but only that their actual values are replaced by some "averaged" ones. For that reason one assigns one spin orbital to each electron (i.e., as many spin orbitals are used as many electrons are in the system) and optimizes these spin orbitals variationally. The variational procedure results in an effective averaged field determining the motion of the individual electrons. Pictorially one can describe the situation as follows. Each electron is moving in an effective field that represents the sum of the field caused by the nuclei and of the averaged field as the result of the other electrons. That means that the field determining the orbitals depends on the orbitals themselves: one has to find orbitals that generate a field in which the solutions are the same orbitals that generate that field. Then one speaks about a "self-consistent field" (SCF), and the procedure of finding such orbitals is called SCF procedure.

In the most common case of treating closed-shell molecules, one usually applies doubly filled *spatial* orbitals, i.e., the spin orbitals are obtained in pairs from the same spatial orbital—one with spin α, another with spin β. This is the "(spin-)restricted Hartree Fock" (RHF) method. If the requirement of double occupancy is lifted, one has the more general "(spin-)unrestricted Hartree–Fock" (UHF) scheme; however, its solutions often coincide with the RHF ones if applied for closed-shell systems (see section 4.3 for a discussion.)

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

As the overall form of the Hartree–Fock wave function (single determinant with or without the restriction of double occupancy) is predetermined, the application of the variation principle means that we should optimize the individual orbitals. This is usually done by solving the Hartree–Fock equations (which give the conditions of the energy to be stationary) in an iterational manner. Usually one starts from some approximate "guess" of the orbitals, determines the effective field they generate, and finds the orbitals that correspond to stationary one-electron states in that field. Then one recalculates the field, and again the orbitals, and repeats the whole procedure until some selected convergence criteria are fulfilled, indicating that the self-consistency is reached with the required accuracy.

1. The variation principle for single determinant wave functions: The Brillouin theorem

The Brillouin theorem is usually considered as a consequence of the Hartree–Fock equations. However, we prefer to proceed in the opposite direction: at first we show that the Brillouin theorem is equivalent to the variation principle for the single determinant wave functions and then derive the Hartree–Fock equations by expanding the Brillouin theorem.

Let us consider the Slater determinant wave function

$$\Psi_0 = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\dots\psi_N(N)] \tag{6.1}$$

that satisfies the variation principle and any of the "singly excited" determinants $\Psi_1 = \Psi_1(\psi_i \to \psi_i')$ that can be obtained from Ψ_0 by replacing the spin orbital ψ_i by a spin orbital ψ_i' that is arbitrary but is orthogonal to each occupied spin orbital $\psi_j, j = 1, 2, \dots, N$. *The Brillouin theorem states that the singly excited determinants Ψ_1 obtained in this manner do not interact with the variational (or "Hartree–Fock") determinant wave function Ψ_0:*

$$H_{01} = \langle \Psi_0 | \hat{H} | \Psi_1 \rangle = 0 \,. \tag{6.2}$$

Here $\hat{H}$ is the Born–Oppenheimer Hamiltonian of the system.

At first we are going to prove this theorem for the UHF determinant having the *absolute minimum* of the energy and then for an RHF or UHF determinant with a *stationary energy* (i.e., satisfying the condition $\delta E = 0$).

1.1. The Brillouin theorem for the determinant with the absolute minimum of the energy

Indirect proof

Let us assume that the determinant wave function Ψ_0 has the absolute minimum of the energy. We shall prove that the Brillouin theorem (6.2) must hold

for it, or else one could construct another single-determinant wave function with an even lower energy, which would contradict the assumptions of the theorem.

One can form the linear combination $\Psi = c_0\Psi_0 + c_1\Psi_1$ and solve the linear variational problem in order to determine the coefficients c_0 and c_1 to get a minimal energy for Ψ. Ψ_0 and Ψ_1 are determinants differing in only one column. Therefore, according to the properties of determinants, their linear combination Ψ can be presented also as a single determinant. It follows from the results in Section 6.1.3 of Chapter 5, that the determinant wave functions Ψ_0 and Ψ_1 are orthogonal. With no loss of generality we may assume them to be normalized as well. Therefore, the secular equation (3.15) can be used to determine the coefficients c_1 and c_2 and the corresponding energy E. The secular determinant is

$$\begin{vmatrix} H_{00} - E & H_{01} \\ H_{10} & H_{11} - E \end{vmatrix} = 0 \,. \tag{6.3}$$

The lower root of (6.3) can be brought by trivial algebra to the form

$$E = H_{00} + \frac{1}{2}(H_{00} - H_{11}) \left[\sqrt{1 + \frac{4|H_{01}|^2}{(H_{00} - H_{11})^2}} - 1 \right] , \tag{6.4}$$

where $H_{ij} = \langle \Psi_i | \hat{H} | \Psi_j \rangle$. $H_{00} - H_{11} < 0$, as Ψ_0 has the absolute minimum of the energy. Then, if we assume $H_{01} \neq 0$, we get from (6.4) $E < H_{00}$ for the energy of the determinant wave function Ψ. But H_{00} is the energy corresponding to Ψ_0 and it was assumed to be the *absolute minimum*. We arrived at a contradiction, therefore H_{01} must be equal to 0. (Then $E = H_{00}$.) Q.E.D.

1.2. The Brillouin theorem for a determinant with a stationary energy

In order to perform a variation of the determinant wave function $\Psi = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)]$, one has to vary the one-electron spin orbitals ψ_i by replacing each ψ_i which $\psi_i + \delta\psi_i = \psi_i + \eta\psi_i'$. Here ψ_i' represents an *arbitrary* spin orbital and η is an arbitrary complex parameter of variation tending to zero. Then the many-electron wave function obtained after the variation is

$$\Psi + \delta\Psi = \hat{\mathcal{A}}\{[\psi_1(1) + \eta\psi_1'(1)][\psi_2(2) + \eta\psi_2'(2)]\ldots[\psi_N(N) + \eta\psi_N'(N)]\} \,. \tag{6.5}$$

This is a determinant, each column of which is the sum of two terms ψ_i and $\eta\psi_i'$. According to the known properties of determinants, such a determinant can be written as a sum of 2^N determinants. However, most of these terms will contain the square or higher powers of η and can be neglected as compared with

the terms of first order in η. Thus we may write

$$\Psi + \delta\Psi = \Psi + \eta \sum_{i=1}^{N} \Psi_1(\psi_i \to \psi_i') \ . \tag{6.6}$$

Here $\Psi_1(\psi_i \to \psi_i')$ differs from Ψ by replacing the spin orbital ψ_i with the arbitrary spin orbital ψ_i'. (We are using no restrictions on the spin orbitals ψ_i' as yet.) Thus one obtains the most general first-order variation of the determinant wave function Ψ as

$$\delta\Psi = \eta \sum_{i=1}^{N} \Psi_1(\psi_i \to \psi_i') \ . \tag{6.7}$$

The variation (6.7) of the determinant wave function should now be substituted into (2.13) that gives the condition for the energy to be stationary ($\delta E = 0$). We get

$$\eta^* \sum_{i=1}^{N} \langle \Psi_1(\psi_i \to \psi_i') | \hat{H} - E | \Psi \rangle = 0 \tag{6.8}$$

$\eta \to 0$ but $\eta \neq 0$, so we may divide (6.8) by η^*. The variations $\delta\psi_i = \eta\psi_i'$ of the individual orbitals are independent, we may choose all but one $\psi_i' \equiv 0$. Therefore, one has to require the fulfillment of the equality

$$\langle \Psi_1(\psi_i \to \psi_i') | \hat{H} - E | \Psi \rangle = 0 \tag{6.9}$$

for every i separately. The variation principle written in the form of (6.9) is usually called the "generalized Brillouin theorem."

As we have seen, every singly excited determinant $\Psi_1(\psi_i \to \psi_i')$ is connected with the variation $\delta\psi_i = \eta\psi_i'$ of the spin orbital ψ_i. This variation can always be written as

$$\delta\psi_i = \delta\psi_{i1} + \delta\psi_{i2} = \eta(\psi_{i1}' + \psi_{i2}') \ , \tag{6.10}$$

where ψ_{i1}' is *orthogonal* to every spin orbital ψ_j occupied in Ψ, whereas ψ_{i2}' can be written as a linear combination of the occupied spin orbitals. (If $\hat{P}$ is the operator of projection on the occupied subspace, then $\psi_{i1}' = (1 - \hat{P})\psi_i'$ and $\psi_{i2}' = \hat{P}\psi_i'$.) It follows from the properties of the determinants that $\Psi_1(\psi_i \to \psi_i') = \Psi_1(\psi_i \to \psi_{i1}') + \Psi_1(\psi_i \to \psi_{i2}')$ and, therefore, the fulfillment of (6.9) should be required for both $\Psi_1(\psi_i \to \psi_{i1}')$ and $\Psi_1(\psi_i \to \psi_{i2}')$. However, the variation $\delta\psi_{i2} = \eta\psi_{i2}'$ leaves invariant the occupied subspace, thus it does not change the determinant wave function or changes only its normalization; therefore it does not cause any variation of the energy. (In other

words, $\Psi_1(\psi_i \to \psi'_{i2})$ is either 0 or is proportional to Ψ; in both cases (6.9) is fulfilled automatically.)

It follows from the previous discussion that in order to obtain the necessary and sufficient conditions for the energy of the determinant wave function Ψ to be stationary, it is enough to consider the case in which all $\delta\psi_{i2} = \eta\psi'_{i2} = 0$, i.e., restrict our consideration to the variations $\delta\psi_i = \eta\psi'_i = \eta\psi'_{i1}$ which are orthogonal to all occupied spin orbitals ψ_j. In that case $\Psi_1(\psi_i \to \psi'_i)$ is also orthogonal to Ψ: $\langle\Psi_1(\psi_i \to \psi'_i)|\Psi\rangle = 0$, and the generalized Brillouin theorem reduces to

$$\langle\Psi_1(\psi_i \to \psi'_i)|\hat{H}|\Psi\rangle = 0 \ . \tag{6.11}$$

This is the *Brillouin theorem*. It should be satisfied for all occupied spin orbitals ψ_i and arbitrary spin orbitals ψ'_i that are *orthogonal* to the occupied ones. In this case the Brillouin theorem is equivalent with the variation principle $\delta E = 0$ for single-determinant wave functions.

The equality (6.11) is satisfied trivially if ψ_i and ψ'_i have different spins, therefore one has to require the fulfillment of the Brillouin theorem for the cases when the spin orbital $\psi_i(\vec{r}, \sigma) = \varphi_i(\vec{r})\gamma_i(\sigma)$ is replaced by an arbitrary spin orbital $\psi'_i = \varphi'_i(\vec{r})\gamma_i(\sigma)$ of the same spin γ_i ($\gamma_i = \alpha$ or β), the spatial part φ'_i of which is orthogonal to all the spatial orbitals that are occupied in Ψ with the same spin. Therefore, the Brillouin theorem can also be written as

$$\langle\Psi_1(\varphi_i\gamma_i \to \varphi'_i\gamma_i)|\hat{H}|\Psi\rangle = 0 \tag{6.12}$$

for all φ'_i satisfying $\langle\varphi'_i|\varphi_j\rangle\delta_{\gamma_i\gamma_j} = 0$ $(i, j = 1, 2, \ldots N)$.

1.3. Algorithm of calculations based on the Brillouin theorem

In principle, the Brillouin theorem not only gives the conditions of stationary energy but it also can be used to construct a procedure for finding the wave functions satisfying this requirement. We shall discuss very briefly this possibility; it is based on considerations similar to those used in the proof of the Brillouin theorem shown in Section 1.1.

Let us assume that we have orthonormalized sets of approximate occupied and virtual spin orbitals $\{\psi_i^o\}$ and $\{\psi_j^v\}$, respectively, and wish to improve the approximation. We may recalculate the spin orbitals cyclically, one at a time, going through all the orbitals in each cycle. When the given orbital ψ_i^o is updated, all the other occupied orbitals are kept fixed. Similarly to the case considered in Section 1.1, one formulates a linear variational problem by requesting the energy of the linear combination

$$\Psi = c_0\Psi_0 + \sum_{j=1}^{m_v} c_j\Psi_1(\psi_j^o \to \psi_j^v) \tag{6.13}$$

to be minimal. This requires the solution of an eigenvalue equation of type (3.15), having the dimension $m_v + 1$, where m_v is the number of virtual orbitals. (Formally our considerations apply also in the limit $m_v \to \infty$.) Now, we utilize the fact that determinant Ψ_0 and the different determinants Ψ_1 are all determinants that differ from each other in the same column. Therefore, the wave function Ψ in (6.13) can also be written as a single determinant, in which the occupied spin orbital ψ_i^o is replaced by the updated one:

$$\psi_i^{o\prime} = c_0 \psi_i^o + \sum_{j=1}^{m_v} c_j \psi_j^v \,. \tag{6.14}$$

Similarly, the other eigenvectors of the secular equation can be used to get the new set of virtual orbitals. Obviously, if one uses doubly filled orbitals, then the two spin orbitals having the same spatial part can be updated at once. (However, the convergence considerations that follow do not necessarily hold in that case.)

In every step of this procedure, the new orbital is obtained by solving a linear variational problem and taking the lowest root. Repeating this procedure cyclically for all the orbitals, one obtains a decreasing (or at least, nonincreasing) sequence of the energy values. This sequence is, in principle, infinite, it is bounded from below (as follows from the variation principle); it is known from mathematics that such a sequence converges to a definite limit (minimum energy value). In the converged case the Brillouin theorem is fulfilled (otherwise we could get a further energy lowering by solving the linear variational problem (6.13)—c.f. Section 1.1)—therefore this scheme represents a necessarily convergent algorithm for solving the SCF problem, at least in the UHF case. However, this scheme is very expensive computationally. Assuming that we are not too far from the desired solution, one could try to reduce the computational cost by using a perturbational approximation instead of solving the eigenvalue problem. (In the latter case special care must be taken to ensure the orthogonality of the new virtual orbitals—c.f. Appendix XI.) Also, one can try to determine the correction for all orbitals simultaneously, which may reduce the necessary computational work. Of course, in this case one loses the strict guarantee of convergence.

2. The Hartree–Fock equations

The Brillouin theorem gives the necessary and sufficient conditions for the energy of the determinant wave function Ψ_0 to be stationary, expressed in terms of many-electron wave functions: the determinant Ψ_0 itself and all the possible singly excited determinants $\Psi_1(\psi_i \to \psi_i')$ in which spin orbital ψ_i is replaced by an arbitrary spin orbital ψ_i', orthogonal to all the occupied orbitals. By expanding the Brillouin theorem in terms of the one-electron orbitals one can derive the equations that the occupied orbitals should satisfy.

2.1. Unrestricted Hartree–Fock equations

At first we shall consider the UHF case, i.e., we shall not yet make any assumption about the double occupancy of the spatial orbitals. We may assume with no loss of generality that the determinant wave function Ψ with a stationary energy is built up of orthonormalized spin orbitals. (As discussed in Section 5 of Chapter 5, the orthonormalization of the orbitals leaves invariant the determinant wave function up to a physically irrelevant constant factor.)

One can expand the Brillouin theorem (6.12) by using the results of Sections 6.2.2 and 6.3.2 of Chapter 5 for the matrix elements of the one- and two-electron parts of the Hamiltonian, calculated between determinants Ψ_0 and Ψ_1 built up of the orthonormalized spin orbitals $\psi_i(\vec{r},\sigma) = \varphi_i(\vec{r})\gamma_i(\sigma)$, and differing in one orbital—in Ψ_1 φ_i is replaced by φ_i'—as:

$$\langle\Psi_1(\varphi_i\gamma_i \to \varphi_i'\gamma_i)|\hat{H}|\Psi\rangle \tag{6.15}$$

$$= \langle\varphi_i'|\hat{h}|\varphi_i\rangle + \sum_{\substack{j=1\\(j\neq i)}}^{N} ([\varphi_i'\varphi_j|\varphi_i\varphi_j] - [\varphi_i'\varphi_j|\varphi_j\varphi_i]\cdot\delta_{\gamma_i\gamma_j}) = 0\ ,$$

where we use the notation

$$[\varphi_i\varphi_j|\varphi_k\varphi_l] = \iint \frac{\varphi_i^*(\vec{r}_1)\varphi_j^*(\vec{r}_2)\varphi_k(\vec{r}_1)\varphi_l(\vec{r}_2)}{r_{12}}\, dv_1 dv_2 \tag{6.16}$$

for the two-electron repulsion integrals. (From now on we shall substitute everywhere the actual interelectronic interaction $\frac{1}{r_{12}}$ for the general two-electron operator $\hat{g}(1,2)$ used in Section 6.3 of Chapter 5.) Equation (6.15) can also be written as

$$\int \varphi_i'^*(\vec{r}_1)\left\{\left[\hat{h} + \sum_{\substack{j=1\\(j\neq i)}}^{N} (\hat{J}_j - \hat{K}_j\cdot\delta_{\gamma_i\gamma_j})\right]\varphi_i(\vec{r}_1)\right\}dv_1 = 0\ , \tag{6.17}$$

where we have introduced the "Coulomb" and "exchange" operators defined by their action on the arbitrary function $\varphi(\vec{r})$ as

$$\hat{J}_j\varphi(\vec{r}_1) = \int \frac{|\varphi_j(\vec{r}_2)|^2}{r_{12}}\, dv_2\varphi(\vec{r}_1) = \int \frac{\varphi_j^*(\vec{r}_2)\varphi_j(\vec{r}_2)}{r_{12}}\, dv_2\varphi(\vec{r}_1) \tag{6.18}$$

and

$$\hat{K}_j\varphi(\vec{r}_1) = \int \frac{\varphi_j^*(\vec{r}_2)\varphi(\vec{r}_2)}{r_{12}}\, dv_2\varphi_j(\vec{r}_1)\ . \tag{6.19}$$

We shall call the attention to the "nonlocal" character of the exchange operator $\hat{K}_j$.

If φ_i' in (6.17) were a completely arbitrary function, then the fundamental lemma of the calculus of variations (a function which is orthogonal to an *arbitrary* function is zero "almost everywhere") would imply that the function in the curly brackets vanishes. But φ_i' is restricted to be orthogonal to all the orbitals φ_j that are occupied in Ψ with the same spin as φ_i. Therefore, the function in the curly brackets can have component proportional to the orbitals filled with the same spin—but no components in the orthogonal complement. Thus we get the UHF equations in the following general form:

$$\left[\hat{h} + \sum_{\substack{j=1 \\ (j\neq i)}}^{N} (\hat{J}_j - \hat{K}_j \cdot \delta_{\gamma_i\gamma_j})\right]\varphi_i = \sum_{j=1}^{N} \lambda_{ji}\varphi_j \cdot \delta_{\gamma_i\gamma_j} \qquad (i = 1, 2, \ldots, N)\ , \tag{6.20}$$

where the coefficients λ_{ij} may be arbitrary.

One can easily see that the derivation can also be performed in the opposite direction, therefore the fulfillment of the UHF equation (6.20) gives not only the necessary but also the sufficient conditions for the energy of a single determinant wave function to be stationary ($\delta E = 0$).

Now we shall concretize the equations for the case when there are n_a orbitals $a_i = a_i(\vec{r})$ filled with spin α and n_b orbitals $b_i = b_i(\vec{r})$ filled with spin β. (That is $\varphi_i = a_i$ or b_i.) Then the wave function Ψ can be written as

$$\Psi = \hat{\mathcal{A}}[a_1(\vec{r}_1)\alpha(\sigma_1)b_1(\vec{r}_2)\beta(\sigma_2)a_2(\vec{r}_3)\alpha(\sigma_3)\ldots]\ . \tag{6.21}$$

We define the Coulomb and exchange operators $\hat{J}_j^a, \hat{J}_j^b, \hat{K}_j^a, \hat{K}_j^b$ corresponding to orbitals a_j, b_j analogously to the definitions (6.18), (6.19) above; for instance

$$\hat{J}_j^a\varphi(\vec{r}_1) = \int \frac{a_j^*(\vec{r}_2)a_j(\vec{r}_2)}{r_{12}}\, dv_2\varphi(\vec{r}_1) \tag{6.22}$$

and

$$\hat{K}_j^b\varphi(\vec{r}_1) = \int \frac{b_j^*(\vec{r}_2)\varphi(\vec{r}_2)}{r_{12}}\, dv_2 b_j(\vec{r}_1)\ . \tag{6.23}$$

Then the UHF equations can be rewritten as

$$\left[\hat{h} + \sum_{\substack{j=1 \\ (j\neq i)}}^{n_a} (\hat{J}_j^a - \hat{K}_j^a) + \sum_{j=1}^{n_b} \hat{J}_j^b\right]a_i = \sum_{j=1}^{n_a} \varepsilon_{ji}^a a_j \qquad (i = 1, 2, \ldots, n_a) \tag{6.24}$$

and we get a similar set of equations for the orbitals b_i—they can be obtained from (6.24) by interchanging a and b everywhere. It is easy to check directly

that $\hat{J}_i^a a_i = \hat{K}_i^a a_i$, therefore one may add to the left-hand side the so called "self-repulsion" term $(\hat{J}_i^a - \hat{K}_i^a)a_i = 0$, permitting one to omit the restriction $j \neq i$ in the summation. As a result, the operator in brackets becomes the same for each orbital.

We introduce the Fock operators ("Fockians") $\hat{F}^a$ and $\hat{F}^b$

$$\begin{aligned} \hat{F}^a &= \hat{h} + \sum_{j=1}^{n_a}(\hat{J}_j^a - \hat{K}_j^a) + \sum_{j=1}^{n_b} \hat{J}_j^b \\ \hat{F}^b &= \hat{h} + \sum_{j=1}^{n_b}(\hat{J}_j^b - \hat{K}_j^b) + \sum_{j=1}^{n_a} \hat{J}_j^a \end{aligned} \tag{6.25}$$

by the use of which the UHF equations—representing the conditions of the energy to be stationary—can be written compactly as

$$\begin{aligned} \hat{F}^a a_i &= \sum_{j=1}^{n_a} \varepsilon_{ji}^a a_j \qquad (i = 1, 2, \ldots, n_a)\ ; \\ \hat{F}^b b_i &= \sum_{j=1}^{n_b} \varepsilon_{ji}^b b_j \qquad (i = 1, 2, \ldots, n_b)\ . \end{aligned} \tag{6.26}$$

Operators $\hat{F}^a$ and $\hat{F}^b$ depend on all the orbitals filled with either spin in the wave function Ψ but in such a manner that they are uniquely determined by the two *subspaces* spanned by the sets of occupied orbitals $\{a_i\}$ and $\{b_i\}$. This follows from the fact that $\hat{F}^a$ and $\hat{F}^b$ are invariant under the unitary transformations between the occupied orbitals of the same spin. (One does not consider the mixing of spin orbitals of different spins because that would result in "general spin orbitals" containing both spin functions α and β.)

Proof of the invariance

Let us consider, for instance, the unitary transformation of the orbitals a_i, and calculate the sum

$$\begin{aligned} \sum_{i=1}^{n_a} \hat{J}_i^a &= \sum_{i=1}^{n_a} \int \frac{a_i^*(\vec{r}_2)a_i(\vec{r}_2)}{r_{12}}\, dv_2 \\ &= \sum_{i=1}^{n_a} \int \frac{\sum_{j=1}^{n_a} U_{ji}^* a_j'^*(\vec{r}_2) \sum_{k=1}^{n_a} U_{ki} a_k'(\vec{r}_2)}{r_{12}}\, dv_2 \\ &= \sum_{i,j,k=1}^{n_a} U_{ki} U_{ji}^* \int \frac{a_j'^*(\vec{r}_2) a_k'(\vec{r}_2)}{r_{12}}\, dv_2\ . \end{aligned} \tag{6.27}$$

Because $U^*_{ji} = (\mathbf{U}^\dagger)_{ij} = (\mathbf{U}^{-1})_{ij}$, we have $\sum_{i=1}^{n_a} U_{ki}U^*_{ji} = \delta_{kj}$, thus the right-hand side of (6.27) simplifies, and we get

$$\sum_{i=1}^{n_a} \hat{J}_i^a = \sum_{j=1}^{n_a} \int \frac{a_j'^*(\vec{r}_2)a_j'(\vec{r}_2)}{r_{12}}\, dv_2 = \sum_{j=1}^{n_a} \hat{J}_j^{a\,\prime} . \tag{6.28}$$

The invariance of the *sums* of the operators $\hat{J}_j^b$, $\hat{K}_j^a$, and $\hat{K}_j^b$, respectively, can be verified exactly in the same manner. Q.E.D.

The wave function Ψ is also invariant with respect to the unitary transformations of the occupied orbitals (except, possibly, an unimportant phase factor of absolute value 1); its energy is invariant, as well. The previous derivation of the UHF equations giving the necessary and sufficient conditions of the energy to be stationary is equally applicable for every set of orthonormalized orbitals spanning the occupied subspace; accordingly, every such set should satisfy the UHF equations in (6.26). The necessary and sufficient conditions of the stationary energy can therefore be formulated by saying that the occupied orbitals a_i and b_i form an *invariant subspace* of the Fock operators $\hat{F}^a$ and $\hat{F}^b$, respectively. (This means that acting by operator $\hat{F}^a$ on any orbital in the occupied subspace $\{a_i\}$ we get a function that is in the same subspace.)

The Coulomb and exchange operators are Hermitian. For instance, one has for arbitrary functions φ and ψ

$$\begin{aligned} \langle\varphi|\hat{J}^a\psi\rangle &= \int \varphi^*(\vec{r}_1)dv_1 \left[\int \frac{|a_i(\vec{r}_2)|^2}{r_{12}}\, dv_2\psi(\vec{r}_1)\right] \\ &= \iint \frac{\varphi^*(\vec{r}_1)\psi(\vec{r}_1)|a_i(\vec{r}_2)|^2}{r_{12}}\, dv_1 dv_2 . \end{aligned} \tag{6.29}$$

To compute $\langle\hat{J}_i^a\varphi|\psi\rangle$ we have to substitute the function $[\hat{J}_i^a\varphi(\vec{r})]^*$ in the integrand. We get

$$\begin{aligned} \langle\hat{J}_i^a\varphi|\psi\rangle &= \int \left[\int \frac{|a_i(\vec{r}_2)|^2}{r_{12}}\, dv_2\varphi(\vec{r}_1)\right]^* \psi(\vec{r}_1)dv_1 \\ &= \iint \frac{|a_i(\vec{r}_2)|^2\varphi^*(\vec{r}_1)\psi(\vec{r}_1)}{r_{12}}\, dv_1 dv_2 = \langle\varphi|\hat{J}_i^a\psi\rangle . \end{aligned} \tag{6.30}$$

Similarly, we have

$$\begin{aligned} \langle\varphi|\hat{K}_i^a\psi\rangle &= \int \varphi^*(\vec{r}_1) \left[\int \frac{a_i^*(\vec{r}_2)\psi(\vec{r}_2)}{r_{12}}\, dv_2 a_i(\vec{r}_1)\right] dv_1 \\ &= \iint \frac{\varphi^*(\vec{r}_1)a_i^*(\vec{r}_2)a_i(\vec{r}_1)\psi(\vec{r}_2)}{r_{12}}\, dv_1 = [\varphi a_i|a_i\psi] \end{aligned} \tag{6.31}$$

and

$$\langle \hat{K}_i^a \varphi | \psi \rangle = \int \left[\int \frac{a_i^*(\vec{r}_2)\varphi(\vec{r}_2)}{r_{12}} dv_2 a_i(\vec{r}_1) \right]^* \psi(\vec{r}_1) dv_1 \tag{6.32}$$

$$= \iint \frac{a_i^*(\vec{r}_1)\varphi^*(\vec{r}_2)\psi(\vec{r}_1)a_i(\vec{r}_2)}{r_{12}} dv_1 dv_2 = [a_i\varphi|\psi a_i] \, .$$

Interchanging the variables of integration in the definition in (6.16) leads to the identity

$$[ab|cd] \equiv [ba|dc] \, . \tag{6.33}$$

Utilizing this, we can transform the right-hand side of (6.32) and get

$$\langle \hat{K}_i^a \varphi | \psi \rangle = [\varphi a_i | a_i \psi] = \langle \varphi | \hat{K}_i^a \psi \rangle \, . \tag{6.34}$$

Q.E.D.

The Fockians $\hat{F}^a$ and $\hat{F}^b$ represent sums of Hermitian operators, therefore they are themselves Hermitian, too. The quantities ε_{ji}^a and ε_{ji}^b can be considered as elements of the $n_a \times n_a$ and $n_b \times n_b$ matrices $\boldsymbol{\varepsilon}^a$ and $\boldsymbol{\varepsilon}^b$, respectively. Matrices $\boldsymbol{\varepsilon}^a$ and $\boldsymbol{\varepsilon}^b$ are also Hermitian.

In fact, let us consider

$$\hat{F}^a a_i = \sum_{j=1}^{n_a} \varepsilon_{ji}^a a_j \, , \tag{6.35}$$

multiply it by a_k^*, and integrate. We get

$$\langle a_k | \hat{F}^a | a_i \rangle = \sum_{j=1}^{n_a} \varepsilon_{ji}^a \underbrace{\langle a_k | a_j \rangle}_{\delta_{kj}} \tag{6.36}$$

i.e.,

$$\varepsilon_{ki}^a = \langle a_k | \hat{F}^a | a_i \rangle \, . \tag{6.37}$$

At the same time one has, taking into account the Hermiticity of operator $\hat{F}^a$,

$$\varepsilon_{ik}^a = \langle a_i | \hat{F}^a | a_k \rangle = (\langle a_k | \hat{F}^a | a_i \rangle)^* = \varepsilon_{ki}^{a*} \, . \tag{6.38}$$

The Hermiticity of matrix $\boldsymbol{\varepsilon}^b$ can be proved in the same manner.
Q.E.D.

According to (6.37), matrix $\boldsymbol{\varepsilon}^a$ is the matrix of the Fock operator $\hat{F}^a$ in the occupied subspace $\{a_i\}$. (Similarly, $\boldsymbol{\varepsilon}^b$ gives the representation of $\hat{F}^b$ in the

subspace $\{b_i\}$.) As matrix $\boldsymbol{\varepsilon}^a$ is Hermitian, it can be diagonalized by a unitary transformation

$$\mathbf{U}^\dagger \boldsymbol{\varepsilon}^a \mathbf{U} = \mathrm{diag}\{\varepsilon_i^{a\,\prime}\} \ . \tag{6.39}$$

Let us subject the orbitals a_i to the unitary transformation $\mathbf{U}$ diagonalizing matrix $\boldsymbol{\varepsilon}^a$. If

$$a_i' = \sum_{j=1}^{n_a} U_{ji} a_j \ , \tag{6.40}$$

then we get, utilizing the invariance of $\hat{F}^a$ with respect to unitary transformations of the orbitals,

$$\begin{aligned} \varepsilon_{ik}^{a\,\prime} &= \langle a_i' | \hat{F}^a | a_k' \rangle = \langle \sum_{j=1}^{n_a} U_{ji} a_j | \hat{F}^a | \sum_{l=1}^{n_a} U_{lk} a_l \rangle = \sum_{j,l=1}^{n_a} U_{ji}^* \underbrace{\langle a_j | \hat{F}^a | a_l \rangle}_{\varepsilon_{jl}^a} U_{lk} \\ &= \sum_{j,l=1}^{n_a} (\mathbf{U}^\dagger)_{ij} \varepsilon_{jl}^a U_{lk} = (\mathbf{U}^\dagger \boldsymbol{\varepsilon}^a \mathbf{U})_{ik} = \varepsilon_i^{a\,\prime} \delta_{ik} \end{aligned} \tag{6.41}$$

in full accord with the previous statement that matrix $\boldsymbol{\varepsilon}^a$ is the matrix representing operator $\hat{F}^a$ in the occupied subspace.

One can proceed analogously with matrix $\boldsymbol{\varepsilon}^b$ and the orbitals b_i, and—omitting the primes—one gets the so called "canonical" form of the UHF equations:

$$\begin{aligned} \hat{F}^a a_i &= \varepsilon_i^a a_i \qquad (i = 1, 2, \ldots, n_a) \ ; \\ \hat{F}^b b_i &= \varepsilon_i^b b_i \qquad (i = 1, 2, \ldots, n_b) \ . \end{aligned} \tag{6.42}$$

These equations represent a system of coupled "pseudo-eigenvalue equations." One uses the word "pseudo" because the equations have the form of eigenvalue equations but the operators $\hat{F}^a$, $\hat{F}^b$ themselves depend on their eigenfunctions we are seeking; the equations are "coupled" because in order to find the solution for a given orbital one needs the other ones, as well. Accordingly, by using the HF method, the $3N$-dimensional Schrödinger equation is reduced approximately to a system of N coupled three-dimensional equations, which is much easier to solve. The eigenvalues ε_i^a, ε_i^a are usually called "orbital energies" for the reasons that will be discussed in Section 3.

Equations of this type are usually solved by iterations: one starts from an initial guess of the orbitals, builds up the Fock operators by their use, gets new orbitals by solving the equations and forms the next Fock operator by using these new orbitals... Symbolically this SCF process may be depicted as

$$\{\varphi^{[0]}\} \to \hat{F}^{[1]} \to \{\varphi^{[1]}\} \to \hat{F}^{[2]} \to \ldots \tag{6.43}$$

The procedure should be repeated until convergence is achieved with a required accuracy. The convergence criterion can be related to the orbitals or (better) to the subspace spanned by the occupied orbitals; usually the energy values are also monitored. If one has a fortunate initial guess $\{\varphi_i^{[0]}\}$ then one can get a quick convergence; in the case of worse choices, the SCF procedure can converge slowly or not at all (diverge).

In Section 5.3. we shall derive a special version of the UHF equations that permits the successive optimization of the orbitals in the spirit of the algorithm discussed in Section 1.3. Again, the problem is that such a method is necessarily convergent but very expensive.

2.2. Alternate derivation by using Lagrangian multipliers

In the most common derivation of the Hartree–Fock equations, the elements of the matrix ε emerge as Lagrangian multipliers corresponding to the requirement that the one-electron orbitals should be orthonormalized. However, this method—which can be found in many textbooks —raises some questions of mathematical character, so we shall discuss the problem with special attention.

In this derivation one starts from the UHF energy expression derived for the case of orthonormalized orbitals; it can be written out by combining Eq. (5.101) and (5.116) and using the simplified notation (6.16) as

$$\begin{aligned} E &= \sum_{i=1}^{n_a}\langle a_i|\hat{h}|a_i\rangle + \sum_{i=1}^{n_b}\langle b_i|\hat{h}|b_i\rangle + \frac{1}{2}\sum_{i,j=1}^{n_a}\left([a_ia_j|a_ia_j] - [a_ia_j|a_ja_i]\right) \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{n_b}\left([b_ib_j|b_ib_j] - [b_ib_j|b_jb_i]\right) + \sum_{i=1}^{n_a}\sum_{j=1}^{n_b}[a_ib_j|a_ib_j]\ , \end{aligned} \tag{6.44}$$

and requires *this* E to be stationary under the requirements of orthonormalization

$$\begin{aligned} \langle a_i|a_j\rangle &= \delta_{ij} \qquad (i,j = 1,2,\ldots,n_a)\ ; \\ \langle b_i|b_j\rangle &= \delta_{ij} \qquad (i,j = 1,2,\ldots,n_b)\ . \end{aligned} \tag{6.45}$$

For that reason one forms the auxiliary functional

$$F = E - \sum_{i,j=1}^{n_a}\varepsilon_{ji}^a(\langle a_i|a_j\rangle - \delta_{ij}) - \sum_{i,j=1}^{n_b}\varepsilon_{ji}^b(\langle b_i|b_j\rangle - \delta_{ij}) \tag{6.46}$$

by the use of the Lagrangian multipliers $-\varepsilon_{ji}^a$ and $-\varepsilon_{ji}^b$ and requires $\delta F = 0$. (For further convenience, the Lagrangian multipliers are introduced with a minus sign.)

The following remark is appropriate concerning (6.46). Each equation (6.45) with $i \neq j$ represents two independent real conditions—one for the real part,

another for the imaginary part of the integral $\langle a_i|a_j\rangle$ (or $\langle b_i|b_j\rangle$); therefore we have to introduce two independent Lagrangian multipliers ε^a_{ij} and ε^a_{ji} for the integral $\langle a_i|a_j\rangle$ and its complex conjugate $\langle a_j|a_i\rangle$. In fact, let us formulate the orthogonality condition for the orbitals a_i and a_j as

$$\begin{aligned} \mathrm{Re}\langle a_i|a_j\rangle &= \frac{1}{2}(\langle a_i|a_j\rangle + \langle a_j|a_i\rangle) = 0\ ; \\ \mathrm{Im}\langle a_i|a_j\rangle &= \frac{1}{2i}(\langle a_i|a_j\rangle - \langle a_j|a_i\rangle) = 0\ . \end{aligned} \tag{6.47}$$

We can introduce the Lagrangian multipliers μ_{ij} and ν_{ij} for these two conditions. Then one will have the following term in the auxiliary functional

$$\begin{aligned} &\mu_{ij}\frac{1}{2}(\langle a_i|a_j\rangle + \langle a_j|a_i\rangle) + \nu_{ij}\frac{1}{2i}(\langle a_i|a_j\rangle - \langle a_j|a_i\rangle) \\ &= \frac{1}{2}\left[(\mu_{ij} - i\nu_{ij})\langle a_i|a_j\rangle + (\mu_{ij} + i\nu_{ij})\langle a_j|a_i\rangle\right]\ . \end{aligned} \tag{6.48}$$

Denoting $\varepsilon^a_{ji} = -\frac{1}{2}(\mu_{ij} - i\nu_{ij})$ and $\varepsilon^a_{ij} = -\frac{1}{2}(\mu_{ij} + i\nu_{ij})$ we arrive at the form used in (6.46)—and also we may conclude that the matrix of Lagrangian multipliers should be Hermitian in problems of this type. (Similar consideration hold for the multipliers ε^b_{ij}, of course.)

Now, we shall perform the variation of some orbital a_k. We get

$$\begin{aligned} \delta F &= \langle \delta a_k|\hat{h}|a_k\rangle + \frac{1}{2}\sum_{j=1}^{n_a}([\delta a_k a_j|a_k a_j] - [\delta a_k a_j|a_j a_k]) \\ &\quad + \frac{1}{2}\sum_{i=1}^{n_a}([a_i\delta a_k|a_i a_k] - [a_i\delta a_k|a_k a_i]) \\ &\quad + \sum_{j=1}^{n_b}([\delta a_k b_j|a_k b_j] - \sum_{j=1}^{n_a}\varepsilon^a_{jk}\langle\delta a_k|a_j\rangle + c.c. = 0\ . \end{aligned} \tag{6.49}$$

Owing to the symmetry of the two-electron integrals, the first two sums can be combined into one. Then using the definition in (6.25) of the operator $\hat{F}^a$, one gets

$$\begin{aligned} \delta F &= \langle \delta a_k|\hat{F}^a a_k\rangle - \sum_{j=1}^{n_a}\varepsilon^a_{jk}\langle\delta a_k|a_j\rangle + c.c. \\ &= \int \delta a_k^* \left[\hat{F}^a a_k - \sum_{j=1}^{n_a}\varepsilon^a_{jk}a_j\right] dv + c.c. = 0\ . \end{aligned} \tag{6.50}$$

Here δa_k is fully arbitrary (it also contains an arbitrary phase factor), therefore the functions in the braces should vanish "almost everywhere," leading to the

UHF equations

$$\hat{F}^a a_k = \sum_{j=1}^{n_a} \varepsilon^a_{jk} a_j \qquad k = 1, 2, \ldots, n_a \,. \tag{6.51}$$

The equations for orbitals b_k can be derived in the same manner.

It should be noted that the previous derivation does not quite correspond to the standard method of using Lagrangian multipliers as prescribed in variational calculus. According to the latter, one has to form the auxiliary functional F' as the sum of the integral $\langle\Psi|\hat{H}|\Psi\rangle$ over the determinant wave function Ψ and of the constraints multiplied by the appropriate Lagrangian multipliers:

$$F' = \langle\Psi|\hat{H}|\Psi\rangle + \sum_{i,j=1}^{n_a} \lambda^a_{ij}(\langle a_i|a_j\rangle - \delta_{ij}) + \sum_{i,j=1}^{n_b} \lambda^b_{ij}(\langle b_i|b_j\rangle - \delta_{ij}) \tag{6.52}$$

and then to perform its variation as for a *free* variational problem; the auxiliary conditions (constraints) must be substituted only *afterward*. This means that the overlap integrals between the orbitals, and the minors of the overlap integral matrix formed by their use—which are implicitly present in the energy formula (6.44) as zero or unity values— should also be varied. It is easy to see by considering some specified variations of one orbital that this standard mathematical treatment leads to results quite different from those of the previous derivation:

a) Let $\delta a_i = \eta a_k$ for $i \neq k$. This means that we multiply one column of the determinant Ψ by a factor η and add it to another column. This does not change the wave function according to the properties of determinants, therefore $\delta\langle\Psi|\hat{H}|\Psi\rangle = 0$, and requiring $\delta F' = 0$ we arrive at $\eta^*\lambda^a_{ik} + \eta\lambda^a_{ki} = 0$. As η contains an arbitrary phase factor, this is possible only if $\lambda^a_{ik} = \lambda^a_{ki} = 0$. The zero value obtained for the off-diagonal matrix elements λ^a_{ik} is in agreement with the fact that the orthogonality of the orbitals can always be achieved and, therefore, it does not represent a true constraint in the variational problem.

b) Let $\delta a_i = \eta a_i$. This means that after the variation one of the columns of the determinant Ψ is multiplied by the factor $1 + \eta$. Owing to the properties of the determinants, this variation changes the wave function Ψ with the same factor, therefore $\delta\Psi = \eta\Psi$. Requiring $\delta F' = 0$, we get $\delta F' = \eta\langle\Psi|\hat{H}|\Psi\rangle + \lambda^a_{ii}\eta\langle a_i|a_i\rangle + c.c. = 0$. Utilizing the condition $\langle a_i|a_i\rangle = 1$, we arrive at $\lambda^a_{ii} = -\langle\Psi|\hat{H}|\Psi\rangle$. If all the auxiliary conditions in (6.45) are satisfied, then $\langle\Psi|\Psi\rangle = 1$ and $\lambda^a_{ii} = -E$, where E is the energy of the determinant wave function Ψ. This result is in full agreement with those obtained in the other derivations performed by using Lagrangian multipliers (Sections 1.3 and 1.5 of Chapter 2 and Section 1 of Chapter 3), because the conditions in (6.45) for $i = j$ are directly connected with the normalization of the wave function Ψ.

It is obvious that these Lagrangian multipliers λ^a_{ij}, λ^b_{ij} are independent of the Lagrangian multipliers ε^a_{ij}, ε^b_{ij} used in (6.46).

The proper variation of the overlap integrals and of different minors of the overlap integral matrix necessary to perform a complete derivation by using the functional (6.52) is very cumbersome. It can be performed in principle (the present author had accomplished such a derivation once in his youth) and, of course, leads to equations equivalent to those derived by using other methods.

One needs an explanation why in such circumstances the rather simple derivation with the use of Lagrangian multipliers given in (6.46)–(6.51) does lead to the correct equations. The point is that one should not consider the variation of E in (6.46) as a variation of the energy, but the procedure should be understood as follows. One defines E in (6.44) derived for the UHF energy by using the auxiliary conditions as a new auxiliary functional of the orbitals, which has, strictly speaking, no immediate physical meaning, but it is *equal* to the energy *when* (and only when) the auxiliary conditions are fulfilled. Then one has to search for the stationary value of this auxiliary quantity, subjected to the same conditions. Within the class of functions satisfying the auxiliary conditions, this quantity and the energy have, of course, identical stationary value(s) realized for the same wave function(s). However, from a formal mathematical point of view, we have two different variational problems (one for the true energy, another for the auxiliary quantity), therefore their Lagrangian multipliers will, of course, be different.

This distinction may appear immaterial from the practical point of view. One should, however, keep it in mind when applying Lagrangian multipliers to avoid mistakes in more complex cases of studying constrained variational problems.

2.3. Alternate derivation by using specified variations

We do already have everything necessary for another alternate derivation of the UHF equations.

Let us again consider the energy formula (6.44) corresponding to the orthonormalized one-electron orbitals, and let us consider only such variations of them that do not violate (at least in first order) the orthonormality requirement. That means that this energy formula remains applicable after the variation is performed. Consequently, under these conditions the variation of the energy is given by that of formula (6.44).

At first let us consider a variation of the orbital a_k as

$$\delta a_k = \eta c \,, \tag{6.53}$$

where $\eta \to 0$ and orbital c satisfies the requirement

$$\langle c|a_i\rangle = 0 \,; \qquad i = 1, 2, \ldots, n_a \,. \tag{6.54}$$

Such a variation conserves the orthonormalization of the orbitals in the zeroth and first order of the variational parameter η.

The variation of the energy expression (6.44) is given in (6.50) as $\langle \delta a_k | \hat{F} | a_k \rangle + c.c.$ Therefore, the requirement $\delta E = 0$ under variation (6.53) leads to

$$\eta^* \langle c | \hat{F} | a_k \rangle = 0 \ . \tag{6.55}$$

As $\eta \to 0$, but $\eta \neq 0$, (6.55) can be rewritten as

$$\int c^* \hat{F} a_k dv = 0 \ . \tag{6.56}$$

Function c is arbitrary but is required to be orthogonal to all the occupied orbitals a_i. As discussed in Section 2.1, this means that the function $\hat{F} a_k$ in (6.56) can have only components proportional to the orbitals occupied with the same spin as a_k:

$$\hat{F} a_k = \sum_{j=1}^{n_a} \varepsilon_{jk} a_j \ ; \qquad k = 1, 2, \ldots, n_a \ . \tag{6.57}$$

One can derive similarly the UHF equations for the orbitals b_k.

Let us now consider some other variations, as well. The variation $\delta a_k = \eta a_i \ \ (i \neq k)$ violates the orthonormalization conditions, but the latter can be restored by considering the simultaneous variations

$$\begin{aligned} \delta a_k &= \eta a_i \\ \delta a_i &= -\eta^* a_k \ . \end{aligned} \tag{6.58}$$

Neglecting the terms quadratic in η, the variation of the energy expression (6.44) under these simultaneous variation is the sum of the variations induced by the variations δa_k and δa_i, respectively.

$$\langle \delta a_k | \hat{F} | a_k \rangle + \langle \delta a_i | \hat{F} | a_i \rangle + c.c. = 0 \ . \tag{6.59}$$

Substituting (6.58) and writing out explicitly the complex conjugate,

$$\eta^* \langle a_i | \hat{F} | a_k \rangle - \eta \langle a_k | \hat{F} | a_i \rangle + \eta \langle a_k | \hat{F}^\dagger | a_i \rangle - \eta^* \langle a_i | \hat{F}^\dagger | a_k \rangle = 0 \ . \tag{6.60}$$

This requirement is fulfilled automatically by virtue of the Hermiticity of the Fockian $\hat{F}$. That is what one could expect, because the variations in (6.58) represent only a linear transformation among the occupied orbitals and do not lead, therefore, to any actual change of the wave function and of the energy.

3. Koopmans theorem

The eigenvalues of the canonical Hartree–Fock equations are usually called "orbital energies," because—to some extent—they resemble the energies of a one-electron system in an external field. This relationship is specified by the "Koopmans theorem" (T. Koopmans, 1933).

We start from the expression (6.44) for the UHF energy of a system consisting of $N = n_a + n_b$ electrons— we shall denote it as E_N—and separate in its expression those terms that contain a selected orbital a_k. We get

$$\begin{aligned} E_N &= \sum_{\substack{i=1\\(i\neq k)}}^{n_a} \langle a_i|\hat{h}|a_i\rangle + \langle a_k|\hat{h}|a_k\rangle + \sum_{i=1}^{n_b}\langle b_i|\hat{h}|b_i\rangle \\ &+ \frac{1}{2}\sum_{\substack{i,j=1\\(i,j\neq k)}}^{n_a} ([a_ia_j|a_ia_j] - [a_ia_j|a_ja_i]) + \frac{1}{2}\sum_{j=1}^{n_a}([a_ka_j|a_ka_j] - [a_ka_j|a_ja_k]) \\ &+ \frac{1}{2}\sum_{i=1}^{n_a}([a_ia_k|a_ia_k] - [a_ia_k|a_ka_i]) + \frac{1}{2}\sum_{j=1}^{n_a}([b_ib_j|b_ib_j] - [b_ib_j|b_jb_i]) \\ &+ \sum_{\substack{i=1\\(i\neq k)}}^{n_a}\sum_{j=1}^{n_b}[a_ib_j|a_ib_j] + \sum_{j=1}^{n_b}[a_kb_j|a_kb_j] \,. \end{aligned} \tag{6.61}$$

One can easily see that the terms that do not contain a_k represent the UHF energy $E_{N-1}(a_k)$ of the determinant wave function for $N-1$ electrons, which one obtains by omitting orbital a_k from the wave function (6.21). Owing to the identities $[ab|cd] \equiv [ba|dc]$, $[a_ka_j|a_ka_j] \equiv \langle a_k|\hat{J}_j^a|a_k\rangle$, $[a_ka_j|a_ja_k] \equiv \langle a_k|\hat{K}_j^a|a_k\rangle$ etc., one can transform the terms containing a_k and write

$$\begin{aligned} E_N &= E_{N-1}(a_k) + \langle a_k|\hat{h}|a_k\rangle + \sum_{j=1}^{n_a}(\langle a_k|\hat{J}_j^a|a_k\rangle - \langle a_k|\hat{K}_j^a|a_k\rangle) \\ &+ \sum_{j=1}^{n_b}\langle a_k|\hat{J}_j^b|a_k\rangle = E_{N-1}(a_k) + \underbrace{\langle a_k|\hat{F}^a|a_k\rangle}_{\varepsilon_{kk}^a} \,. \end{aligned} \tag{6.62}$$

Thus we get the *Koopmans theorem*

$$E_N = E_{N-1}(a_k) + \varepsilon_{kk}^a \,. \tag{6.63}$$

According to the result obtained, the diagonal elements of the matrix $\boldsymbol{\varepsilon}^a$ (and similarly of the matrix $\boldsymbol{\varepsilon}^b$) are equal with the energy change taking place when

deleting the electron from the given orbital. ε^a_{kk} is usually negative, indicating that one has to invest energy to ionize the system; therefore the quantity $-\varepsilon^a_{kk}$ can be considered as the approximation to the respective ionization potential.

As we have discussed, matrix $\boldsymbol{\varepsilon}^a$ gives the representation of the operator $\hat{F}^a$ in the occupied subspace. As $\boldsymbol{\varepsilon}^a$ is Hermitian, its eigenvalues have the stationary behavior discussed in Chapter 3 in connection with the linear variational problem. It is to be pointed out that not only the lowest eigenvalue represents the possible most negative ε^a_{kk} value, but also the highest eigenvalue gives the possible less negative one. (We recall that we are considering the occupied subspace only.) As the energy of the ionized species according to (6.63) is

$$E_{N-1}(a_k) = E_N - \varepsilon^a_{kk} \,, \tag{6.64}$$

the lowest energy of the ion that can be obtained by using the orbitals of the original system (i.e., without allowing "orbital relaxation" upon ionization) is that which corresponds to the possible highest (less negative) ε^a_{kk} value, which is achieved for the highest orbital energy in the canonical representation. This is the property—besides the computational convenience of using pseudo-eigenvalue equations—that gives a privileged importance to the canonical version of the Hartree–Fock equations. In addition to these mathematical aspects, one favors canonical orbitals because they usually reflect different *global* physical properties of the system well, in particular the ionization and—to a lesser extent—excitation processes. At the same time the canonical orbitals often do not correlate well with such local characteristics of the ground state as the individual chemical bonds. (The use of the localized orbitals is more adequate for that purpose.)

According to the previous discussion, the orbital energies provide us an estimate of the "vertical" ionization potentials—i.e., ionization potentials computed at a fixed geometry. Not only the highest eigenvalue but also often the lower ones reflect well the different specific ionization processes. The numerical values obtained are usually accurate enough to distinguish between different ionization processes of the same atom or molecule, as e.g., ionization from a core, σ- or π-level.

The value of the ionization potential estimated on the basis of the Koopmans theorem is always larger (in absolute value) than the so-called "ΔSCF" value that one obtains by performing independent HF calculations for the ion and for the molecule and computing the difference of the total energies. (The SCF energy of the ion accounts for orbital relaxation and is necessarily lower than the value obtained by using (6.64) corresponding to frozen orbitals.) One expects that this error may be partly compensated by electron correlation effects: the energy lowering one obtains when taking into account electron correlation is expected somewhat less for the ion in which the number of electron is less by one.

3.1. The orbital energies and the total energy

An important remark must be made: although the energy necessary to delete one electron from the system equals the respective orbital energy (taken with the opposite sign), the total electronic energy of the system (i.e., the energy necessary to delete all the electrons at a given—fixed—nuclear configuration) does *not* equal the sum of the orbital energies. The point is that the orbital energy ε_k^a includes the whole interaction energy of the electron on orbital a_k with all the other electrons: the energy $E_{N-1}(a_k)$ does not contain any terms connected with orbital a_k. Therefore, by forming the sum of the orbital energies one accounts *twice* for the electron–electron interaction energy, whereas that sum contains the kinetic energy and the energy of the electron–nuclear attraction only once. Accordingly, one may write

$$E = \sum_{i=1}^{n_a} \varepsilon_i^a + \sum_{i=1}^{n_b} \varepsilon_i^b - E_{\text{el.-el.}} \,. \tag{6.65}$$

Instead of subtracting the electron–electron energy from the sum of the orbital energies, one may also add once more the one-electron part of the energy (the expectation value of the one-electron part of the Hamiltonian) and divide the result by two:

$$E = \frac{1}{2}\left[\sum_{i=1}^{n_a}(\varepsilon_i^a + \langle a_i|\hat{h}|a_i\rangle) + \sum_{i=1}^{n_b}(\varepsilon_i^b + \langle b_i|\hat{h}|b_i\rangle)\right] \,. \tag{6.66}$$

One may substitute here the expression of the orbital energies in terms of the Fock operators:

$$E = \frac{1}{2}\left[\sum_{i=1}^{n_a}\langle a_i|\hat{F}^a + \hat{h}|a_i\rangle + \sum_{i=1}^{n_b}\langle b_i|\hat{F}^b + \hat{h}|b_i\rangle\right] \,. \tag{6.67}$$

In the actual calculations one usually applies some variant of the last equation.

The Fockian is a Hermitian operator, irrespective of whether or not it is constructed by using the converged Hartree–Fock orbitals. Accordingly, it has eigenfunctions that are also outside the subspace of the occupied orbitals—the eigenvectors of a Hermitian operator form a complete orthonormalized set. These additional eigenvectors (the "empty," "unoccupied," or "virtual") orbitals do not, of course, influence the process of building the Fock operator. Before the iteration process converges, the eigenvectors of the Fockian do not recover exactly the occupied orbitals used for its construction. Therefore, the process of iterational solution of the SCF equations may be viewed so that in every iteration step one computes some mixing between the occupied and virtual orbitals of the previous step—although usually one does not proceed explicitly in that manner.

Normally the occupied orbitals belong to the lowest orbital energies (lowest eigenvalues of the Fockian); this is often called the "Aufbau" ($\sim$ "building-up") principle, and usually one observes a considerable gap between the orbital energy of the HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital). In most cases of calculations performed for neutral systems, the orbital energy of the HOMO is negative and that of the LUMO is positive. This is connected with the "self-repulsion" term $\hat{J}_k^a a_k - \hat{K}_k^a a_k = 0$ that we have added to the Fock operator in order to get it in the form of (6.25), which is a common equation for all orbitals a_k. However, when this Fockian is applied to a *virtual* orbital a_j^v, then it will contain n_a terms $\hat{J}_k^a a_j^v - \hat{K}_k^a a_j^v \neq 0$; one can interpret this as the electron put on a virtual orbital "sees" one more electron than does an electron put on an occupied orbital. This "pushes up" the virtual orbital energies. As a consequence, if one starts the SCF procedure by using a given orbital-filling scheme (e.g., a given number of σ- and π-electrons) then, in most cases, the overall character of this scheme does not change during the iterations: the orbital that was not filled in the starting guess will have a high orbital energy because of the self-repulsion term and, consequently, it will not be filled in the next iteration steps either. Choosing different types of starting guesses, one can in some cases arrive at different converged SCF solutions approximating different exact electronic states of the system. (Usually this happens for systems of high symmetry.)

In some cases one observes divergence in the SCF procedure—e.g., such that a pair of occupied and virtual orbitals is interchanged in each subsequent cycle. In such cases it may be useful to introduce an additional "level shift" increasing artificially the orbital energies of the virtual orbitals. This can be achieved by adding a term like $\Delta\varepsilon(1-\hat{P}_{occ})$ to the Fockian, with $\Delta\varepsilon$ being the desired orbital energy shift and $\hat{P}_{occ}$ being the operator of projection on the occupied subspace. It is easy to see that in the converged case the introduction of such a term does not influence anything except the orbital energies attributed to the empty virtual orbitals, lacking any real significance. Simple qualitative considerations of perturbational character indicate that if one increases the difference between the orbital energies of the occupied and virtual orbitals then it will decrease their "mixing" (or prevent their interchange) in the next iteration step and "damp" accordingly the oscillators of the iteration procedure.

Using a sufficiently large but finite level shift, one can force every SCF problem to converge to a solution of the HF equation, which may or may not be the solution with the lowest total energy. Because of the presence of the self-repulsion term in the Fock operator, most solutions will obey the Aufbau principle after the convergence is achieved, i.e., the occupied orbitals will have the lowest orbital energies even without using an energy shift, and the latter can be "switched off" as the convergence is approached. (This cannot be proved in

general, however.) Of course, if the SCF procedure is convergent without the use of the level shift, then the use of the latter will slow down the iteration process. It is to be noted that there are numerous other interpolation and extrapolation schemes devoted to facilitate or speed the convergence in the SCF calculations.

As noted previously, the inclusion of the self-repulsion term in the Fock operator implies that the virtual orbital energies contain the interaction with one more electron. Therefore the virtual orbitals correspond to some states of a negative ion, rather than to the excited states of the original system. Accordingly, the virtual orbital energies could be related to the electron affinity of the system. Proper description of negative ions, however, is usually a more delicate problem than that of positive ions and neutral species. In most cases one cannot get actual bonded states of negative ions without accounting for the orbital relaxation; usually explicit consideration of the electron correlation is necessary, as well.

Taking properly into account the self-repulsion term, one can use the orbital energies also for getting a crude estimate of the excitation energy. As discussed previously, the effective potential determining the orbital energy of the virtual orbital (say a_j^v) contains the field which is the result of the electron occupying the orbital (say a_k^o) from which the excitation took place; therefore the excitation energy is not $\varepsilon_j^{av} - \varepsilon_k^{ao}$, but one has to subtract from this difference the self-repulsion term $\langle a_j^v | \hat{J}_k^a - \hat{K}_k^a | a_j^v \rangle = [a_j^v a_k^o | a_j^v a_k^o] - [a_j^v a_k^o | a_k^o a_j^v]$. (This formula is pertinent to the single-determinant UHF case. In the RHF case the situation is similar, but the formula is different, because one usually considers two-determinant spin-adapted excited states—see Section 4.4.)

Owing to the inadequacy of the virtual canonical orbitals to describe excited states, one usually does not treat excited states at the single determinant level. The simplest method that can be used reasonably well for this purpose is the so-called "all singles CI", i.e., solving the linear variational problem for all the determinant wave functions that can be obtained by a single excitation from a Hartree–Fock ground state. (Sometimes only a part of such configurations is considered, omitting excitations that involve the lowest occupied orbitals and/or the highest virtual ones; that means that one performs a preselection based on the orbital energy values.)

Comparisons of individual orbital energies are often used in qualitative considerations of chemical nature without proper attention always being paid to the difference between the sum of orbital energies and the total electronic energy. Such considerations can be justified to some extent by referring to the result of (6.63) in those cases when the states of all electrons but one do not change appreciably during the process considered; in that case the total energy changes more or less parallel the single orbital energy undergoing a significant change.

It may be mentioned in this respect that it is possible to construct such effective (SCF) operators, the diagonal matrix elements of which within the occupied subspaces sum to the total electronic energy. It is trivial to check that such an operator can be obtained by introducing a factor of $\frac{1}{2}$ to every Coulomb operator and exchange operator in the Fockian (6.25). Using such operators, one can present the Hartree–Fock total electronic energy as a sum of one-electron contributions (increments). If one would use this operator in an SCF calculation of canonical type, then the total energy would be equal to the sum of the orbital energies—but, of course, such a procedure would converge to a wave function with an energy higher than the HF one. Nevertheless, one can imagine such an operator as giving some theoretical background to many qualitative paper-and-pencil considerations of practical chemists.

4. The RHF method

4.1. The RHF and UHF schemes

The most common version of the Hartree–Fock method for treating singlet states of closed-shell systems is the "restricted" HF (RHF) method using doubly filled spatial orbitals. The wave function accordingly is

$$\Psi = \hat{\mathcal{A}}[\varphi_1(1)\alpha(1)\varphi_1(2)\beta(2)\ldots\varphi_n(2n-1)\alpha(2n-1)\varphi_n(2n)\beta(2n)] \quad (6.68)$$

(In this section we consider only the case of an even number N=2n of electrons, with an equal number n of spins α and β.)

All the considerations given previously concerning the properties and transformations of the one-electron orbitals, the invariance of the determinant functions, etc., hold with no changes in this case, as well. Also, the equivalence of the variation principle ($\delta E = 0$) with the Brillouin theorem can trivially be established for the present case. In fact, the introduction of the double occupancy requirement causes a single difference as compared with the derivation given in Sections 1 and 2: in the RHF case we have to consider simultaneously the variation of both spin orbitals that are derived from a given spatial orbital φ_i. Accordingly, the variation of orbital φ_i as $\varphi_i \to \varphi_i + \delta\varphi_i = \varphi_i + \eta\varphi_i'$ leads to a first-order variation of the determinant Ψ that is a sum of two terms:

$$\delta\Psi = \eta\Psi_1(\varphi_i\alpha \to \varphi_i'\alpha) + \eta\Psi_1(\varphi_i\beta \to \varphi_i'\beta)\ . \quad (6.69)$$

However, the matrix elements between determinant Ψ and the two singly excited determinants on the right-hand side of (6.69) are the same, as they differ only by interchanging spins α and β. Therefore, the presence of two terms in (6.69) leads only to a factor of 2 on the left-hand side of the equation $\langle\delta\Psi|\hat{H} - E|\Psi\rangle = 0$, giving the variation condition (c.f. Section 1.3 of Chapter 2). Thus the proof of the Brillouin theorem and the derivation of the HF equations by expanding the latter in terms of the one-electron orbitals remain

exactly the same as were considered previously for the UHF case. Accordingly, one obtains the RHF equations from the UHF ones by replacing everywhere $a_i = b_i = \varphi_i$; $n_a = n_b = n$; $\varepsilon_i^a = \varepsilon_i^b = \varepsilon_i$, etc. Thus one obtains from (6.42) the canonical RHF equations as

$$\hat{F}\varphi_i = \varepsilon_i \varphi_i \qquad (i = 1, 2, \ldots, n) \ . \tag{6.70}$$

One can omit the superscripts a and b also in the definitions of the Coulomb operators and exchange operators and write

$$\hat{F} = \hat{h} + \sum_{j=1}^{n} (2\hat{J}_j - \hat{K}_j) \ . \tag{6.71}$$

Obviously, one obtains the same RHF equations starting from the UHF equations for orbitals a_i and from those for orbitals b_i: the number of different occupied spatial orbitals and thus the number of equations is halved as compared with the UHF case. The RHF equations represent a special case of the UHF ones and, therefore, the solutions of the RHF equations always are solutions of the UHF ones, as well. Often the UHF solution coincides with the RHF one—and this mostly happens near the equilibrium geometry of the closed-shell molecules. In such cases the RHF wave function represents a minimum not only among the determinant wave function built up of doubly filled orbitals but also among all single-determinant wave functions. In these cases one cannot get any energy lowering by admitting the "splitting" of the orbitals, i.e., by the use of different occupied subspaces spanned by the orbitals filled with spins α and β, respectively.

In other cases—e.g., at large interatomic separations—there exists a UHF solution with an energy lower than the RHF one. (More precisely, the genuine UHF solutions always occur in pairs having strictly equal energies and differing from each other by interchanging all spins α and β.) In such cases the RHF solution represents a minimum only among the determinants with doubly filled orbitals but it is only a saddle point for the more general UHF determinants having no double-occupancy restriction. This means that the RHF wave function is a minimum for the case of doubly occupied orbitals but represents a maximum with respect of some changes of orbital coefficients that lead to a "splitting" of the orbitals and result in different orbitals occupied with spins α and β. (For that reason the UHF method is often also called the "different orbitals for different spin," or DODS, scheme.) If there exists a UHF solution with an energy lower than the closed-shell RHF one, then one often speaks about the "triplet instability" of the latter. This nomenclature is connected with the fact that a DODS determinant is not a pure singlet eigenfunction of the total spin-square operator $\hat{S}^2$, but also contains some admixture of triplet (etc.) components.

The UHF wave function contains twice as many freely optimized one-electron spatial orbitals as does the RHF one, so it represents a much more

flexible variational trial function. The question of whether or not one eventually obtains an orbital splitting is determined by the relative importance of two effects of opposite character. From one side, the splitting of the orbitals in the UHF wave function permits an energy lowering because the average distance between the electrons increases as the electrons are not forced to occupy by pairs the same spatial orbitals. (This means that the UHF method can account for *some* electron correlation.) From the other side, there is an increase of energy because the average electron–nuclear distances also increase to some extent and/or an increase of the kinetic energy takes place.

4.2. Symmetry and the RHF method

In most cases of symmetric molecules the RHF wave functions automatically correspond to the spatial symmetry of the system. That means that the symmetry operations (e.g., reflection in a symmetry plane or rotation around an axis of symmetry) leave invariant the molecular RHF wave function. In such cases the canonical molecular orbitals are also symmetry adapted. The non-degenerate orbitals remain invariant or change their phases under symmetry operations whereas degenerate orbitals transform between themselves.

It would be, however, a mistake to assume that the Slater determinant wave function built up of doubly filled orbitals is *always* symmetry adapted. That can be guaranteed only in the case when the starting guess of the orbitals was selected in the manner as to provide the Fockian also to reflect the full symmetry of the system, i.e., to commute with all the symmetry operations of the point group in question. Then its eigenfunctions—the canonical orbitals—can also be classified according to their symmetry properties. This behavior is conserved during the consecutive iterations if we are keeping this symmetry-adapted way of constructing the Fock operator. That requires each degenerate set of orbitals to be either completely filled or to be left completely unoccupied.

It may happen that for a given system there exists a wave function constructed by using doubly filled orbitals that does not have the full symmetry of the system but has an energy that is lower than that of the symmetry-adapted determinant. In this case these wave functions are called symmetry-unrestricted and symmetry-restricted HF functions, respectively, and one speaks about "singlet instability" of the latter.

Special attention must be paid to those points in which the symmetry of the system is higher than in their environment. As Musher had pointed out, by forcing the full symmetry in these isolated points, one can obtain a point-like discontinuity of the calculated potential energy surface. (An example of such a behavior is the cyclobutadiene molecule at a conformation of a regular square.)

Using group-theoretical language, the previous discussion can briefly be summarized as follows. If the Fockian corresponds to a wave function built up by "filling whole irreducible representations" (i.e., the functions forming

a basis of an irreducible representation either are all doubly filled or are all left unoccupied) then the Fockian will commute with the symmetry operations of the point group and its eigenfunctions can be classified according to the irreducible representations. In such a case the overall RHF wave function also provides a fully symmetric one-dimensional representation of the point group. The crucial point is the proper symmetry of the initial guess, as the symmetry properties of the initial guess are usually conserved during the iterations.

4.3. The "dissociation catastrophe"; different Hartree–Fock methods

The RHF method reflects well the main features characterizing the electronic structures of stable molecules; it does, however, completely fail when the "homolytical" dissociation of a closed-shell molecule into two open-shell fragments is considered. The point is that the use of doubly filled orbitals does not allow the description of two separated odd-electron atoms. Upon analyzing the RHF wave function applied in the case of large interatomic separation, one establishes that it represents an equal mixture of components corresponding to the neutral species and to pairs of ions in which one electron is transferred either to one atom or to another. (In practice the standard RHF procedure usually fails to converge in most cases of large interatomic separation but, at least in principle, the RHF solution exists and can be found by using symmetry considerations or other means.)

Example

Let us consider the H_2 molecule at large interatomic separations. The RHF wave function is

$$\Psi = \hat{\mathcal{A}}[\varphi(1)\alpha(1)\varphi(2)\beta(2)] \ . \tag{6.72}$$

The orbital φ is symmetric with respect to the two atoms, so it can be written as

$$\varphi = \mathcal{N}[\chi_A + \chi_B] \ , \tag{6.73}$$

where functions χ_a and χ_B are connected by a reflection, and each of them is located around the respective nucleus. If both are normalized and are orthogonal to each other, i.e., $\langle\chi_A|\chi_A\rangle = \langle\chi_B|\chi_B\rangle = 1$, and $\langle\chi_A|\chi_B\rangle = 0$, then the normalization coefficient $\mathcal{N}$ equals $\frac{1}{\sqrt{2}}$. Thus

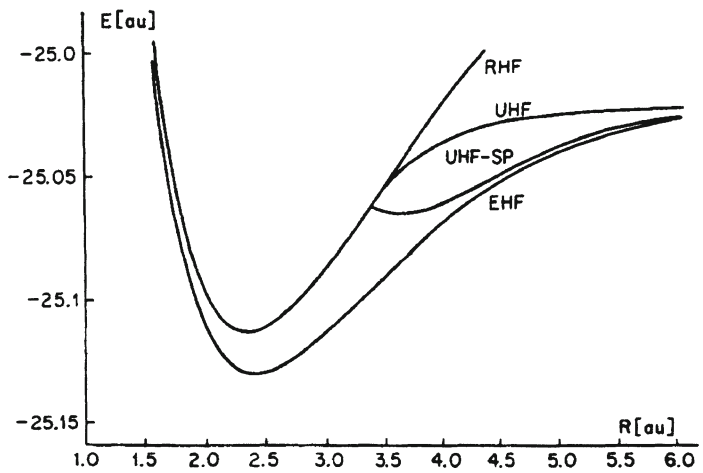
$$\Psi = \frac{1}{2}\hat{\mathcal{A}}\{[\chi_A(1) + \chi_B(1)]\alpha(1)[\chi_A(2) + \chi_B(2)]\beta(2)\} \ . \tag{6.74}$$

This wave function represents a determinant, each column of which is a sum of two terms; it can therefore be written as a sum of four determinants:

$$\Psi = \frac{1}{2}\Big\{\hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_A(2)\beta(2)] + \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_B(2)\beta(2)]$$
$$+ \hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_B(2)\beta(2)] + \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_A(2)\beta(2)]\Big\} \ . \quad (6.75)$$

The first two terms of this sum represent ionic structures corresponding to the situation in which two electrons are around the same atom. The third and fourth terms correspond to the neutral atoms; they are usually called "covalent structures." (In this context the term "covalent" simply means "nonionic.")

As the energy of the ionic terms is very high (two electrons on the same atom exhibit a large repulsion), the RHF energy obtained at large distances is much higher than the sum of the energies of the neutral free atoms. In such systems the RHF method always exhibits a "triplet instability" at distances, larger than a "critical" one, i.e., at sufficiently large interatomic separations there exists a UHF solution that has an energy lower than the RHF one and describes dissociation into neutral fragments.



Potential curves of the BH molecule obtained by using a double zeta Gaussian lobe basis set[1].

The Figure displays the typical behavior of the potential energy curves given by the different SCF schemes. The RHF method predicts the position of the minimum of the curve (the equilibrium interatomic distance) with an acceptable accuracy, whereas overestimates its curvature by some 10–20%. (Accordingly,

[1]Reproduced from my paper "Spin-Projected EHF Method. IV. Comparison of Potential Curves Given by Different One-Electron Methods", I. Mayer, Int. J. Quantum Chem. **14**, 29 (1978); Copyright 1978 Wiley Periodicals.

the RHF vibrational frequencies corresponding to bond stretching should be systematically "scaled down.") At larger interatomic distances the RHF curve sharply increases, owing to the presence of the ionic terms in the wave function. The UHF method usually does not give a solution differing from the RHF one at the equilibrium distance or even somewhat larger ones. When the interatomic distance is further increased, there is a "critical point" (bifurcation point) in which the distinct, genuine UHF solution suddenly appears and the UHF curve quite smoothly departs from the RHF one. In the critical point the UHF and RHF solutions coincide. As the gradient of the SCF energy is determined by the orbitals themselves (c.f. Section 2.1 of Chapter 2 and Section 7), the first derivatives of the UHF and RHF curves also coincide at the point of the separation. As the UHF wave function represents a new type of solution, the second derivative of the UHF curve differs from that of the RHF one. After departing from the RHF solution, the UHF curve tends to a qualitatively correct dissociation limit; one usually observes that it approaches the asymptotic energy value somewhat too quickly.

One may consider the UHF curve as coinciding with the RHF one at distances smaller than the critical one. Then the UHF curve is continuous and continuously differentiable but exhibits a discontinuity of the second and higher derivatives at the critical point. (As pointed out by Pulay, there must be a discontinuity on either the RHF or the UHF curves, because they coincide at and around the equilibrium distance. If both curves were continuous everywhere with continuous derivatives of any order, then they would have a common Taylor series and would coincide everywhere.)

As the RHF wave function is a special case of the UHF solutions, the (first) variation of the energy as a functional of orbitals vanishes, i.e., the energy is stationary. The question of the existence of a specific UHF solution can be studied by considering the second variations. If it is positive definite even admitting "orbital splitting", then the RHF solution is a true minimum even for the manifold of UHF wave functions. If it is not positive definite, then one can find a UHF solution with an energy lower than the RHF one. This means that the critical point in which the specific UHF solution appears is the point in which the second variation of the energy becomes positive *semi*definite. If, as it is usual in practice, one applies a finite basis set to expand the orbitals, then one can compute the Hessian matrix (matrix of the second derivatives of the energy as functions of the orbital coefficients) and diagonalize it. The point in which the lowest eigenvalue of the Hessian matrix crosses the zero level is the bifurcation point in which an SCF solution of a new kind appears. This permits, at least in principle, to localize precisely the position of the critical point.

The closed-shell RHF wave function built up of doubly filled orbitals is automatically a pure singlet, i.e., it is an eigenfunction of operators $\hat{S}_z$ and

$\hat{S}^2$ with eigenvalues 0. This can be seen trivially for $\hat{S}_z = \sum_{i=1}^{N} \hat{s}_z(i)$, because the spin functions $\alpha(i)$ and $\beta(i)$ are eigenfunctions of the operator $\hat{s}_z(i)$ with eigenvalues $+\frac{1}{2}$ and $-\frac{1}{2}$, respectively. As the number of α and β spin functions is equal in a closed-shell RHF determinant, the latter is an eigenfunction of $\hat{S}_z$ with an eigenvalue 0.

The closed-shell RHF determinant wave function is an eigenfunction of operators $\hat{S}_x = \sum_{i=1}^{N} \hat{s}_x(i)$ and $\hat{S}_y = \sum_{i=1}^{N} \hat{s}_y(i)$, as well, also with eigenvalues 0. (The case $S_x = S_y = S_z = 0$ is exceptional in the sense that all the three components of the spin angular momentum can be measured simultaneously.) To see that the closed-shell determinant is an eigenfunction of $\hat{S}_x$ and $\hat{S}_y$, first one should observe that $\hat{S}_x$ and $\hat{S}_y$ are symmetric with respect to the electrons and, therefore, commute with the antisymmetrizer. So, for instance, we have for $\hat{S}_x$, denoting by n the number of doubly occupied orbitals ($N = 2n$):

$$\begin{aligned}
\hat{S}_x\Psi &= \sum_{i=1}^{N} \hat{s}_x(i)\hat{\mathcal{A}}[\varphi_1(1)\alpha(1)\varphi_1(2)\beta(2)\ldots \\
&\qquad\qquad \times\psi_n(2n-1)\alpha(2n-1)\varphi_n(2n)\beta(2n)] \\
&= \sum_{i=1}^{N} \hat{\mathcal{A}}\{\hat{s}_x(i)[\varphi_1(1)\alpha(1)\varphi_1(2)\beta(2)\ldots \qquad (6.76) \\
&\qquad\qquad \times\psi_n(2n-1)\alpha(2n-1)\varphi_n(2n)\beta(2n)]\} \,.
\end{aligned}$$

Now, according to (5.6) and (5.7), operator $\hat{s}_x(i)$ "flips" the spin function $\gamma(i)$, i.e., if it was α then becomes $\frac{1}{2}\beta$; if it was β then becomes $\frac{1}{2}\alpha$. This may be denoted as $\hat{s}_x(i)\gamma(i) = \frac{1}{2}\bar{\gamma}(i)$, where $\bar{\gamma}(i)$ is the spin function directed oppositely to $\gamma(i)$. Originally in Ψ we have the same spatial orbital filled with both spins $\gamma(i)$ and $\bar{\gamma}(i)$; after applying $\hat{s}_x(i)$ it becomes twice filled with spin $\bar{\gamma}(i)$. This means that each $\hat{s}_x(i)$ produces a determinant with two identical columns, which vanishes according to the properties of determinants. Therefore, the closed-shell RHF determinant Ψ is an eigenfunction of $\hat{S}_x$ with an eigenvalue 0. Similar considerations hold for operator $\hat{S}_y$, as well. Accordingly, Ψ is an eigenfunction also of $\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2$ with an eigenvalue 0.
Q.E.D.

The single DODS determinant UHF wave function is an eigenfunction of operator $\hat{S}_z$ but not of $\hat{S}^2$. This means that one cannot assign a definite spin multiplicity to it, but it represents a mixture of different singlet, triplet, etc., components. (One often uses the term "spin contamination" when the wave function contains unwanted spin components.)

In order to solve this problem, Löwdin introduced his "spin projection operator"

$$\hat{O}^S = \prod_{l \neq S} \frac{\hat{S}^2 - l(l+1)}{S(S+1) - l(l+1)} \tag{6.77}$$

permitting to select from a wave function that pure spin component of it that belongs to the desired eigenvalue $S(S+1)$ of the operator S^2. (In accord with the general rules pertinent to the angular momentum in quantum mechanics, the selected state has the spin multiplicity $2S+1$.) The spin projection operator annihilates all unwanted spin components, leaving unchanged the required one. In fact, any wave function Ψ can be presented as a sum $\Psi = \sum_l {}^l\Psi$ of terms ${}^l\Psi$ with different spin multiplicities $2l+1$. The term $\hat{S}^2 - l(l+1)$ in the operator $\hat{O}^S$ annihilates the component ${}^l\Psi$:

$$[\hat{S}^2 - l(l+1)]\,{}^l\Psi = l(l+1)\,{}^l\Psi - l(l+1)\,{}^l\Psi = 0\,. \tag{6.78}$$

All components ${}^l\Psi$ with $l \neq S$ are subsequently annihilated. When the operator $\hat{S}^2 - l(l+1)$ acts on the term ${}^S\Psi$ with the desired multiplicity $2S+1$, then we get

$$[\hat{S}^2 - l(l+1)]\,{}^S\Psi = [S(S+1) - l(l+1)]\,{}^S\Psi]\,. \tag{6.79}$$

The denominators in (6.77) are introduced to compensate this coefficient; the component ${}^S\Psi$ therefore remains unchanged, while one gets rid of all the other components ${}^l\Psi$.

The spin-projected DODS determinant is not a single determinant anymore, which makes its use tedious. By subjecting the UHF determinant to a spin projection, not only does one get a pure spin function but usually one obtains also an energy lowering with respect to the UHF energy—of course, only in those points where there exists a specific UHF solution differing from the RHF wave function representing a pure singlet. The Figure shown previously illustrates that this can lead to quite pathological potential curves. The source of this problem is that in this case one had optimized variationally the single-determinant wave function *before* the spin projection had been applied and not the pure spin component produced by the spin projection ("UHF method with subsequent spin projection"). It is much more consistent to consider the spin-projected determinant itself as a variational trial wave function ("variation after projection"). This procedure leads to the spin-projected "extended" Hartree–Fock (EHF) method; as the Figure also shows, it is able to give a well-balanced description of the homolytic dissociation of *one* chemical bond.

The special property of the spin-projected wave functions is that they are particularly well suited to describe antiferromagnetic and antiaromatic systems

that can be considered as containing alternating spins forming two sublattices of the highest possible spin that are coupled into a resulting singlet.

The RHF wave function again represents a special solution of the EHF equations but, contrary to the UHF case, an EHF solution with an energy lower than the RHF one exists everywhere. The EHF method can still be considered as a generalized one-electron model because there are as many spin orbitals in the starting DODS determinant as the number of electrons. Its serious disadvantage—besides the rather complex equations and high computation cost—that it is not size consistent and, as a consequence, cannot give a similarly good description of a system with several chemical bonds as given for a single bond.

As the EHF scheme is rather complex, two simplified versions of it were also introduced; these still can be encountered. One of them is the "simple annihilation" method in which one applies only one term of the product in (6.77), corresponding to the lowest—and in most cases largest—unwanted spin component (usually the triplet). Another is the so-called "half-projected Hartree–Fock method" in which the wave function is a sum or difference of only two determinants that differ from each other by interchanging all spin functions α and β. It contains only each second spin component (for instance singlet, quintet, etc., but not triplet and septuplet, etc.) and usually represents an almost pure spin state.

The difficulties connected with the spin eigenfunctions represent the practically most important, but not the only, example of the problem that Löwdin called the "symmetry dilemma." He pointed out that the exact solutions of the Schrödinger equations are automatically eigenfunctions of all operators commuting with the Hamiltonian, i.e., are automatically symmetry adapted—e.g., are pure spin eigenfunctions. (Exceptions are some special cases of degeneracy, like that of the singlet and triplet wave functions of the H_2 system at an infinite internuclear separation. Even in such cases, however, the eigenfunctions may be chosen to be symmetry adapted.) At the same time, this is not the case for the approximate wave functions determined by using the variation principle. Requesting a symmetry property to be fulfilled reduces the flexibility of the trial wave function used in the variational procedure, and this can cause an increase of the energy as compared with the wave function obtained without using the symmetry restriction. As a consequence, one is faced with the symmetry dilemma of choosing between the wave function of lower energy and that of the correct symmetry properties.

4.4. Singlet and triplet excitations

The considerations concerning the energy of a singly excited determinant, given in Section 3.1 are valid in the RHF case, as well. However, as the RHF ground state is a pure singlet, it is meaningful to discuss how one can obtain

pure singlets and triplets by considering single excitations from the RHF wave function.

If one considers the excitation of one electron from the doubly occupied spatial orbital φ_i to the virtual spatial orbital φ_v (without the change of the overall spin projection S_z) then one gets two different singly excited determinants depending on whether the excitation involves an electron of spin α or spin β:

$$\begin{aligned}
\Psi_1^{iv,\alpha} &= \Psi_1(\varphi_i\alpha \to \varphi_v\alpha) \\
&= \hat{\mathcal{A}}[\varphi_1(1)\alpha(1)\varphi_1(2)\beta(2)\ldots\varphi_v(2i-1)\alpha(2i-1)\varphi_i(2i)\beta(2i)\ldots] \\
\Psi_1^{iv,\beta} &= \Psi_1(\varphi_i\beta \to \varphi_v\beta) \\
&= \hat{\mathcal{A}}[\varphi_1(1)\alpha(1)\varphi_1(2)\beta(2)\ldots\varphi_i(2i-1)\alpha(2i-1)\varphi_v(2i)\beta(2i)\ldots] \,.
\end{aligned} \tag{6.80}$$

Neither of these determinant is a pure spin function, but their normalized sum and difference are:

$$^1\Psi = \frac{1}{\sqrt{2}}(\Psi_1^{iv,\alpha} + \Psi_1^{iv,\beta}) \tag{6.81}$$

is a singlet and

$$^3\Psi = \frac{1}{\sqrt{2}}(\Psi_1^{iv,\alpha} - \Psi_1^{iv,\beta}) \tag{6.82}$$

is a triplet. These formulae are quite similar to Eq. (5.19), as the doubly filled orbitals form singlet shells which need not to be considered when discussing spin functions. Some minor remarks, however, are appropriate. In (6.81), (6.82) there are—somewhat unusually—sign "+" at the singlet and sign "−" at the triplet. This is because in the two determinants in (6.80) the order of the spin functions is the same and not that of the spatial orbitals as in (5.19). If one changes the order of the orbitals in $\Psi_1^{iv,\beta}$, then one gets a change of the sign and arrives at the more conventional form with $\ldots\varphi_v(2i-1)\varphi_i(2i)\frac{1}{\sqrt{2}}$ $[\alpha(2i-1)\beta(2i) - \beta(2i-1)\alpha(2i)]\ldots$ for $^1\Psi$ and similarly at a form of type $\frac{1}{\sqrt{2}}(\alpha\beta + \beta\alpha)$ for $^3\Psi$. It is to be noted that the $S_z = \pm 1$ components of the triplet can be written in the form of a single determinant—c.f. Eq. (5.18). We recall also that the factor $\frac{1}{2}$ in Eq. (5.19) contains also the $\frac{1}{\sqrt{2!}}$ coming from the definition of the antisymmetrizer.

Let us determine the energy of these wave functions:

$$E_{S,T} = \langle \frac{1}{\sqrt{2}}(\Psi_1^{iv,\alpha} \pm \Psi_1^{iv,\beta})|\hat{H}|\frac{1}{\sqrt{2}}(\Psi_1^{iv,\alpha} \pm \Psi_1^{iv,\beta})\rangle \,. \tag{6.83}$$

The upper signs correspond to the singlet, the lower ones to the triplet.

Expanding (6.83) we get

$$\begin{aligned} E_{S,T} &= \frac{1}{2}\Big(\langle\Psi_1^{iv,\alpha}|\hat{H}|\Psi_1^{iv,\alpha}\rangle + \langle\Psi_1^{iv,\beta}|\hat{H}|\Psi_1^{iv,\beta}\rangle \\ &\quad \pm \langle\Psi_1^{iv,\beta}|\hat{H}|\Psi_1^{iv,\alpha}\rangle \pm \langle\Psi_1^{iv,\alpha}|\hat{H}|\Psi_1^{iv,\beta}\rangle\Big) \ . \end{aligned} \tag{6.84}$$

The first two terms are equal; they differ from each other only by interchanging spins α and β. Each of these terms represents the energy of the singly excited single determinant $\Psi_1^{iv,\alpha}$ or $\Psi_1^{iv,\beta}$. According to the discussion in Section 3.1, the energy of such a determinant differs from the Hartree–Fock ground-state energy E_0 by the difference of the orbital energies and by a correction for the self-repulsion term $\hat{J}_i - \hat{K}_i$ present in the Fock operator:

$$\begin{aligned} \langle\Psi_1^{iv,\alpha}|\hat{H}|\Psi_1^{iv,\alpha}\rangle &= \langle\Psi_1^{iv,\beta}|\hat{H}|\Psi_1^{iv,\beta}\rangle = E_0 + \varepsilon_v - \varepsilon_i - \langle\varphi_v|\hat{J}_i - \hat{K}_i|\varphi_v\rangle \\ &= E_0 + \varepsilon_v - \varepsilon_i - J_{iv} + K_{iv} \ . \end{aligned} \tag{6.85}$$

Here $J_{iv} = [\varphi_i\varphi_v|\varphi_i\varphi_v]$ and $K_{iv} = [\varphi_i\varphi_v|\varphi_v\varphi_i]$ are the Coulomb and exchange integrals for the pair of orbitals φ_i and φ_v.

The third and fourth terms of (6.84) are complex conjugates; if real orbitals are used, they are also equal. The term $\langle\Psi_1^{iv,\alpha}|\hat{H}|\Psi_1^{iv,\beta}\rangle$ represents the matrix element of the Hamiltonian between determinants built up of orthonormalized orbitals and differing in two orbitals: the "bra" contains $\varphi_v\alpha\varphi_i\beta$, the "ket" $\varphi_i\alpha\varphi_v\beta$.

We use Slater rules (Section 6.3.2 of Chapter 5)

$$\begin{aligned} \langle\Psi_1^{iv,\alpha}|\hat{H}|\Psi_1^{iv,\beta}\rangle &= \langle\varphi_v(1)\alpha(1)\varphi_i(2)\beta(2)|\frac{1}{r_{12}}|\varphi_i(1)\alpha(1)\varphi_v(2)\beta(2)\rangle \\ &\quad -\langle\varphi_v(1)\alpha(1)\varphi_i(2)\beta(2)|\frac{1}{r_{12}}|\varphi_v(1)\beta(1)\varphi_i(2)\alpha(2)\rangle \ . \end{aligned} \tag{6.86}$$

The second integral vanishes, owing to the orthogonality of the spin functions. Performing the summations over the spins in the first integral, we get

$$\langle\Psi_1^{iv,\alpha}|\hat{H}|\Psi_1^{iv,\beta}\rangle = [\varphi_v\varphi_i|\varphi_i\varphi_v] = K_{iv} \ . \tag{6.87}$$

By substituting (6.85) and (6.87) into (6.84) we arrive at

$$E_{S,T} = \frac{1}{2}[2(E_0 + \varepsilon_v - \varepsilon_i - J_{iv} + K_{iv}) \pm 2K_{iv}] \ , \tag{6.88}$$

that is

$$\begin{aligned} E_S &= E_0 + \varepsilon_v - \varepsilon_i - J_{iv} + 2K_{iv} \\ E_T &= E_0 + \varepsilon_v - \varepsilon_i - J_{iv} \ . \end{aligned} \tag{6.89}$$

Determinants $\Psi_1^{iv,\alpha}$ and $\Psi_1^{iv,\beta}$ are, according to (6.81) and (6.82) one-to-one mixtures of the singlet ${}^1\Psi$ and triplet ${}^3\Psi$. In agreement with this, their energy (6.85) is the average of the singlet and triplet energies in (6.89).

The exchange integral $K_{iv} > 0$ because it represents the electrostatic self-interaction of the exchange charge density $\varphi_i(\vec{r})\varphi_v(\vec{r})$, and one gets from (6.89) $E_S > E_T$, in agreement with the general observation that the first triplet excited state of a closed-shell molecule usually has a lower energy than that of the first excited singlet.

Of course, it is better to consider not only a particular excitation from one occupied orbital to a virtual one, but to perform a linear variational ("configuration interaction"—CI) calculation between all the singly excited determinants as discussed in Section 3.1 in connection with the UHF scheme. If the RHF wave function is used as a reference, then the solutions will be automatically either singlets or triplets (except, possibly, "pathological" degenerate cases). Alternately, one can use the pure singlet and triplet configurations like those discussed here; then the calculation of the matrix elements becomes more complicated but one has to diagonalize smaller matrices—one for the singlet, another for the triplet case.

5. Finite basis Hartree–Fock theory

5.1. Hartree–Fock–Roothaan equations

The Hartree–Fock equations represent a coupled set of complex integro–differential equations. They can be solved by numerical methods for the case of free atoms, and possibly also systems like H_2. When molecules are considered, it is usual to approximate the solutions—the one-electron orbitals—in the form of linear combinations of some predetermined basis functions (basis orbitals). This is an approach quite similar to that considered in Chapter 3, but here the problem is not linear. The basis orbitals are usually defined with respect to the atoms in the molecule; earlier they were often taken from calculations performed for the free atoms; this explains why the abbreviation LCAO (linear combination of atomic orbitals) is often used for this approach.

Accordingly, we are going to use the expansion of the one-electron orbitals φ_i as

$$\varphi_i = \sum_{\mu=1}^{m} c_\mu^i \chi_\mu \; ; \qquad i = 1, 2, \ldots, m \; , \tag{6.90}$$

where the m orbitals χ_μ ($\mu = 1, 2, \ldots, m$) form the LCAO basis set, and c_μ^i is the μth orbital coefficient of the i-th molecular orbital (MO) φ_i. The coefficients of a given orbital φ_i can be collected to form a column-vector $\mathbf{c}^i$ with the elements c_μ^i. The basis functions χ_μ are usually (but not necessarily) normalized to unity; usually they are not assumed orthogonal to each other but,

of course, should be linearly independent. (In practice it may happen that the basis orbitals are nearly linearly dependent. In such case Löwdin's canonical orthogonalization described in Section 3.2 of Chapter 3 can be used to reduce the dimension of the matrix diagonalization problem.)

When one uses a finite basis set, then the integro–differential pseudo-eigenvalue equations of the HF method are replaced by generalized matrix pseudo-eigenvalue equations. These were originally derived by Roothaan — and also, independently, by Hall—(1951), and are usually called Hartree–Fock–Roothaan (HFR) equations. Roothaan has written out explicitly the total energy of the determinant wave function as a function of the orbital coefficients c^i_μ and performed the differentiations necessary for finding the stationary value of the energy under the constraints the orbitals being orthonormalized. Instead of this (somewhat lengthy) approach we shall again use the Brillouin theorem (Section 1), equivalent to the variation principle for the single determinant wave functions:

$$\langle \Psi_1(\varphi_i \to \varphi_i')|\hat{H}|\Psi\rangle = 0\,; \qquad i = 1, 2, \ldots, n\,. \tag{6.91}$$

Here $\langle \varphi_i'|\varphi_j\rangle = 0$, $j = 1, 2, \ldots, n$; i.e., φ_i' is orthogonal to all the occupied orbitals, but *it is otherwise arbitrary*. (For sake of simplicity we shall consider the RHF case; the generalization to the UHF one is trivial.)

The only principal difference with respect of the derivation given in Section 2.1 is that now all the orbitals are of the LCAO form (6.90), and this property must be conserved during the variation, as well. This requirement means that the function φ_i' must also represent a linear combination of the same basis orbitals χ_μ:

$$\varphi_i' = \sum_{\mu=1}^{m} q_\mu \chi_\mu\,. \tag{6.92}$$

The coefficients q_μ are not quite arbitrary because of the orthogonality requirements $\langle \varphi_i'|\varphi_j\rangle = 0$. We may construct the most general arbitrary orbital of type (6.92) satisfying the orthogonality requirement in the following manner. We introduce the projector on the occupied subspace

$$\hat{P} = \sum_{j=1}^{m} |\varphi_j\rangle\langle\varphi_j|\,. \tag{6.93}$$

Then $1 - \hat{P}$ is the projector on the orthogonal complement of that subspace, i.e., any function $|\varphi_i'\rangle = (1 - \hat{P})|\sum_\nu p_\nu \chi_\nu\rangle$ with arbitrary coefficients p_ν is orthogonal to all the occupied orbitals. The coefficients p_ν are arbitrary and independent of each other; we may assume, for instance, that only one of them

differs from zero at a time. Therefore, one has to request the fulfillment of the Brillouin theorem for every

$$|\varphi_i'\rangle = (1 - \hat{P})|\chi_\nu\rangle \; ; \qquad \nu = 1, 2, \ldots, m \; . \tag{6.94}$$

The expansion of the Brillouin theorem in term of the one-electron orbitals again leads to the equality

$$\langle\varphi_i'|\hat{F}|\varphi_i\rangle = 0 \; , \tag{6.95}$$

where the meaning of the Fockian $\hat{F}$ is the same as previously—of course, it is now built up of orbitals that can be expanded by using the given one-electron basis set. By taking the adjoint of $|\varphi_i'\rangle$ in (6.94) we get

$$\langle\varphi_i'| = [|\varphi_i'\rangle]^\dagger = [(1 - \hat{P})|\chi_\nu\rangle]^\dagger = \langle\chi_\nu|(1 - \hat{P}) \; . \tag{6.96}$$

We substitute (6.96) into (6.95):

$$\langle\chi_\nu|(1 - \hat{P})\hat{F}|\varphi_i\rangle = 0 \; , \tag{6.97}$$

i.e.,

$$\langle\chi_\nu|\hat{F}|\varphi_i\rangle - \langle\chi_\nu|\sum_{j=1}^{m}|\varphi_j\rangle \underbrace{\langle\varphi_j|\hat{F}|\varphi_i\rangle}_{\varepsilon_{ji}} = 0 \; . \tag{6.98}$$

We may again assume that the orbitals have already been subjected to the unitary transformation diagonalizing matrix $\boldsymbol{\varepsilon}$, and thus $\varepsilon_{ji} = \varepsilon_i\delta_{ij}$. This leads to

$$\langle\chi_\nu|\hat{F}|\varphi_i\rangle = \varepsilon_i\langle\chi_\nu|\varphi_i\rangle \; . \tag{6.99}$$

By substituting the LCAO expansion (6.90) of orbital φ_i and introducing the notations $F_{\nu\mu} = \langle\chi_\nu|\hat{F}|\chi_\mu\rangle$ and $S_{\nu\mu} = \langle\chi_\nu|\chi_\mu\rangle$, we get the HFR equations

$$\sum_{\mu=1}^{m} F_{\nu\mu}c_\mu^i = \varepsilon_i \sum_{\mu=1}^{m} S_{\nu\mu}c_\mu^i \qquad i = 1, 2, \ldots, m \; . \tag{6.100}$$

Considering $F_{\nu\mu}$ and $S_{\nu\mu}$ as elements of the $m \times m$ Fock matrix $\mathbf{F}$ and overlap matrix $\mathbf{S}$, respectively, (6.100) can be written in matrix form as

$$\mathbf{F}\mathbf{c}^i = \varepsilon_i\mathbf{S}\mathbf{c}^i \; . \tag{6.101}$$

By introducing matrix $\mathbf{C}$, the columns of which are the vectors $\mathbf{c}^i$, then (6.101) can also be written as

$$\mathbf{F}\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\varepsilon} \; , \tag{6.102}$$

where $\boldsymbol{\varepsilon} = \mathrm{diag}(\varepsilon_i)$ is the diagonal matrix of the orbital energies. One usually writes (6.102) by including in $\mathbf{C}$ the coefficients of both occupied and virtual orbitals; then both $\mathbf{C}$ and $\boldsymbol{\varepsilon}$ are $m \times m$ matrices. (One may restrict the consideration to the occupied orbitals; then $\mathbf{C}$ is a $m \times n$ rectangular matrix and $\boldsymbol{\varepsilon}$ is a $n \times n$ diagonal quadratic one.)

If the basis set were infinite and complete, then the change from the integro–differential form to the matrix form would only require the expansion of two functions (the two sides of the equation $\hat{F}\varphi_i = \varepsilon_i\varphi_i$) in the same complete basis set. It is not evident *a priori* that in the case of a finite basis set one obtains the conditions of the stationary energy simply by "cutting" this infinite set of equations and keeping only the first m equations corresponding to the basis set actually applied. However, this was always the case, even when considering more complex types of SCF equations than the HFR ones, although no general proof is known to the author.

It is worthwhile to consider briefly the latter approach, too, because it permits to obtain the HFR equations very simply—even if this *cannot be considered a true derivation.*

One substitutes expansion (6.90) into the HF equation $\hat{F}\varphi_i = \varepsilon\varphi_i$:

$$\hat{F}\sum_{\mu=1}^{m} c_\mu^i \chi_\mu = \varepsilon_i \sum_{\mu=1}^{m} c_\mu^i \chi_\mu \ , \tag{6.103}$$

multiplies with χ_ν^*, and integrates:

$$\sum_{\mu=1}^{m} \langle \chi_\nu | \hat{F} | \chi_\mu \rangle c_\mu^i = \varepsilon_i \sum_{\mu=1}^{m} \langle \chi_\nu | \chi_\mu \rangle c_\mu^i \ , \tag{6.104}$$

i.e.,

$$\sum_{\mu=1}^{m} F_{\nu\mu} c_\mu^i = \varepsilon_i \sum_{\mu=1}^{m} S_{\nu\mu} c_\mu^i \tag{6.105}$$

or, in matrix notation:

$$\mathbf{F}\mathbf{c}^i = \varepsilon_i \mathbf{S}\mathbf{c}^i \ . \tag{6.106}$$

When the basis set is increased and becomes increasingly complete, then the solutions of the HFR equation (orbitals, orbital, and total energies) tend to the so-called "Hartree–Fock limit" that one would obtain by exact solution of the integro–differential HF equations. Calculations performed by using finite but rather large basis sets are often called "near Hartree–Fock" ones.

In practical calculations one usually applies some standard sets of basis orbitals assigned to the atoms in the system; the number of basis functions adapted

for a given atom may vary in wide limits. Accordingly, the accuracy (even within the Hartree–Fock model) of the different *ab initio* calculations may also vary very much. We shall stress that the term *ab initio* does only mean that no empirical parameters—except the values of the universal constants—are applied and all integrals occurring in the calculation are computed with an adequate precision. At the present time, it is likely that the smallest and largest "standard" basis sets for the H_2 molecule are those denoted by the acronyms "STO–3G" and "aug–cc–pV5Z", consisting of only 2 ("minimal basis") and as much as 160 basis functions, respectively, and giving the SCF energies at the same interatomic distance 0.74 Å: -1.116759 and -1.133619 atomic units. (The latter value may be considered as the Hartree–Fock limit; numerical Hartree–Fock calculations can hardly provide better accuracy.)

5.2. The P-matrix

The single-determinant wave function is uniquely determined (up to a physically irrelevant constant factor) by the subspace of the occupied orbitals. That subspace can be specified by the operator of projection on it. Here we are going to consider this point in some detail for the case when the orbitals are expressed by using a finite basis set.

Let us consider the occupied orbitals φ_i; $i = 1, 2, \ldots, n$ of an RHF problem and express the operator $\hat{P}$ of projection on the occupied subspace in terms of the LCAO coefficients used in the expansion in (6.90):

$$\begin{aligned} \hat{P} &= \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i| = \sum_{i=1}^{n} |\sum_{\mu=1}^{m} c_\mu^i \chi_\mu\rangle\langle\sum_{\nu=1}^{m} c_\nu^i \chi_\nu| \\ &= \sum_{\mu,\nu=1}^{m} \sum_{i=1}^{n} c_\mu^i c_\nu^{i*} |\chi_\mu\rangle\langle\chi_\nu| = \sum_{\mu,\nu=1}^{m} P_{\mu\nu} |\chi_\mu\rangle\langle\chi_\nu| \,, \end{aligned} \tag{6.107}$$

where we have introduced the "P-matrix" having the elements

$$P_{\mu\nu} = \sum_{i=1}^{n} c_\mu^i c_\nu^{i*} \,. \tag{6.108}$$

In matrix notations matrix **P** can be expressed as a sum of dyadic products:

$$\mathbf{P} = \sum_{i=1}^{n} \mathbf{c}^i \mathbf{c}^{i\dagger} \,. \tag{6.109}$$

The Hermitian matrix **P** plays a significant role in the theory. It is generally called "density matrix" as well, although this nomenclature can be considered not quite fortunate if a nonorthogonal basis set is used.

In the case of the closed-shell RHF theory one often applies matrix P multiplied by a factor of 2; we may denote this as

$$\mathbf{D} = 2\mathbf{P} = 2\sum_{i=1}^{n} \mathbf{c}^i \mathbf{c}^{i\dagger} \ . \tag{6.110}$$

Although, as we have seen, matrix $\mathbf{P}$ is immediately related to the operator of projection on the occupied subspace, it is a true projection matrix only if the basis $\{\chi_\mu\}$ is orthonormalized. In the general case matrix $\mathbf{P}$ is Hermitian, but it is not idempotent if $S_{\mu\nu} = \langle\chi_\mu|\chi_\nu\rangle \neq \delta_{\mu\nu}$. In that case it is the matrix $\mathbf{PS}$ that has the correct idempotency property, i.e., $(\mathbf{PS})^2 = \mathbf{PS}$. In fact

$$\begin{aligned}(\mathbf{PS})^2 &= \mathbf{PSPS} = \sum_{i=1}^{n} \mathbf{c}^i \mathbf{c}^{i\dagger}\mathbf{S} \sum_{j=1}^{n} \mathbf{c}^j \mathbf{c}^{j\dagger}\mathbf{S} \\ &= \sum_{i,j=1}^{n} \mathbf{c}^i \underbrace{\mathbf{c}^{i\dagger}\mathbf{S}\mathbf{c}^j}_{\delta_{ij}} \mathbf{c}^{j\dagger}\mathbf{S} = \sum_{i=1}^{n} \mathbf{c}^i \mathbf{c}^{i\dagger}\mathbf{S} = \mathbf{PS} \ .\end{aligned} \tag{6.111}$$

Q.E.D.

In (6.111) we have utilized the fact that the occupied orbitals are assumed orthonormalized, $\langle\varphi_i|\varphi_j\rangle = \delta_{ij}$:

$$\langle\varphi_i|\varphi_j\rangle = \langle\sum_{\mu=1}^{m} c_\mu^i \chi_\mu| \sum_{\nu=1}^{m} c_\nu^j \chi_\nu\rangle = \sum_{\mu,\nu=1}^{m} c_\mu^{i*} S_{\mu\nu} c_\nu^j = \mathbf{c}^{i\dagger}\mathbf{S}\mathbf{c}^j = \delta_{ij} \ . \tag{6.112}$$

Often the idempotency property in (6.111) is expressed by writing $\mathbf{PSP} = \mathbf{P}$.

In the finite basis theory, one-electron orbitals are represented by the column-vectors of their orbital coefficients. In that framework, matrix $\mathbf{PS}$ is the matrix of projection on the occupied subspace, in agreement with (6.111). In fact, let $|\psi\rangle = \sum_{\varrho=1}^{m} q_\varrho |\chi_\varrho\rangle$ be an arbitrary one-electron orbital and let us look for the LCAO coefficients p_μ of the orbital $\hat{P}|\psi\rangle = \sum_{\mu=1}^{m} p_\mu |\chi_\mu\rangle$ by using the expansion (6.107):

$$\begin{aligned}\sum_{\mu=1}^{m} p_\mu |\chi_\mu\rangle &\equiv \hat{P}|\psi\rangle = \sum_{\mu,\nu=1}^{m} P_{\mu\nu}|\chi_\mu\rangle\langle\chi_\nu| \sum_{\varrho=1}^{m} q_\varrho \chi_\varrho\rangle \\ &= \sum_{\mu,\nu,\varrho=1}^{m} |\chi_\mu\rangle P_{\mu\nu} S_{\nu\varrho} q_\varrho = \sum_{\mu,\varrho=1}^{m} (\mathbf{PS})_{\mu\varrho} q_\varrho |\chi_\mu\rangle \ ,\end{aligned} \tag{6.113}$$

i.e., in matrix notations:

$$\mathbf{p} = \mathbf{PSq} \,. \tag{6.114}$$

Q.E.D.

In the case of an orthonormalized basis, matrix **S** is equal to the unit matrix and matrix **P** itself is the idempotent projection matrix.

In the UHF case one has to introduce two P-matrices, one for the orbitals of spin α, another for orbitals of spin β. Assuming that the expansion coefficients of the orbitals $|a_i\rangle$ and $|b_i\rangle$ are given by the vectors $\mathbf{a}^i$ and $\mathbf{b}^i$, respectively, i.e.,

$$|a_i\rangle = \sum_{\mu=1}^{m} a_\mu^i |\chi_\mu\rangle \;; \qquad |b_i\rangle = \sum_{\mu=1}^{m} b_\mu^i |\chi_\mu\rangle \tag{6.115}$$

we have

$$\mathbf{P}^a = \sum_{i=1}^{n_a} \mathbf{a}^i \mathbf{a}^{i\dagger} \;; \qquad \mathbf{P}^b = \sum_{i=1}^{n_b} \mathbf{b}^i \mathbf{b}^{i\dagger} \,. \tag{6.116}$$

(So, in components one has, for instance, $P_{\mu\nu}^a = \sum_{i=1}^{n_a} a_\mu^i a_\nu^{i*}$.)

A warning may be appropriate at this point. If one has a complex operator expression containing some projection operator $\hat{Q}$ and one wishes to get the LCAO equivalent of that expression, then one has to construct the respective matrix **Q** analogously to that as we have matrix **P** previously. However, it is not true in general that in all cases one should use matrix **QS** as the LCAO equivalent of operator $\hat{Q}$. One must not make any substitutions mechanically but always must consider explicitly the LCAO expansions of all the expressions because the proper matrix to be used for operator $\hat{Q}$ may appear to be **Q**, **QS**, **SQ**, or even **SQS**, depending on the environment in which operator $\hat{Q}$ occurs in the expression considered. In some cases the LCAO equivalent of the identity operator— which is also a projector—is not the unit matrix but the overlap ("metric") matrix **S**. Thus the use of the level-shift technique mentioned in Section 3.1 requires adding matrix $\Delta\varepsilon(\mathbf{S} - \mathbf{SPS}) = \Delta\varepsilon\mathbf{S}(\mathbf{1} - \mathbf{PS})$ to the Fock matrix **F** as an LCAO representation of the operator $\Delta\varepsilon(1 - \hat{P})$.

5.3. Example of using projectors in the LCAO framework: The UHF equations for successive optimization of the orbitals

The derivation of the UHF equations presented in Section 2.1 can be used to derive that version of the UHF equations that can be utilized also to set up a necessarily convergent algorithm of successive optimization of the orbitals, analogous to that discussed in Section 1.3.

We recall that (6.26) gives the condition of the energy to be stationary under the variations of the orbital $|a_i\rangle$. In the derivation of that equation the fact that the other orbitals are already the converged ones was not utilized, therefore by solving the equation for $|a_i\rangle$ we can get this orbital minimizing the total energy for the case of all the other orbitals being fixed. Assuming that the self-repulsion term $\hat{J}_i^a - \hat{K}_i^a$ has *not* been added to the Fockian (thus, instead of the common Fockian $\hat{F}^a$ we use the "i-dependent" one, $\hat{F}_i^a$) then the problem of finding the optimal orbital a_i *for the given (fixed) other orbitals* can be solved without iterations. We can, therefore, set up a necessarily convergent procedure of solutions by solving these equations cyclically for all the orbitals.

The equation to be solved for orbital $|a_i\rangle$ is accordingly

$$\hat{F}_i^a|a_i\rangle = \sum_{j=1}^{n_a} \varepsilon_{ji}^a|a_j\rangle \ , \tag{6.117}$$

where orbitals $|a_j\rangle$, $j \neq i$ are some fixed orthonormalized orbitals, and we are looking for an orbital $|a_i\rangle$ that is also orthogonal to them. Multiplying (6.117) with $\langle a_l|$ and utilizing the orthogonality of the orbitals, we get, similarly to (6.37),

$$\varepsilon_{li}^a = \langle a_l|\hat{F}_i^a|a_i\rangle \ . \tag{6.118}$$

Substituting (6.118) into (6.117) and bringing the off-diagonal terms to the left-hand side:

$$\hat{F}_i^a|a_i\rangle - \sum_{\substack{j=1\\ j\neq i}}^{n_a} |a_j\rangle\langle a_j|\hat{F}_i^a|a_i\rangle = \varepsilon_{ii}^a|a_i\rangle \tag{6.119}$$

or

$$\left[1 - \sum_{\substack{j=1\\ (j\neq i)}}^{n_a} |a_j\rangle\langle a_j|\right]\hat{F}_i^a|a_i\rangle = \varepsilon_{ii}^a|a_i\rangle \ . \tag{6.120}$$

Koopmans theorem (Section 3) indicates that the lowest total energy of the system will correspond to that solution $|a_i\rangle$ that is the eigenvector with the lowest eigenvalue of this eigenvalue equation. (It is assumed here that all the other orbitals $|a_j\rangle$, $j \neq i$ are fixed.) By introducing the notation

$$\hat{P}_{(i)}^a = \sum_{\substack{j=1\\ (j\neq i)}}^{n_a} |a_j\rangle\langle a_j| \tag{6.121}$$

we can write

$$(1 - \hat{P}_{(i)}^a)\hat{F}_i^a|a_i\rangle = \varepsilon_{ii}^a|a_i\rangle \ . \tag{6.122}$$

Similar to the case discussed in Section 1.5 of Chapter 2, all solutions of this equation that belong to $\varepsilon^a_{ii} \neq 0$ are automatically orthogonal to all the $|a_j\rangle$s that were included when building projector $\hat{P}^a_{(i)}$. Therefore, one may insert an additional projector $1 - \hat{P}_{(i)}$ without destroying the equality, and get the Hermitian eigenvalue equation

$$(1 - \hat{P}^a_{(i)})\hat{F}^a_i(1 - \hat{P}^a_{(i)})|a_i\rangle = \varepsilon^a_{ii}|a_i\rangle \ . \tag{6.123}$$

In order to rewrite this equation to a finite basis $\{\chi_\mu\}$, we shall use for brevity the simplified procedure discussed at the end of Section 5.1. We introduce the LCAO expansion of the orbitals

$$|a_l\rangle = \sum_{\mu=1}^{m} a^l_\mu|\chi_\mu\rangle \tag{6.124}$$

and multiply (6.123) by $\langle\chi_\nu|$. We get

$$\langle\chi_\nu|\left[1 - \sum_{\substack{j=1\\(j\neq i)}}^{n_a} |\sum_{\varrho=1}^{m} a^j_\varrho\chi_\varrho\rangle\langle\sum_{\tau=1}^{m} a^j_\tau\chi_\tau|\right]\hat{F}^a_i\left[1 - \sum_{\substack{k=1\\(k\neq i)}}^{n_a} |\sum_{\eta=1}^{m} a^k_\eta\chi_\eta\rangle\langle\sum_{\lambda=1}^{m} a^k_\lambda\chi_\lambda|\right]$$
$$\times |\sum_{\mu=1}^{m} a^i_\mu\chi_\mu\rangle = \varepsilon^a_{ii}\langle\chi_\nu|\sum_{\mu=1}^{m} a^i_\mu\chi_\mu\rangle \ . \tag{6.125}$$

Expanding this equation and introducing the matrix

$$\mathbf{P}^a_{(i)} = \sum_{\substack{j=1\\(j\neq i)}}^{n_a} \mathbf{a}^j\mathbf{a}^{j\dagger} \tag{6.126}$$

we get trivially the Hermitian generalized matrix eigenvalue equation

$$(\mathbf{1} - \mathbf{SP}^a_{(i)})\mathbf{F}^a_i(1 - \mathbf{P}^a_{(i)}\mathbf{S})\mathbf{a}^i = \varepsilon^a_{ii}\mathbf{Sa}^i \tag{6.127}$$

in which the LCAO equivalent of operator $\hat{P}^a_{(i)}$ is matrix $\mathbf{SP}^a_{(i)}$ in one case and matrix $\mathbf{P}^a_{(i)}\mathbf{S}$ in another.

5.4. The Fock matrix and the energy

According to (6.100), we need the elements of the Fock matrix

$$\begin{aligned} F_{\nu\mu} &= \langle\chi_\nu|\hat{F}|\chi_\mu\rangle = \langle\chi_\nu|\hat{h} + \sum_{i=1}^{n}(2\hat{J}_i - \hat{K}_i)|\chi_\mu\rangle \\ &= h_{\nu\mu} + \sum_{i=1}^{n}(2\langle\chi_\nu|\hat{J}_i|\chi_\mu\rangle - \langle\chi_\nu|\hat{K}_i|\chi_\mu\rangle) \ . \end{aligned} \tag{6.128}$$

(The RHF case is considered.) According to the definition in (6.22) of the Coulomb operator, one can write

$$\begin{aligned}
\langle\chi_\nu|\hat{J}_i|\chi_\mu\rangle &= \int \chi_\nu^*(1)\int \frac{\varphi_i^*(2)\varphi_i(2)}{r_{12}}\,dv_2\chi_\mu(1)dv_1 \qquad (6.129)\\
&= \iint \frac{\chi_\nu^*(1)\sum_{\varrho=1}^{m} c_\varrho^{i*}\chi_\varrho^*(2)\sum_{\tau=1}^{m} c_\tau^i\chi_\tau(2)\chi_\mu(1)}{r_{12}}\,dv_1dv_2\\
&= \sum_{\varrho,\tau=1}^{m} c_\tau^i c_\varrho^{i*}\iint \frac{\chi_\nu^*(1)\chi_\varrho^*(2)\chi_\mu(1)\chi_\tau(2)}{r_{12}}\,dv_1dv_2\\
&= \sum_{\varrho,\tau=1}^{m} c_\tau^i c_\varrho^{i*}[\chi_\nu\chi_\varrho|\chi_\mu\chi_\tau]\,.
\end{aligned}$$

Quite analogously, we get also the matrix element of the exchange operator (6.23)

$$\langle\chi_\nu|\hat{K}_i|\chi_\mu\rangle = \sum_{\varrho,\tau=1}^{m} c_\tau^i c_\varrho^{i*}[\chi_\nu\chi_\varrho|\chi_\tau\chi_\mu]\,. \qquad (6.130)$$

By substituting (6.129) and (6.130) into (6.128) and using the definition in (6.108) of the elements of the P-matrix, we get

$$\begin{aligned}
F_{\nu\mu} &= h_{\nu\mu} + \sum_{i=1}^{n}\left(2\sum_{\varrho,\tau=1}^{m} c_\tau^i c_\varrho^{i*}[\chi_\nu\chi_\varrho|\chi_\mu\chi_\tau] - \sum_{\varrho,\tau=1}^{m} c_\tau^i c_\varrho^{i*}[\chi_\nu\chi_\varrho|\chi_\tau\chi_\mu]\right)\\
&= h_{\nu\mu} + \sum_{\varrho,\tau=1}^{m} P_{\tau\varrho}(2[\chi_\nu\chi_\varrho|\chi_\mu\chi_\tau] - [\chi_\nu\chi_\varrho|\chi_\tau\chi_\mu])\,. \qquad (6.131)
\end{aligned}$$

This result indicates that for the calculation of the Fock matrix one needs matrix $\mathbf{P}$ and the one- and two-electron integrals $h_{\nu\mu}$ and $[\chi_\nu\chi_\varrho|\chi_\mu\chi_\tau]$, respectively.

The UHF Fock matrices can be obtained by performing quite similar substitutions. One gets

$$F_{\nu\mu}^a = h_{\nu\mu} + \sum_{\varrho,\tau=1}^{m}\left[(P_{\tau\varrho}^a + P_{\tau\varrho}^b)[\chi_\nu\chi_\varrho|\chi_\mu\chi_\tau] - P_{\tau\varrho}^a[\chi_\nu\chi_\varrho|\chi_\tau\chi_\mu]\right]\,. \qquad (6.132)$$

The elements of matrix $\mathbf{F}^b$ can be obtained by replacing $P_{\tau\varrho}^a$ with $P_{\tau\varrho}^b$ in the last term.

When performing theoretical analyzes, it is quite convenient to use the "[12|12] convention for the integrals," i.e., the notation $[\chi_\mu\chi_\nu|\chi_\varrho\chi_\tau] \equiv \langle\chi_\mu(1)\chi_\nu(2)|\frac{1}{r_{12}}|\chi_\varrho(1)\chi_\tau(2)\rangle$ of the two-electron integrals, because one can

distinguish between the terms coming from the "bra" and the "ket." Also, it is the author's opinion that in the abstract theoretical developments it is worthwhile to keep the possibility of using complex wave functions. However, when the programming work is considered, one has to take the advantage that the basis functions used in practice are real. Therefore, it is needless to distinguish between functions with and without complex conjugation, and it is more convenient to use the notations of the two-electron integrals in the $(11|22)$ convention:

$$(\chi_\mu\chi_\nu|\chi_\varrho\chi_\tau) = \iint \frac{\chi_\mu(1)\chi_\nu(1)\chi_\varrho(2)\chi_\tau(2)}{r_{12}}\, dv_1 dv_2 \,. \tag{6.133}$$

(One can see that $(\chi_\mu\chi_\nu|\chi_\varrho\chi_\tau) \equiv [\chi_\mu\chi_\varrho|\chi_\nu\chi_\tau]$ for real orbitals.)

Inspection of the definition (6.133) indicates that it is invariant with respect to the interchange of the orbitals χ_μ and χ_ν, χ_ϱ and χ_τ as well as of the two sets of the orbitals corresponding to the variables of integration "1" and "2." By denoting for brevity the integral $(\chi_\mu\chi_\nu|\chi_\varrho\chi_\tau)$ as $(\mu\nu|\varrho\tau)$, these invariances result in the equality

$$\begin{aligned} &(\mu\nu|\varrho\tau) = (\nu\mu|\varrho\tau) = (\mu\nu|\tau\varrho) = (\nu\mu|\tau\varrho) \\ = \;& (\varrho\tau|\mu\nu) = (\varrho\tau|\nu\mu) = (\tau\varrho|\mu\nu) = (\tau\varrho|\nu\mu) \,. \end{aligned} \tag{6.134}$$

These symmetries can easily be understood if one considers the two-electron integral (6.133) as the electrostatic interaction between two charge densities $\chi_\mu(\vec{r}_1)\chi_\nu(\vec{r}_1)$ and $\chi_\varrho(\vec{r}_2)\chi_\tau(\vec{r}_2)$.

According to (6.134), if all four indices μ, ν, ϱ, and τ are different, then there are eight integrals with different order of subscripts in the definition in (6.133), the values of which are equal. (This number is reduced to four if there is one pair of coinciding subscripts, two if two pairs coincide and one if all the subscripts are the same.)

By using the notations in (6.133) and the elements of matrix $\mathbf{D}$ defined in (6.110), the elements of the Fock matrix can be written in the RHF case as

$$F_{\nu\mu} = h_{\nu\mu} + \sum_{\varrho,\tau=1}^{m} D_{\tau\varrho}[(\chi_\nu\chi_\mu|\chi_\varrho\chi_\tau) - \frac{1}{2}(\chi_\nu\chi_\varrho|\chi_\mu\chi_\tau)] \,. \tag{6.135}$$

(This equation can be obtained from its UHF counterpart in (6.132) by taking into account that in the RHF case $\mathbf{P}^a = \mathbf{P}^b = \frac{1}{2}\mathbf{D}$. In the earlier literature the differences $(\chi_\nu\chi_\mu|\chi_\varrho\chi_\tau) - \frac{1}{2}(\chi_\nu\chi_\varrho|\chi_\mu\chi_\tau)$ were called the elements of a "hypermatrix.")

According to (6.67) the single-determinant electronic energy in the UHF case can be written as

$$E = \frac{1}{2}\left(\sum_{i=1}^{n_a}\langle a_i|\hat{h} + \hat{F}^a|a_i\rangle + \sum_{i=1}^{n_b}\langle b_i|\hat{h} + \hat{F}^b|b_i\rangle\right) \,. \tag{6.136}$$

By substituting the LCAO expansion of the orbitals a_i, b_i we get

$$
\begin{aligned}
E &= \frac{1}{2}\left(\sum_{i=1}^{n_a}\langle\sum_{\mu=1}^{m} a_\mu^i\chi_\mu|\hat{h}+\hat{F}^a|\sum_{\nu=1}^{m} a_\nu^i\chi_\nu\rangle\right. \\
&\qquad\left. + \sum_{i=1}^{n_b}\langle\sum_{\mu=1}^{m} b_\mu^i\chi_\mu|\hat{h}+\hat{F}^b|\sum_{\nu=1}^{m} b_\nu^i\chi_\nu\rangle\right) \\
&= \frac{1}{2}\left(\sum_{i=1}^{n_a}\sum_{\mu,\nu=1}^{m} a_\nu^i a_\mu^{i*}\langle\chi_\mu|\hat{h}+\hat{F}^a|\chi_\nu\rangle\right. \\
&\qquad\left. + \sum_{i=1}^{n_b}\sum_{\mu,\nu=1}^{m} b_\nu^i\, b_\mu^{i*}\langle\chi_\mu|\,\hat{h}+\hat{F}^b\,|\chi_\nu\rangle\right) \\
&= \frac{1}{2}\sum_{\mu,\nu=1}^{m}\left[P_{\nu\mu}^a(h_{\mu\nu}+F_{\mu\nu}^a)+P_{\nu\mu}^b(h_{\mu\nu}+F_{\mu\nu}^b)\right] \\
&= \frac{1}{2}\left\{\mathrm{Tr}\left[\mathbf{P}^a(\mathbf{h}+\mathbf{F}^a)\right]+\mathrm{Tr}\left[\mathbf{P}^b(\mathbf{h}+\mathbf{F}^b)\right]\right\}\,.
\end{aligned}
\tag{6.137}
$$

In the RHF case this expression reduces simply to

$$
E=\frac{1}{2}\mathrm{Tr}[\mathbf{D}(\mathbf{h}+\mathbf{F})]\,. \tag{6.138}
$$

Of course, one can express the energy directly in terms of the one- and two-electron integrals by substituting the explicit expression of the Fock matrix elements. Thus, for instance, by substituting (6.135) into (6.138) one gets the RHF electronic energy as

$$
E=\sum_{\mu,\nu=1}^{m} D_{\mu\nu}h_{\nu\mu}+\frac{1}{2}\sum_{\mu,\nu,\varrho,\tau=1}^{m}(D_{\mu\nu}D_{\tau\varrho}-\frac{1}{2}D_{\varrho\nu}D_{\tau\mu})(\chi_\nu\chi_\mu|\chi_\varrho\chi_\tau)\,. \tag{6.139}
$$

(The last term, originating from the exchange part of the Fock matrix, has been obtained by interchanging the summation indices μ and ϱ.)

6. Restricted open-shell Hartree–Fock method

Open shell-systems (radicals, triplet states, etc.) are often treated with the single-determinant UHF method. As discussed previously, it has the disadvantage that the wave function is not a pure spin eigenfunction. Another problem appears in the case when one wishes to compare the energies and other properties of a closed-shell and some related open-shell system(s). As already discussed, the UHF method often fails to give a solution differing from the RHF one

around the equilibrium geometry of a molecule. This means that the closed-shell system under study is actually described by using doubly filled orbitals. At the same time, the open-shell system obtained by ionization/excitation of the starting molecule or representing an open-shell fragment of it contains a different number of electrons with spins α and β. Consequently—because of the difference in the exchange term of the Fockian—the UHF treatment of the open-shell system cannot result in strictly doubly filled orbitals. (There will be some orbital splitting in those parts of the system that may be considered identical in the closed-shell and open-shell cases.) The use of different orbitals for different spins represents the application of a more flexible variational trial wave function that may result in accounting for some electron correlation effects. From the other side, one can study some of these problems based on the calculations performed for the parent closed-shell system, as it has been considered for the case of ionization by using Koopmans theorem (Section 3) or singlet and triplet excitations from an RHF ground state (Section 4.4). In this case, however, one loses the possibility of accounting for the effect of orbital relaxation during the ionization/excitation.

These and similar problems are dealt with in the "restricted open shell Hartree–Fock (ROHF) method" in which one distinguishes between closed and open shells and treats them differently. The closed shells consists of doubly occupied orbitals; their splitting is not admitted. Depending on the nature of the open shell, the wave function may be a single determinant as is the case of a doublet or of a high-spin ($S_z = 1$) component of a triplet, or it may be a linear combination of two (or more) determinants, as is the case for an open-shell singlet (singlet biradical) or for the $S_z = 0$ component of a triplet. Multideterminant wave functions can arise also if a high spatial symmetry is requested, e.g., in the case of free atoms.

It is to be noted that there are problems in which forcing the correct spin symmetry by introducing as many doubly filled orbitals as possible may have also disadvantages: the ROHF wave function cannot account for the phenomenon of spin polarization, which is inevitable for understanding hyperfine splittings in the EPR spectra, etc.

In the ROHF scheme the different open-shell systems are treated in a common framework on the basis of introducing a sufficiently general energy formula containing some variable numerical parameters the values of which should be specified for each actual type of problem. In all cases it is assumed that we are dealing with some number N of orthonormalized (spatial) orbitals φ_i, each of which can be assigned an occupation number ω_i and that the energy can be written as

$$E = \sum_{i=1}^{N} \omega_i h_{ii} + \sum_{i,j=1}^{N} \left(\alpha_{ij} J_{ij} - \beta_{ij} K_{ij}\right) , \tag{6.140}$$

where the "coupling coefficients" α_{ij} and β_{ij} should be determined for every type of open-shell systems separately. In (6.140) $h_{ii} = \langle\varphi_i|\hat{h}|\varphi_i\rangle$ is the matrix element of the one-electron part of the Hamiltonian and $J_{ij} = [\varphi_i\varphi_j|\varphi_i\varphi_j]$ and $K_{ij} = [\varphi_i\varphi_j|\varphi_j\varphi_i]$ are the Coulomb and exchange integrals, respectively. Note that in (6.140) the sums over i and j run independently.

Of course, Eq. (6.140) is sufficiently general as to include the case of the usual closed-shell RHF wave function (with the values of parameters $N = n$; $\omega_i = 2$; $\alpha_i = 2$; $\beta_{ij} = 1$) and even the UHF wave function ($N = n_a + n_b$; $\omega_i = 1$; $\alpha_{ij} = \frac{1}{2}$; $\beta_{ij} = \frac{1}{2}$ if i and j both belong to the same set $\{a_i\}$ or $\{b_i\}$, and $\beta_{ij} = 0$ otherwise). At the end of this section we shall derive the values of the coupling coefficients for the most important open-shell cases mentioned previously.

6.1. The ROHF equations

The ROHF scheme is essentially defined by the energy formula of the general form (6.140) and the requirement of all the orbitals to be orthonormalized. Therefore, the most adequate treatment of the variational problem is that with the use of the Lagrangian multiplier technique. The procedure is practically the same as applied in Section 2.2, the main difference being the presence of the general coefficients ω_i, α_i, and β_i in the energy formula (6.140). We define the auxiliary functional

$$F = E - \sum_{i,j=1}^{N} 2\varepsilon_{ji}(\langle\varphi_i|\varphi_j\rangle - \delta_{ij}) \tag{6.141}$$

and require its variation to vanish: $\delta F = 0$ under the variations $\delta\varphi_i$ of the orbitals φ_i; $i = 1, 2, \ldots, N$. In (6.141) E stands for the energy expression (6.140) and, as it is discussed in Section 2.2 in detail, the Lagrangian multiplier matrix $\boldsymbol{\varepsilon}$ should be Hermitian and the solutions φ_i should be orthonormalized. (We have defined the Lagrangian multipliers as $-2\varepsilon_{ji}$ for further convenience.)

We get the variation δF under the variation of some orbital φ_k as

$$\begin{aligned}\delta F = {} & \omega_k\langle\delta\varphi_k|\hat{h}|\varphi_k\rangle + \sum_{j=1}^{N}(\alpha_{kj}\langle\delta\varphi_k|\hat{J}_j|\varphi_k\rangle - \beta_{kj}\langle\delta\varphi_k|\hat{K}_j|\varphi_k\rangle) \\ & + \sum_{i=1}^{N}(\alpha_{ik}\langle\delta\varphi_k|\hat{J}_i|\varphi_k\rangle - \beta_{ik}\langle\delta\varphi_k|\hat{K}_i|\varphi_k\rangle) - \sum_{j=1}^{N} 2\varepsilon_{jk}\langle\delta\varphi_k|\varphi_j\rangle + \text{c.c.} = 0\,.\end{aligned} \tag{6.142}$$

Observing that the sums over i and j are equal and introducing the "k-dependent" ROHF Fockian

$$\hat{F}_k = \frac{\omega_k}{2}\,\hat{h} + \sum_{j=1}^{N}(\alpha_{kj}\hat{J}_j - \beta_{kj}\hat{K}_j) \tag{6.143}$$

we rewrite condition (6.142) in the form of the ROHF equations

$$\hat{F}_k|\varphi_k\rangle = \sum_{j=1}^{N} \varepsilon_{jk}|\varphi_j\rangle \qquad k = 1, 2, \ldots, N \ . \tag{6.144}$$

The k-dependence of the Fockian means that one has a different Fockian for each orbital or, at least, each class of orbitals. In the ROHF case one also has to introduce explicitly the requirement of Hermiticity of the Lagrange multipliers

$$\varepsilon_{jk} = \varepsilon_{kj}^* \tag{6.145}$$

because in general it is not fulfilled automatically owing to the k-dependence of the Fockian in (6.143). In the ROHF case—contrary to the RHF or UHF ones—unitary transformations are permitted only within the given class of the orbitals and, consequently, one cannot get rid completely of the off-diagonal Lagrangian multipliers connecting different orbital sets.

It is often considered desirable to have, at least formally, a common eigenvalue equation of which all the orbitals φ_i are eigenfunctions. We shall consider briefly two possible equations of this type.

In the first case we start with similar manipulations as used already in Section 5.3. We multiply (6.144) with $\langle\varphi_l|$ and get the value of the Lagrangian multiplier ε_{lk}:

$$\langle\varphi_l|\hat{F}_k|\varphi_k\rangle = \sum_{j=1}^{N} \varepsilon_{jk}\langle\varphi_l|\varphi_j\rangle = \sum_{j=1}^{N} \varepsilon_{jk}\delta_{lj} = \varepsilon_{lk} \ . \tag{6.146}$$

Substituting (6.146) into (6.144) and bringing all terms except that with $j = k$ to the left-hand side, we have

$$\hat{F}_k|\varphi_k\rangle - \sum_{\substack{j=1 \\ (j\neq k)}}^{N} |\varphi_j\rangle\langle\varphi_j|\hat{F}_k|\varphi_k\rangle = \varepsilon_{kk}|\varphi_k\rangle \tag{6.147}$$

i.e.,

$$(1 - \hat{P}_{(k)})\hat{F}_k|\varphi_k\rangle = \varepsilon_{kk}|\varphi_k\rangle \ , \tag{6.148}$$

where, similar to the case discussed in Section 5.3, we use the notation $\hat{P}_{(k)}$ for the projector:

$$\hat{P}_{(k)} = \sum_{\substack{j=1 \\ (j\neq k)}}^{N} |\varphi_j\rangle\langle\varphi_j| \ . \tag{6.149}$$

Owing to the equality $(1 - \hat{P}_{(k)})|\varphi_k\rangle = |\varphi_k\rangle$, we may Hermitize (6.149) as

$$(1 - \hat{P}_{(k)})\hat{F}_k(1 - \hat{P}_{(k)})|\varphi_k\rangle = \varepsilon_{kk}|\varphi_k\rangle \ . \tag{6.150}$$

Each occupied orbital $|\varphi_i\rangle$, $i \neq k$, is an eigenvector of this equation with an eigenvalue of 0. Therefore we may form the sum of these operators and obtain a common Hermitian eigenvalue equation that is satisfied by all the orbitals φ_i:

$$\hat{G}|\varphi_i\rangle = \varepsilon_{ii}|\varphi_i\rangle \ , \tag{6.151}$$

where

$$\hat{G} = \sum_{k=1}^{N}(1 - \hat{P}_{(k)})\hat{F}_k(1 - \hat{P}_{(k)}) \ . \tag{6.152}$$

Unfortunately, the fulfillment of this equation does not guarantee that of the Hermiticity requirements in (6.145). These requirements can be reformulated in the following manner: the ROHF orbitals should be *simultaneously* eigenvectors also of another Hermitian operator

$$\hat{Q} = \sum_{i,j=1}^{N} \hat{P}_i(\hat{F}_i - \hat{F}_j)\hat{P}_j \ . \tag{6.153}$$

In (6.153) we use the notation of the projector

$$\hat{P}_i = |\varphi_i\rangle\langle\varphi_i| \tag{6.154}$$

—it should be distinguished from the notation $\hat{P}_{(i)}$ introduced in (6.149). In the fact, let $|\varphi_l\rangle$ be an eigenvector of $\hat{Q}$ with an eigenvalue κ_l:

$$\hat{Q}|\varphi_l\rangle = \sum_{i,j=1}^{N} \hat{P}_i(\hat{F}_i - \hat{F}_j)\hat{P}_j|\varphi_l\rangle = \sum_{i=1}^{N} \hat{P}_i(\hat{F}_i - \hat{F}_l)|\varphi_l\rangle = \kappa_l|\varphi_l\rangle \ . \tag{6.155}$$

Multiplying (6.155) by some $\langle\varphi_k|$ $(k \neq l)$, and utilizing the orthogonality of the orbitals $|\varphi_k\rangle$ and $|\varphi_l\rangle$, we get

$$\langle\varphi_k|\sum_{i=1}^{N} \hat{P}_i(\hat{F}_i - \hat{F}_l)|\varphi_l\rangle = \langle\varphi_k|\hat{F}_k|\varphi_l\rangle - \langle\varphi_k|\hat{F}_l|\varphi_l\rangle = \kappa_l\langle\varphi_k|\varphi_l\rangle = 0 \ . \tag{6.156}$$

According to (6.146), one has $\varepsilon_{lk} = \langle\varphi_l|\hat{F}_k|\varphi_k\rangle$ and, analogously, $\varepsilon_{kl} = \langle\varphi_k|\hat{F}_l|\varphi_l\rangle$. Therefore (6.156) indicates that the equality $\hat{Q}|\varphi_l\rangle = \kappa_l|\varphi_l\rangle$ $(l = 1, 2, \ldots, N)$ is equivalent with the Hermiticity requirement in (6.145).

It is not so trivial to ensure the simultaneous diagonalization of two Hermitian operators $\hat{G}$ and $\hat{Q}$. It may be of interest, therefore, to look for

an eigenvalue equation that automatically ensures the fulfillment of the conditions in (6.145), as well. For that reason, one has to introduce explicitly these conditions into the ROHF equations (6.144) by replacing ε_{jk} with $\varepsilon^*_{kj} = \langle\varphi_k|\hat{F}_j|\varphi_j\rangle^* = \langle\varphi_j|\hat{F}_j|\varphi_k\rangle$. One gets

$$\hat{F}_k|\varphi_k\rangle = \sum_{\substack{j=1\\(j\neq k)}}^{N} |\varphi_j\rangle\langle\varphi_j|\hat{F}_j|\varphi_k\rangle + \varepsilon_{kk}|\varphi_k\rangle \,, \tag{6.157}$$

i.e.,

$$\left(\hat{F}_k - \sum_{\substack{j=1\\(j\neq k)}}^{N} |\varphi_j\rangle\langle\varphi_j|\hat{F}_j\right)|\varphi_k\rangle = \varepsilon_{kk}|\varphi_k\rangle \,. \tag{6.158}$$

This equation differs from (6.148) by the appearance of the operators $\hat{F}_j$ in the sum instead of $\hat{F}_k$. We multiply (6.158) with the corresponding sides of the orthonormalization condition

$$\langle\varphi_k|\varphi_i\rangle = \delta_{ki} \tag{6.159}$$

and sum over k:

$$\sum_{k=1}^{N}\left[\left(\hat{F}_k - \sum_{\substack{j=1\\(j\neq k)}}^{N} \hat{P}_j\hat{F}_j\right)\hat{P}_k\right]|\varphi_i\rangle = \varepsilon_{ii}|\varphi_i\rangle \,. \tag{6.160}$$

By multiplying this equation with $\langle\varphi_l|$ $(l \neq i)$, we get, by taking into account that $\hat{P}_k|\varphi_i\rangle = \delta_{ki}|\varphi_i\rangle$ and $\langle\varphi_l|\hat{P}_j = \delta_{jl}\langle\varphi_l|$,

$$\begin{aligned}\langle\varphi_l|\sum_{k=1}^{N}&\left[\left(\hat{F}_k - \sum_{\substack{j=1\\(j\neq k)}}^{N} \hat{P}_j\hat{F}_j\right)\hat{P}_k\right]|\varphi_i\rangle \\ &= \langle\varphi_l|\sum_{k=1}^{N}\hat{F}_k\hat{P}_k|\varphi_i\rangle - \langle\varphi_l|\sum_{k=1}^{N}\sum_{\substack{j=1\\(j\neq k)}}^{N}\hat{P}_j\hat{F}_j\hat{P}_k|\varphi_i\rangle \\ &= \langle\varphi_l|\hat{F}_i|\varphi_i\rangle - \langle\varphi_l|\hat{F}_l|\varphi_i\rangle = 0 \,.\end{aligned} \tag{6.161}$$

This means that the solutions of (6.160) automatically fulfill the Hermiticity requirement in (6.145). Unfortunately, however, (6.160) is not Hermitian; if one performs a simple Hermitization of it, then the automatic fulfillment of the requirement in (6.145) may be lost.

We shall not enter into the consideration of technical problems connected with the ROHF scheme, only note that different programs in common use produce the same ROHF orbitals by assigning them quite different "orbital energies" (eigenvalues).

6.2. The coupling coefficients for some open-shell systems

The simplest case with a single (high-spin) open-shell (doublet, $S_z = 1$ component of a triplet, etc.) can be described by using a single-determinant wave function. In this case we consider n_c doubly filled (closed-shell) orbitals and n_o singly filled (open-shell) ones having the same spin (say α). Then the energy can trivially be obtained using the general formula in (6.44) of the single determinant energy by substituting in it $n_b = n_c$, $n_a = n_c + n_o$, $a_p = b_p = \varphi_p$ ($p \leq n_c$), and $a_p = \varphi_p$ ($n_c + 1 \leq p \leq n_c + n_o$). We get

$$E = \sum_{p=1}^{n_c} 2h_{pp} + \sum_{q=n_c+1}^{n_c+n_o} h_{qq} + \sum_{p,q=1}^{n_c} (2J_{pq} - K_{pq}) + \sum_{p=1}^{n_c} \sum_{q=n_c+1}^{n_c+n_o} (2J_{pq} - K_{pq})$$
$$+ \frac{1}{2} \sum_{p,q=n_c+1}^{n_c+n_o} (J_{pq} - K_{pq}) \tag{6.162}$$

Accordingly, the coefficients to be substituted in (6.140) are $\omega_c = 2$, $\omega_o = 1$, $\alpha_{cc} = 2$, $\alpha_{co} = \alpha_{oc} = 1$, $\alpha_{oo} = 1/2$; $\beta_{cc} = 1$, and $\beta_{co} = \beta_{oc} = \beta_{oo} = 1/2$, where the subscripts c or o indicate the *class* $\{C\}$ or $\{O\}$ to which orbitals φ_i, φ_j belong. (C=closed, O=open.) When determining these coefficients, we have taken into account the fact that the sums over i and j in (6.140) run independently, so there is a case when $i \in \{C\}$ and $j \in \{O\}$ and also one with $j \in \{C\}$ and $i \in \{O\}$. In the doublet case ($n_o = 1$) there is only one open-shell orbital and the last sum of (6.162) vanishes ($J_{pp} \equiv K_{pp}$).

In the case of an open-shell singlet or of an $S_z = 0$ component of a triplet, one can proceed exactly in the same manner as has been done when considering singlet and triplet excitations from an RHF wave function (Section 4.4). We have

$$^{1,3}\Psi = \frac{1}{\sqrt{2}}[\Psi_1 \pm \Psi_2] \,, \tag{6.163}$$

where the two determinants Ψ_1 and Ψ_2 are defined as

$$\Psi_1 = \mathcal{A}[\varphi_1^2, \varphi_1^2 \ldots \varphi_{n_c}^2 \varphi_i^\alpha \varphi_j^\beta] \tag{6.164}$$

and

$$\Psi_2 = \mathcal{A}[\varphi_1^2, \varphi_1^2 \ldots \varphi_{n_c}^2 \varphi_j^\alpha \varphi_i^\beta] \tag{6.165}$$

and the upper sign again corresponds to the singlet and the lower one to the triplet. In (6.164), (6.165) we have introduced the simplified notation φ_p^2 for a doubly filled orbital and have omitted the arguments (coordinates of the individual electrons).

We have to calculate

$$^{1,3}E = \frac{1}{2}\langle \Psi_1 \pm \Psi_2 | \hat{H} | \Psi_1 \pm \Psi_2 \rangle \,. \tag{6.166}$$

Here $\langle\Psi_1|\hat{H}|\Psi_1\rangle = \langle\Psi_2|\hat{H}|\Psi_2\rangle$ is the single-determinant energy of Ψ_1 and Ψ_2; it can be calculated using the same general formula in (6.44) by substituting in it $n_a = n_b = n_c + 1$, $a_p = b_p = \varphi_p$ $(p \leq n_c)$, $a_{n_c+1} = \varphi_i$, and $b_{n_c+1} = \varphi_j$. We get

$$\begin{aligned}\langle\Psi_1|\hat{H}|\Psi_1\rangle &= \sum_{p=1}^{n_c} 2h_{pp} + h_{ii} + h_{jj} + \sum_{p,q=1}^{n_c} (2J_{pq} - K_{pq}) \\ &+ \sum_{p=1}^{n_c} [2(J_{pi} + J_{pj}) - (K_{pi} + K_{pj})] + J_{ij} \ .\end{aligned} \tag{6.167}$$

Assuming real orbitals, we get, similar to the case discussed in Section 4.4,

$$\langle\Psi_1|\hat{H}|\Psi_2\rangle = \langle\Psi_2|\hat{H}|\Psi_1\rangle = K_{ij} \ . \tag{6.168}$$

Thus, introducing the notation $i = n_c+1, j = n_c+2$ to simplify the formula:

$$\begin{aligned}{}^{1,3}E &= \sum_{p=1}^{n_c} 2h_{pp} + \sum_{q=n_c+1}^{n_c+2} h_{qq} + \sum_{p,q=1}^{n_c} (2J_{pq} - K_{pq}) \\ &+ \sum_{p=1}^{n_c} \sum_{q=n_c+1}^{n_c+2} (2J_{pq} - K_{pq}) + J_{ij} \pm K_{ji} \ .\end{aligned} \tag{6.169}$$

The values of the coefficients are accordingly as follows. Common values: $\omega_c = 2$, $\omega_o = 1$, $\alpha_{cc} = 2$, $\alpha_{co} = \alpha_{oc} = 1$, $\alpha_{oo} = 1/2$; $\beta_{cc} = 1$; $\beta_{co} = 1/2$; specific values: $\beta_{oo} = -1/2$ for the singlet and $\beta_{oo} = 1/2$ for the triplet.

As expected, the $S_z = 1$ and $S_z = 0$ components of the triplet have the same energy—although the former is a single determinant but the latter is not.

7. The gradient of the energy

One of the basic tasks in practical quantum chemistry is the determination of the minima and other stationary points on the the Born–Oppenheimer potential energy surface of molecular systems. The knowledge of the minima is necessary to study the geometrical parameters of stable molecules, their different conformations, etc.; the saddle points are important for studying different conversion processes and chemical reactions. Although the study of potential surfaces is, in principle, possible by calculating the energy in a sufficiently large number of different geometrical arrangements of the nuclei, the task of optimizing the geometry can be performed much more effectively if one can compute the gradient of the energy, i.e., its derivatives according to the different nuclear coordinates.

It has already been discussed in Section 2.1 of Chapter 2 that the Hellmann–Feynman theorem is not applicable to the HFR wave functions, if—as it is

usual—the basis orbitals "follow" the nuclei, and a change of the geometry generates also a change of the basis functions applied. In other words, in the finite basis case the variations of the basis orbitals give contributions to the variation of the energy δE taking place under the infinitesimal change of the geometry $\delta\mathbf{R}$. (Here $\mathbf{R}$ stands for the set of all the nuclear coordinates.) These contributions give rise to the so-called "wave function forces," as opposed to the "Hellmann–Feynman forces" pertinent to the case when the Hellmann–Feynman theorem holds.

We shall consider[2] an SCF-type wave function Ψ that may be a single determinant or a linear combination of determinants; in the latter case it is assumed that the different determinants either have fixed coefficients (as in ROHF or projected HF theories) or the coefficients of the individual determinants are optimized variationally (as in different MC–SCF theories). The wave function Ψ depends on N orbitals ψ_i that are expanded in the basis set $\{\chi_\mu\}$ as

$$\psi_i = \sum_{\mu=1}^{m} c_\mu^i \chi_\mu \ . \tag{6.170}$$

We rewrite Eq. (2.55) in the form

$$\delta E = \frac{\langle\Psi|\delta\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} + \frac{1}{\langle\Psi|\Psi\rangle}\{\langle\delta\Psi|\hat{H} - E|\Psi\rangle \ + \ \text{c.c.}\ \} \ . \tag{6.171}$$

Here both $\delta\hat{H}$ and $\delta\Psi$ are the result of the variation $\delta\mathbf{R}$. Obviously, the first term on the right-hand side of (6.171) corresponds to the Hellmann–Feynman force, the second to the "wave function force."

The change of the geometry $\delta\mathbf{R}$ induces a variation of the orbital ψ_i, which consists of two terms:

$$\delta\psi_i = \delta\psi_i' + \delta\psi_i'' = \sum_{\mu=1}^{m} \delta c_\mu^i \chi_\mu + \sum_{\mu=1}^{m} c_\mu^i \delta\chi_\mu \tag{6.172}$$

The first term $(\delta\psi_i')$ contains the variation of the orbital coefficients c_μ^i, the second one $(\delta\psi_i'')$ those of the basis orbitals χ_μ.

Owing to the linearity of the first-order variations and the properties of the determinants, the first order variation of Ψ can be presented as a sum

$$\delta\Psi = \sum_{i=1}^{N}(\delta\Psi_i' + \delta\Psi_i'') + \delta\Psi_C \ , \tag{6.173}$$

[2]The derivation that follows is reprinted with permission of Elsevier Science from our paper M. B. Ruiz and I. Mayer, "Deriving Gradient Formulae for SCF Methods by Using Brillouin-type Theorems. Gradients in the HPHF Method", Chemical Physics Letters, **236**, pages 217–228, Copyright 1995 Elsevier Science.

where $\delta\Psi_i'$ and $\delta\Psi_i''$ are the variations of the many-electron wave function Ψ taking place when the orbital ψ_i undergoes the variations $\delta\psi_i'$ and ψ_i'', respectively. If Ψ is a multideterminant (MC–SCF) wave function with variable coefficients, then $\delta\Psi_C$ gives the variation of Ψ connected with the variation of these coefficients. If Ψ is a single determinant or a sum of determinants with fixed coefficients, (RHF, UHF, ROHF, projected HF cases) then the term $\delta\Psi_C$ is absent.

Now, terms $\delta\Psi_i'$ and $\delta\Psi_C$ (if any) are variations of the type that are also permitted when the wave function is optimized at a fixed geometry according to the variation principle, so the equality $\langle\delta\Psi|\hat{H} - E|\Psi\rangle = 0$ holds for them (see Section 1.3 of Chapter 2). Accordingly, when calculating the gradient from a *variational* wave function, it is sufficient to consider in (6.171) only variations $\delta\Psi_i''$ that are connected with the changes of the basis orbitals, but not those that are related to the changes in orbital coefficients (or coefficients at the different determinants).

As $\delta\Psi_i''$ is none other than the singly excited configuration $\Psi_1(\psi_i \to \delta\psi_i'')$, the general formula for the first-order energy change of wave function Ψ, taking place under the influence of the geometry change $\delta\vec{R}$, can be written as

$$\delta E = \frac{\langle\Psi|\delta\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} + \frac{1}{\langle\Psi|\Psi\rangle}\sum_{i=1}^{N}\{\langle\Psi_1(\psi_i \to \delta\psi_i'')|\hat{H} - E|\Psi\rangle + \text{c.c.}\}\ . \quad (6.174)$$

When using (6.174), one has to remember that in the usual Born–Oppenheimer Hamiltonian only the electron–nuclear attraction term

$$\hat{V}_{eN} = -\sum_{\alpha=1}^{N_N}\sum_{i=1}^{N_e}\frac{Z_\alpha}{r_{\alpha i}} \quad (6.175)$$

depends on the nuclear geometry, so $\delta\hat{H} \equiv \delta\hat{V}_{eN}$.

Now we shall concretize the previous results for the UHF case, when Ψ is a single determinant built up of n_a orbitals a_i filled with spin α and n_b orbitals b_i filled with spin β, having the finite basis expansions

$$a_i = \sum_{\mu=1}^{m} a_\mu^i \chi_\mu\ ; \qquad b_i = \sum_{\mu=1}^{m} b_\mu^i \chi_\mu \quad (6.176)$$

Equation (6.174) can be specified for this case as

$$\begin{aligned}\delta E = \langle\Psi|\delta\hat{V}_{Ne}|\Psi\rangle \ &+ \ \sum_{i=1}^{n_a}\{\langle\Psi_1(a_i \to \delta a_i'')|\hat{H} - E|\Psi\rangle \ + \ \text{c.c.}\} \\ &+ \ \sum_{i=1}^{n_b}\{\langle\Psi_1(b_i \to \delta b_i'')|\hat{H} - E|\Psi\rangle \ + \ \text{c.c.}\}\ .\end{aligned} \quad (6.177)$$

(The orbitals are assumed orthonormalized, so $\langle\Psi|\Psi\rangle = 1$.) As $\delta\hat{V}_{eN}$ represents a sum of one-electron operators, Slater rules give

$$\langle\Psi|\delta\hat{H}|\Psi\rangle = \langle\Psi|\delta\hat{V}_{Ne}|\Psi\rangle = \sum_{i=1}^{n_a}\langle a_i|\delta\hat{h}|a_i\rangle + \sum_{i=1}^{n_b}\langle b_i|\delta\hat{h}|b_i\rangle \ , \tag{6.178}$$

with

$$\delta\hat{h} = \delta\sum_{\alpha=1}^{N_N}\frac{-Z_\alpha}{r_\alpha} \ . \tag{6.179}$$

Now, we shall consider the variation $\delta a_i'' = \sum_\mu a_\mu^i \delta\chi_\mu$ in detail. We introduce a resolution of identity in the form $1 \equiv \hat{P}^a + (1 - \hat{P}^a)$, where $\hat{P}^a$ is the projector onto the subspace of molecular orbitals $\{a_j\}$ occupied with spin α. Then the variation $\delta a_i''$ can be presented as a sum of two terms:

$$\delta a_i'' = \delta a_{i,1}'' + \delta a_{i,2}'' = \hat{P}^a\delta a_i'' + (1 - \hat{P}^a)\delta a_i'' \ , \tag{6.180}$$

the first of which lies completely in the subspace of the occupied orbitals a_j, whereas the second is orthogonal to that subspace. Therefore, we have also

$$\Psi_1(a_i \to \delta a_i'') = \Psi_1(a_i \to \delta a_{i,1}'') + \Psi_1(a_i \to \delta a_{i,2}'') \tag{6.181}$$

according to the properties of determinants. As $\delta a_{i,1}'' = \hat{P}^a\delta a_i''$ is a linear combination of the occupied orbitals a_j, the variation $\delta a_{i,1}''$ does not change the determinant wave function Ψ except, possibly, its phase and/or normalization. Consequently $\Psi_1(a_i \to \delta a_{i,1}'')$ is either zero or is proportional to Ψ, and $\langle\Psi_1(a_i \to \delta a_{i,1}'')|\hat{H} - E|\Psi\rangle = 0$, so only the term $\langle\Psi_1(a_i \to \delta a_{i,2}'')|\hat{H} - E|\Psi\rangle$ should actually be evaluated. However, $\delta a_{i,2}'' = (1 - \hat{P}^a)\delta a_i''$ is an orbital *orthogonal* to all the orbitals occupied in Ψ, and, therefore, $\langle\Psi_1(a_i \to \delta a_{i,2}'')|\Psi\rangle = 0$ according to Slater rules, and the terms of (6.177) containing the total energy E also disappear.

Similar considerations apply for the orbitals of spin β as well, so the energy variation induced by the variation $\delta\vec{R}$ in the UHF case can be expressed as

$$\begin{aligned}\delta E \ = \ & \sum_{i=1}^{n_a}\Big[\langle a_i|\delta\hat{h}|a_i\rangle + \{\langle\Psi_1(a_i \to \delta a_{i,2}'')|\hat{H}|\Psi\rangle \ + \ \text{c.c.}\ \}\Big] \\ + \ & \sum_{i=1}^{n_b}\Big[\langle b_i|\delta\hat{h}|b_i\rangle + \{\langle\Psi_1(b_i \to \delta b_{i,2}'')|\hat{H}|\Psi\rangle \ + \ \text{c.c.}\ \}\Big] \ .\end{aligned} \tag{6.182}$$

We can observe a close relationship with the technique of deriving the UHF equations through the Brillouin theorem (Section 2.1).

Now we shall discuss in some detail the term $\langle\Psi_1(a_i \to \delta a''_{i,2})|\hat{H}|\Psi\rangle$ in the right-hand side of (6.182). Expanding it by using Slater rules, one gets

$$\langle\Psi_1(a_i \to \delta a''_{i,2})|\hat{H}|\Psi\rangle = \langle\delta a''_{i,2}|\hat{F}^a|a_i\rangle \ , \tag{6.183}$$

with the usual definition (6.25) of the Fock operator, including the self-repulsion terms.

Substituting the definition of $\delta a''_{i,2}$ into the "bra," we have

$$\langle\delta a''_{i,2}| = \langle\delta a''_i|(1-\hat{P}^a) = \langle\delta a''_i|\left(1 - \sum_{k=1}^{n_a}|a_k\rangle\langle a_k|\right) \ , \tag{6.184}$$

and the right-hand side of (6.183) can be transformed to get

$$\langle\Psi_1(a_i \to \delta a''_{i,2})|\hat{H}|\Psi\rangle = \langle\delta a''_i|\hat{F}^a|a_i\rangle - \sum_{k=1}^{n_a}\langle\delta a''_i|a_k\rangle\langle a_k|\hat{F}^a|a_i\rangle \ . \tag{6.185}$$

Now, assuming that we are dealing with *canonical* UHF orbitals, for which $\langle a_k|\hat{F}^a|a_i\rangle = \varepsilon_i^a\delta_{ik}$, we obtain

$$\langle\Psi_1(a_i \to \delta a''_{i,2})|\hat{H}|\Psi\rangle = \langle\delta a''_i|\hat{F}^a|a_i\rangle - \varepsilon_i^a\langle\delta a''_i|a_i\rangle \ , \tag{6.186}$$

and a similar expression holds for the orbitals of spin β. Using this result and introducing the the P-matrices $\mathbf{P}^a, \mathbf{P}^b$ and defining the "orbital energy weighted density matrices" $\mathbf{W}^a, \mathbf{W}^b$ with the elements

$$W^a_{\mu\nu} = \sum_{i=1}^{n_a}\varepsilon_i^a a_\mu^i a_\nu^{i*} \ , \quad W^b_{\mu\nu} = \sum_{i=1}^{n_b}\varepsilon_i^b b_\mu^i b_\nu^{i*} \ , \tag{6.187}$$

and their spin-free counterparts $\mathbf{D} = \mathbf{P}^a + \mathbf{P}^b$; $\mathbf{W} = \mathbf{W}^a + \mathbf{W}^b$, we get by elementary manipulations the energy variation induced by the variation $\delta\vec{R}$

$$\begin{aligned}\delta E \ &= \sum_{\mu,\nu}\left[P^a_{\nu\mu}\langle\chi_\mu|\delta\hat{h}|\chi_\nu\rangle + \{P^a_{\nu\mu}\langle\delta\chi_\mu|\hat{F}^a|\chi_\nu\rangle - W^a_{\nu\mu}\langle\delta\chi_\mu|\chi_\nu\rangle \ + \ \text{c.c.}\}\right] \\ &+ \sum_{\mu,\nu}\left[P^b_{\nu\mu}\langle\chi_\mu|\delta\hat{h}|\chi_\nu\rangle + \{P^b_{\nu\mu}\langle\delta\chi_\mu|\hat{F}^b|\chi_\nu\rangle - W^b_{\nu\mu}\langle\delta\chi_\mu|\chi_\nu\rangle \ + \ \text{c.c.}\}\right].\end{aligned} \tag{6.188}$$

It follows from the definition of the Fock operator that

$$\langle\delta\chi_\mu|\hat{F}^a|\chi_\nu\rangle = \langle\delta\chi_\mu|\hat{h}|\chi_\nu\rangle + \sum_{\rho,\tau}\left\{D_{\tau\rho}[\delta\chi_\mu\chi_\rho|\chi_\nu\chi_\tau] - P^a_{\tau\rho}[\delta\chi_\mu\chi_\rho|\chi_\tau\chi_\nu]\right\} \ , \tag{6.189}$$

where the [12|12] convention is used for the two-electron integrals. A similar expression holds for $\langle\delta\chi_\mu|\hat{F}^b|\chi_\nu\rangle$ as well. Substituting (6.189) into

(6.188) and utilizing the symmetry of the two-electron integrals $[\chi_\mu\chi_\rho|\chi_\nu\chi_\tau] = [\chi_\rho\chi_\mu|\chi_\tau\chi_\nu]$, we get, after simple manipulations (such as decomposing a sum into two identical halves and interchanging some indices) the expression

$$\begin{aligned}\delta E &= \sum_{\mu,\nu}\Big[D_{\nu\mu}\langle\chi_\mu|\delta\hat{h}|\chi_\nu\rangle + (D_{\nu\mu}\langle\delta\chi_\mu|\hat{h}|\chi_\nu\rangle - W_{\nu\mu}\langle\delta\chi_\mu|\chi_\nu\rangle \;+\; \text{c.c.})\Big] \\ &+ \frac{1}{2}\sum_{\mu,\nu,\rho,\tau}\{([\delta\chi_\mu\chi_\rho|\chi_\nu\chi_\tau] + [\chi_\mu\delta\chi_\rho|\chi_\nu\chi_\tau])\,(D_{\nu\mu}D_{\tau\rho} \\ &\qquad -P^a_{\nu\rho}P^a_{\tau\mu} - P^b_{\nu\rho}P^b_{\tau\mu}\;) + \text{c.c.}\}\;. \end{aligned} \tag{6.190}$$

Now, we write out explicitly the complex conjugates, observe that

$$\begin{aligned}\delta S_{\mu\nu} &= \delta\langle\chi_\mu|\chi_\nu\rangle = \langle\delta\chi_\mu|\chi_\nu\rangle + \langle\chi_\mu|\delta\chi_\nu\rangle\;; \\ \delta h_{\mu\nu} &= \delta\langle\chi_\mu|\hat{h}|\chi_\nu\rangle = \langle\delta\chi_\mu|\hat{h}|\chi_\nu\rangle + \langle\chi_\mu|\delta\hat{h}|\chi_\nu\rangle + \langle\chi_\mu|\hat{h}|\delta\chi_\nu\rangle\;; \\ \delta[\chi_\mu\chi_\rho|\chi_\nu\chi_\tau] &= [\delta\chi_\mu\chi_\rho|\chi_\nu\chi_\tau] + [\chi_\mu\delta\chi_\rho|\chi_\nu\chi_\tau] \\ &\qquad + [\chi_\mu\chi_\rho|\delta\chi_\nu\chi_\tau] + [\chi_\mu\chi_\rho|\chi_\nu\delta\chi_\tau]\;, \end{aligned} \tag{6.191}$$

interchange some indices, and get the final formula for the change δE of the UHF energy induced by the change $\delta\mathbf{R}$ of the geometry:

$$\begin{aligned}\delta E &= \sum_{\mu,\nu}(D_{\nu\mu}\delta h_{\mu\nu} - W_{\nu\mu}\delta S_{\mu\nu}) \\ &+ \frac{1}{2}\sum_{\mu,\nu,\rho,\tau}(D_{\nu\mu}D_{\tau\rho} - P^a_{\nu\rho}P^a_{\tau\mu} - P^b_{\nu\rho}P^b_{\tau\mu})\delta[\chi_\mu\chi_\rho|\chi_\nu\chi_\tau]\,. \end{aligned} \tag{6.192}$$

In the RHF case this reduces simply to

$$\begin{aligned}\delta E &= \sum_{\mu,\nu}(D_{\nu\mu}\delta h_{\mu\nu} - W_{\nu\mu}\delta S_{\mu\nu}) \\ &+ \frac{1}{2}\sum_{\mu,\nu,\rho,\tau}(D_{\nu\mu}D_{\tau\rho} - \frac{1}{2}D_{\nu\rho}D_{\tau\mu})\delta[\chi_\mu\chi_\rho|\chi_\nu\chi_\tau]\,. \end{aligned} \tag{6.193}$$

Alternate derivation by using Lagrangian multipliers

The auxiliary functional F given in (6.46), which was used for deriving the UHF equations applying the method of Lagrangian multipliers, has the property that its values are always equal to those of the single determinant energy whenever the auxiliary orthonormalization condition are satisfied Therefore the gradients of the energy and of the auxiliary functional F must also coincide for that class of wave functions. When considering the "physical" variation of the orbitals induced by the variation $\delta\mathbf{R}$ of the geometry, we deal with orbitals

changing from one set of orthonormalized Hartree–Fock orbitals, corresponding to the starting geometry, to another such set that one has after the variation $\delta\mathbf{R}$. This means that the orthonormalization of the orbitals is conserved during this variation.

We shall stress that the quantity E in (6.46) represents the energy expression (6.44) derived for the case of orthonormalized orbitals. If orbital a_i undergoes variation δa_i, then the variation of this energy expression is, in accord with (6.50) $\langle \delta a_i | \hat{F}^a | a_i \rangle +$ c.c. In the present case every orbital undergoes variation because of the variation $\delta\mathbf{R}$ of the geometry, so we have to sum over all the orbitals. Also, we have to take into account that the one-electron part of the Hamiltonian also changes with the geometry. Thus we get, performing variations also in the terms containing the Lagrangian multipliers,

$$\begin{aligned} \delta E = \delta F &= \sum_{i=1}^{n_a} \left[\langle \delta a_i | \hat{F}^a | a_i \rangle - \sum_{j=1}^{n_a} \varepsilon_{ji}^a \langle \delta a_i | a_j \rangle + \text{c.c.} \right] \\ &+ \sum_{i=1}^{n_b} \left[\langle \delta b_i | \hat{F}^b | b_i \rangle - \sum_{j=1}^{n_b} \varepsilon_{ji}^b \langle \delta b_i | b_j \rangle + \text{c.c.} \right] \\ &+ \sum_{i=1}^{n_a} \langle a_i | \delta \hat{h} | a_i \rangle + \sum_{i=1}^{n_b} \langle b_i | \delta \hat{h} | b_i \rangle \,. \end{aligned} \tag{6.194}$$

Now, we may again present the variation of the orbitals in the form of two terms, one containing the variations of the orbital coefficients, the other the variations of the basis orbitals:

$$\delta a_i = \delta a_i' + \delta a_i'' = \sum_{\mu=1}^{m} \delta a_\mu^i \chi_\mu + \sum_{\mu=1}^{m} a_\mu^i \delta \chi_\mu \tag{6.195}$$

and similarly for δb_i. The variations $\delta a_i'$ and $\delta b_i'$ are those "mathematical" variations that are admitted when the HF equations are derived requiring $\delta F = 0$ at a given geometry, so they do not contribute to the gradient. Therefore, only variations $\delta a_i''$ and $\delta b_i''$ should be considered in (6.194). We may again assume that we are using canonical orbitals, i.e., $\varepsilon_{ji}^a = \varepsilon_i^a \delta_{ji}$; $\varepsilon_{ji}^b = \varepsilon_i^b \delta_{ij}$. Then we obtain

$$\begin{aligned} \delta E &= \sum_{i=1}^{n_a} \left(\langle \delta a_i'' | \hat{F}^a | a_i \rangle - \varepsilon_i^a \langle \delta a_i'' | a_i \rangle + \text{c.c.} \right) \\ &+ \sum_{i=1}^{n_b} \left(\langle \delta b_i'' | \hat{F}^b | b_i \rangle - \varepsilon_i^b \langle \delta b_i'' | b_i \rangle + \text{c.c.} \right) \\ &+ \sum_{i=1}^{n_a} \langle a_i | \delta \hat{h} | a_i \rangle + \sum_{i=1}^{n_b} \langle b_i | \delta \hat{h} | b_i \rangle \,. \end{aligned} \tag{6.196}$$

By substituting here the expansion of the orbitals and their variations, and introducing matrices $\mathbf{P}^a, \mathbf{P}^b, \mathbf{W}^a, \mathbf{W}^b$, we arrive at (6.188) .

Notes

Section 1.
Both proofs of the Brillouin theorem were published in [1]. (Also see [2].) An algorithm resembling that described in Section 1.3. has been used by Daudey [3] to obtain localized orbitals directly from a variational procedure—and not by transformation of the canonical orbitals. (A similar algorithm based on the generalized Brillouin theorem of the spin projected "extended" Hartree–Fock theory was used in [4].)

The Brillouin theorem was originally [5] considered as a consequence of the Hartree–Fock equations and it is still treated in this way by many authors.

Section 2.1.
The derivation follows largely that in [1]. The possibility of deriving the Hartree–Fock equations via the Brillouin theorem has also been discussed in [6,7], but our derivation is more general.

Section 2.2.
The derivation by using Lagrangian multipliers is the most standard one (e.g., [8]). The problems connected with the use of Lagrangian multipliers were discussed in [9] and [2].

Section 2.3.
Present author's unpublished results. (A technique of this type has been used by the author to give an alternative derivation [10] of the spin projected extended Hartree–Fock equations.)

Section 3.
Koopmans theorem has an enormous importance in understanding electronic spectra of any kind as well as ionization processes. It was proposed by Koopmans [11] (Nobel prize in *economics,* 1975). According to Koopmans' autobiography [12], the proof was provided by Kramers.

Section 4.
In the first decades Hartree–Fock method was understood in the sense of RHF. (For atoms a spherically averaged effective potential was also usually applied.) The UHF (DODS) scheme has been introduced by Pople and Nesbet [13] and Berthier [14]. (Also see the important work by Slater [15,16] who was probably the first to study the problem of "dissociation catastrophe" and the appearance of a distinct UHF solution at larger distances. For a study of the mathematical aspects of the "critical point" where the bifurcation of the RHF and UHF solutions takes place, we refer to [17]; also see [18]. Musher pointed out the possibility of pointlike discontinuities in the case of full symmetry forcing in [19].

The spin projection operator devoted to solve the "symmetry dilemma" was introduced by Löwdin [20]. It was first extensively utilized in the π-electron"Alternant Molecular Orbital Method" [21]. The spin-projected extended Hartree–Fock method [22] can be considered as a generalization of the one-electron approximation, in which the wave function is a *spin projected* Slater determinant. For a survey see [2]. (The general existence of the distinct EHF solution can be deduced from the results in [23].)

Section 5.1.
The HFR equations were independently derived by Roothaan [24] and Hall [25]. The present derivation belongs to the author [26].

Section 5.2.
The P-matrix (more precisely the "charge–bond order matrix" $\mathbf{D} = 2\mathbf{P}$) was introduced by

Coulson [27] in the framework of the simple π-electron (Hückel) theory. For a discussion of the relation between the matrix **P** and the "first order density matrix" $\rho_1(x, x')$ in the case of nonorthogonal basis sets we refer to [28].

Section 5.3.
These equations were first derived by another method [29]. The Hermitization scheme leading from Eq. (6.122) to Eq. (6.123) was proposed by Kuprievich [30].

Section 6.
The ROHF equations were first studied systematically by Roothaan [31]. Our presentation broadly follows that in [32], except the transformation leading from Eq. (6.158) to Eq. (6.160), which is similar to that used in [33] and is based on an unpublished result of Sklenar [34].

Section 7.
The derivation of the gradient described first is from [35], the second (standard) one is based on [36,37].

Suggested readings

The Hartree–Fock method has a central importance in quantum chemistry; accordingly all books consider it in detail and contain useful information about it. On a very subjective basis we could point out the classical work of Slater [38], the exhaustive monograph of Huzinaga [39] and the excellent textbooks by McWeeny [40] and by Szabo and Ostlund [41]. For the comparison of the different Hartree–Fock methods, the author's survey [2] may also be recommended. The tables of Clementi and Roetti [42] of near-Hartree–Fock wave functions of free atoms and ions calculated by using Slater-type orbitals are very useful sources of reference data.

References

1 I. Mayer, Acta Phys. Hung. **30**, 373 (1971)
2 I. Mayer, Adv. Quantum Chem. **12**, 189 (1980)
3 J.P. Daudey, Chem. Phys. Letters **24**, 574 (1974)
4 R. Lefebvre and Y.G. Smeyers, Int. J. Quantum Chem. **1**, 403 (1967)
5 L. Brillouin, *Actual. Sci. Ind.*, **71** (1933), **159** (1934)
6 J. P. Dahl, H. Johansen, D. R. Truax and T. Ziegler, Chem. Phys. Letters **6**, 64 (1970)
7 R. Lefebvre, Cahiers Phys. (Paris) **13**, 369 (1959)
8 H.A. Bethe, *Intermediate Quantum Mechanics*, Benjamin, New York 1964.
9 W.A. Goddard III., T.H. Dunning Jr. and W. J. Hunt, Chem. Phys. Letters **4**, 231 (1969)
10 I. Mayer. Acta Phys. Hung. **37**, 39 (1974)
11 T. C. Koopmans, Physica **1**, 104 (1933)
12 See the web-site *http://www.nobel.se/economics/laureates/1975/koopmans-autobio.html*
13 J.A. Pople and R.K. Nesbet, J. Chem. Phys. **22**, 571 (1954)
14 G. Berthier, J. Chim. Phys. **51**, 363 (1954)
15 J.C. Slater, Phys. Rev. **35**, 509 (1930)
16 J.C. Slater, Phys. Rev. **82**, 538 (1951)
17 I. Mayer, Acta Phys. Hung. **54**, 249 (1983)
18 P.-O. Löwdin and I. Mayer, Adv. Quantum Chem. **24**, 79 (1992)
19 J. I. Musher. Chem. Phys. Lett. **7**, 397 (1970)
20 P.-O. Löwdin, Phys. Rev. **97**, 1510 (1855)
21 R. Pauncz, *Alternant Molecular Orbital Method* Saunders, Philadelphia 1967.

22 P.-O. Löwdin, in *Quantum Theory of Atoms, Molecules, and the Solid State*, ed. P.-O. Löwdin, Academic Press, New York 1966.

23 P. Karadakov and O. Castaño, Int. J. Quant. Chem. **24**, 453 (1983)

24 C.C.J. Roothaan, Rev. Mod. Phys. **23**, 69 (1951)

25 G.G. Hall, Proc. Roy. Soc. (London) **A205**, 541 (1951)

26 I. Mayer, Acta Phys. Hung. **36**, 11 (1974)

27 C.A. Coulson, Proc. Roy. Soc. (London) **A169**, 413 (1939)

28 I. Mayer, J. Mol. Struct. (Theochem) **255**, 1 (1992)

29 I. Mayer, Acta Phys. Hung. **34**, 83 (1973)

30 V.A. Kuprievich, personal communication to the author, Kiev, 1971.

31 C.C.J. Roothaan, Rev. Mod. Phys. **32**, 179 (1960)

32 R. Carbo and O. Gropen, Adv. Quantum Chem. **12**, 159 (1980)

33 I. Mayer, J. Ladik and G. Biczó, Int. J. Quantum Chem. **7**, 583 (1973)

34 H. Sklenar, personal communication to the author, Budapest, 1972.

35 M.B. Ruiz and I. Mayer, Chem. Phys. Letters **236**, 217 (1995)

36 P. Pulay in *Modern Theoretical Chemistry* vol. 4., ed. Schaefer III, Plenum Press, New York 1977.

37 P. Pulay, Adv. Chem. Phys. **69**, 241 (1987)

38 J.C. Slater, *Quantum Theory of Molecules and Solids vol. 1. Electronic Structure of Molecules*, McGraw-Hill, New York 1963.

39 S. Fudzinaga (S. Huzinaga) *Metod Molekularnikh Orbitaley* (Russian translation from Japanese), Mir, Moscow, 1983.

40 R. McWeeny, *Methods of Molecular Quantum Mechanics*, Academic Press, London 1992.

41 A. Szabo and N.S. Ostlund *Modern Quantum Chemistry*, McGraw Hill, New York 1989.

42 E. Clementi and C. Roetti, *Roothaan–Hartree–Fock Atomic Wavefunctions*, Academic Press, New York 1974. (Atomic Data and Nuclear Data Tables **14**, 177, 1974)

Chapter 7

POPULATION ANALYSIS, BOND ORDERS, AND VALENCES

When performing quantum chemical investigations, one usually concentrates on the values of the total energy and on different quantities related to the latter, like the geometrical parameters corresponding to an energy minimum, the vibrational frequencies that are determined by the shape of the potential energy surface around the energy minimum, etc. Often one wishes to get a better understanding of the system studied by utilizing not only the energetic data but also the information contained in the wave function. However, the wave function is usually defined by too big a set of numerical data to be directly used for that purpose, and one needs a sort of "data compression" to make any interpretation of the result possible.

A very important, *measurable physical quantity* is the electron density, i.e., the distribution of the electrons in the physical space. Its detailed analysis—as done e.g., in Bader's "Atoms in molecules" theory—represents an interesting field, which is, however, out of the scope of the present book. Alternately, it is also of interest to consider the distribution of the electrons between the individual basis orbitals and, especially, between the different atoms and define in this manner the resulting charge that can be attributed to the individual atoms in the molecule. One may also be interested in different parameters characterizing the valence state of an atom in a molecule as well as the chemical bonding between the individual atoms. We are going to describe here some parameters of this type. They are often very useful in interpreting the results of calculations and may even have some predictive power. Nevertheless, one has to keep in mind that neither of quantities discussed in this chapter—except the spatial electron and spin densities—are true measurable physical quantities; they merely serve the purpose of presenting the results of calculations in a relatively simple and

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

compact manner by using terms that may be related to the genuine chemical concepts. (Neither a basis orbital nor its population is a measurable physical quantity.)

As opposed to an analysis of the electron distribution in the physical space, the methods described in this chapter give an analysis in the Hilbert space of the basis orbitals.

1. Mulliken's population analysis

1.1. Electron density

In Section 2.1 of Chapter 2 we have already given an expression of the electron density $\varrho(\vec{r})$ that one obtains by integrating over the coordinates of all the electrons but one and summing over all the electrons. The electron density at the point $\vec{r}$ can be considered as the expectation value of the operator $\hat{\varrho}(\vec{r})$, representing the sum of one-electron operators, symmetric with respect of the individual electrons:

$$\hat{\varrho}(\vec{r}) = \sum_{i=1}^{N} \delta(\vec{r} - \vec{r}_i) \ . \tag{7.1}$$

As an electron has a negative charge, the electron density $\varrho(\vec{r})$ corresponds to an electrostatic charge density $-\varrho(\vec{r})$ if atomic units are used. (Otherwise one has to write $-e_0\varrho(\vec{r})$, where e_0 is the elementary charge.)

If the many-electron wave function is a Slater determinant $\Psi = \hat{\mathcal{A}}[\psi_1(1)\psi_2(2)\ldots\psi_N(N)]$ built up of orthonormalized spin orbitals, then we may use the Slater rule in (5.100), and—similarly to other one-electron quantities—the electron density $\varrho(\vec{r})$ can be presented as the sum over the individual occupied spin orbitals $\psi_i(\vec{r},\sigma) = \varphi_i(\vec{r})\gamma(\sigma)$ as

$$\begin{aligned} \varrho(\vec{r}) &= \langle\Psi|\hat{\varrho}(\vec{r})|\Psi\rangle = \sum_{j=1}^{N}\langle\psi_j(\vec{r}_1,\sigma_1)|\delta(\vec{r}-\vec{r}_1)|\psi_j(\vec{r}_1,\sigma_1)\rangle \\ &= \sum_{j=1}^{N}\langle\varphi_j(\vec{r}_1)|\delta(\vec{r}-\vec{r}_1)|\varphi_j(\vec{r}_1)\rangle = \sum_{j=1}^{N}\varphi_j^*(\vec{r})\varphi_j(\vec{r}) = \sum_{j=1}^{N}|\varphi_j(\vec{r})|^2 \ . \end{aligned} \tag{7.2}$$

If, in particular, Ψ contains n_a orbitals a_i occupied with spin α and n_b orbitals b_i occupied with spin β, (7.2) becomes

$$\varrho(\vec{r}) = \sum_{j=1}^{n_a}|a_j(\vec{r})|^2 + \sum_{j=1}^{n_b}|b_j(\vec{r})|^2 \ . \tag{7.3}$$

Of course, the electron density is invariant under the unitary transformations of the sets of orbitals $\{a_i\}$ or $\{b_i\}$ as they leave invariant the wave function Ψ.

For instance, if

$$a_j(\vec{r}) = \sum_{k=1}^{n_a} U_{kj} a'_k(\vec{r}) \ , \tag{7.4}$$

and matrix $\mathbf{U}$ is uniter ($\mathbf{U}^\dagger = \mathbf{U}^{-1}$), then

$$\begin{aligned}
\sum_{j=1}^{n_a} |a_j(\vec{r})|^2 &= \sum_{j=1}^{n_a} a_j^*(\vec{r}) a_j(\vec{r}) = \sum_{j=1}^{n_a} \sum_{k=1}^{n_a} U_{kj}^* a_k'^*(\vec{r}) \sum_{l=1}^{n_a} U_{lj} a'_l(\vec{r}) \\
&= \sum_{j,k,l=1}^{n_a} a_k'^*(\vec{r}) a'_l(\vec{r}) U_{lj} (\mathbf{U}^\dagger)_{jk} = \sum_{k,l=1}^{n_a} a_k'^*(\vec{r}) a'_l(\vec{r}) \delta_{kl} \\
&= \sum_{k=1}^{n_a} a_k'^*(\vec{r}) a'_k(\vec{r}) = \sum_{k=1}^{n_a} |a'_k(\vec{r})|^2 \ . \qquad (7.5)
\end{aligned}$$

Q.E.D.

According to the previous results, the electron density corresponding to a determinant wave function behaves like that of a set of independent electrons occupying orthonormalized orbitals. However, we shall stress again that one should not attribute too much physical meaning to these *individual* orbitals because any set of orthonormalized orbitals spanning the given occupied subspace will exhibit the same behavior and give the same resulting electron density. (The same holds for any other one-electron property, as well.)

If one uses an LCAO expansion of the type (6.90), (6.115) then the electron density can be expressed also in terms of the basis orbitals χ_μ:

$$\begin{aligned}
\varrho(\vec{r}) &= \sum_{j=1}^{n_a} a_j^*(\vec{r}) a_j(\vec{r}) + \sum_{j=1}^{n_b} b_j^*(\vec{r}) b_j(\vec{r}) \qquad (7.6) \\
&= \sum_{j=1}^{n_a} \sum_{\nu=1}^{m} a_\nu^{j*} \chi_\nu^*(\vec{r}) \sum_{\mu=1}^{m} a_\mu^j \chi_\mu(\vec{r}) + \sum_{j=1}^{n_b} \sum_{\nu=1}^{m} b_\nu^{j*} \chi_\nu^*(\vec{r}) \sum_{\mu=1}^{m} b_\mu^j \chi_\mu(\vec{r}) \ .
\end{aligned}$$

By using the elements of the P-matrices defined in (6.116), this may be rewritten as

$$\varrho(\vec{r}) = \sum_{\mu,\nu=1}^{m} (P_{\mu\nu}^a + P_{\mu\nu}^b) \chi_\nu^*(\vec{r}) \chi_\mu(\vec{r}) = \sum_{\mu,\nu=1}^{m} D_{\mu\nu} \chi_\nu^*(\vec{r}) \chi_\mu(\vec{r}) \ . \tag{7.7}$$

We have introduced the "spinless density matrix" $\mathbf{D}$:

$$\mathbf{D} = \mathbf{P}^a + \mathbf{P}^b \ . \tag{7.8}$$

In the RHF case this definition coincides with that in (6.110). Equation (7.7) obviously gives the sum of the densities that one can assign to the electrons

with spin α and β, respectively. Analogously, one can define the spin density $\varrho_s(\vec{r})$ as the difference of these densities:

$$\varrho_s(\vec{r}) = \sum_{\mu,\nu=1}^{m} (P^a_{\mu\nu} - P^b_{\mu\nu})\chi^*_\nu(\vec{r})\chi_\mu(\vec{r}) = \sum_{\mu,\nu=1}^{m} P^s_{\mu\nu}\chi^*_\nu(\vec{r})\chi_\mu(\vec{r}) \ , \tag{7.9}$$

where

$$\mathbf{P}^s = \mathbf{P}^a - \mathbf{P}^b \tag{7.10}$$

is the "spin-density matrix."

If one integrates the electron density $\varrho(\vec{r})$ over the whole space, then the number of electrons $N = n_a + n_b$ is recovered. This follows from the fact that orbitals a_i and b_i are normalized, so each term in the sums in (7.3) contributes unity to the integral $\int \varrho(\vec{r})dv$. By using the expansion in (7.7), the integral of the electron density can also be expressed through the traces of matrices $\mathbf{P}^a\mathbf{S}$, $\mathbf{P}^b\mathbf{S}$, and $\mathbf{DS}$:

$$\begin{aligned}
\sum_{\mu,\nu=1}^{m} P^a_{\mu\nu}S_{\nu\mu} &= \sum_{\mu=1}^{m}(\mathbf{P}^a\mathbf{S})_{\mu\mu} = \mathrm{Tr}(\mathbf{P}^a\mathbf{S}) = n_a \\
\sum_{\mu,\nu=1}^{m} P^b_{\mu\nu}S_{\nu\mu} &= \sum_{\mu=1}^{m}(\mathbf{P}^b\mathbf{S})_{\mu\mu} = \mathrm{Tr}(\mathbf{P}^b\mathbf{S}) = n_b \\
\sum_{\mu,\nu=1}^{m} D_{\mu\nu}S_{\nu\mu} &= \sum_{\mu=1}^{m}(\mathbf{DS})_{\mu\mu} = \mathrm{Tr}(\mathbf{DS}) = N \ .
\end{aligned} \tag{7.11}$$

We should note that the expansion in (7.7) and the equalities in (7.11) are valid for any N-electron wave function with n_a electrons of spin α and n_b electrons of spin β—i.e., with the resulting spin projection $S_z = \frac{1}{2}(n_a - n_b)$—because the expansion in (7.7) has the most general possible form, provided that the wave function is built up of the m basis orbitals χ_μ. However, the idempotency properties $(\mathbf{P}^a\mathbf{S})^2 = \mathbf{P}^a\mathbf{S}$; $(\mathbf{P}^b\mathbf{S})^2 = \mathbf{P}^b\mathbf{S}$ hold only if the wave function is a single determinant.

1.2. Population analysis

According to (7.11) the trace of matrix $\mathbf{DS}$ equals the number of electrons in the system, and this equation defines a distribution of the electrons between the basis orbitals χ_μ. In accord with this, the quantity

$$q_\mu = (\mathbf{DS})_{\mu\mu} = \sum_{\nu=1}^{m} D_{\mu\nu}S_{\nu\mu} \tag{7.12}$$

is called "Mulliken's gross orbital population" of the basis orbital χ_μ, and the q_μs sum to the number of electrons.

One can consider q_μ as consisting of terms of two types:

$$q_\mu = D_{\mu\mu} + \sum_{\substack{\nu=1 \\ (\nu\neq\mu)}}^{m} D_{\mu\nu} S_{\nu\mu} \,. \tag{7.13}$$

(The basis functions are assumed normalized $S_{\mu\mu} = 1$.)

The first term is called "Mulliken's net orbital population" of orbital χ_μ; it comes from integrating the term $D_{\mu\mu}\chi_\mu^*(\vec{r})\chi_\mu(\vec{r})$ in the expansion in (7.7). The second term comes from integrating the terms $D_{\mu\nu}\chi_\nu^*(\vec{r})\chi_\mu(\vec{r})$ with $\nu \neq \mu$. It is connected with the existence of an "overlap density" $\chi_\nu^*(\vec{r})\chi_\mu(\vec{r})$ that does not integrate to 0, which would be the case if the basis set were orthogonal.

The given pair of orbitals χ_μ and χ_ν give the contribution $\chi_\nu^*(\vec{r})\chi_\mu(\vec{r}) + \chi_\mu^*(\vec{r})\chi_\nu(\vec{r})$ to the total electron density in (7.7). The first of these terms is taken into account when computing the gross orbital population q_μ, whereas the second contributes to the gross orbital population q_ν. In the case of real orbitals these two terms are equal, and this is the reason why one often speaks about the "halving" of the total overlap density $2\chi_\mu(\vec{r})\chi_\nu(\vec{r})$ and the corresponding "overlap populations" $2D_{\mu\nu}S_{\mu\nu}$ when computing Mulliken's gross populations q_μ, q_ν. (If the orbitals are real, $D_{\mu\nu} = D_{\nu\mu}$, as well.) The present author cannot agree with the opinion that this halving is "arbitrary"; it follows from (7.11) and therefore it is the only possibility that is consistent with the mathematical structure of the formalism of using overlapping basis orbitals (matrices $\mathbf{P}^a\mathbf{S}$ and $\mathbf{P}^b\mathbf{S}$ are the idempotent projection matrices).

If one applies a basis set each orbital of which is centered on a given atom (as is common), then one may sum the gross orbital populations q_μ for all the orbitals assigned to a given atom A (this assignment will be denoted as $\mu \in A$) and get Mulliken's gross atomic population of that atom:

$$Q_A = \sum_{\mu\in A} q_\mu = \sum_{\mu\in A} (\mathbf{DS})_{\mu\mu} = \sum_{\mu\in A}\sum_{\nu=1}^{m} D_{\mu\nu}S_{\nu\mu} \,. \tag{7.14}$$

The gross atomic population Q_A can naturally be decomposed as

$$Q_A = d_A + \sum_{\substack{B \\ (B\neq A)}} d_{AB} \,, \tag{7.15}$$

where the quantity d_A containing the terms in which both indices correspond to atom A

$$d_A = \sum_{\mu,\nu\in A} D_{\mu\nu}S_{\nu\mu} \tag{7.16}$$

is called Mulliken's net atomic population, whereas the quantity of diatomic nature

$$d_{AB} = \sum_{\mu \in A} \sum_{\nu \in B} D_{\mu\nu} S_{\nu\mu} \tag{7.17}$$

is called "Mulliken's overlap population." (Often this name is used for the quantity $d_{AB} + d_{BA}$, which is equal to $2d_{AB}$ if real orbitals are used.) It is to be stressed that the net atomic population d_A is not simply the sum of the net orbital populations but it also contains the overlap populations of all pairs of the orbitals that both are centered on the same atom.

Mulliken's overlap population is a significant positive number (say of order $\sim$0.5) for chemically bonded pairs of atoms and reflects the accumulation of the electronic charge in the bonding region. Its value usually correlates well with the strength of the bond between the atoms involved. From the other side, a negative value of the overlap population between chemically nonbonded atoms indicates an antibonding situation and makes probable a repulsive interaction.

Invariance

In practice one usually applies basis orbitals that are centered on the given atom and are oriented according to the external "laboratory" coordinate frame—e.g., one uses orbitals p_x, p_y, p_z that are oriented along the axes x, y, and z. Obviously, if one rotates the molecule with respect to the laboratory frame, then the expansion coefficients of each molecular orbital will change, although the molecular wave function remains essentially the same. This is especially evident if one considers this process as the laboratory frame being rotated around the molecule. The change of the orbital coefficients results in the change of the matrix **D**, as well, although the physical situation—the state of the molecule—remains invariant. (In particular, only the D-matrix elements between orbitals of s-type remain the same; all the others will change under rotations.)

One often uses the concept of "hybrid atomic orbitals" representing linear combinations of the basis orbitals of type s and p (sometimes also d) for the discussion of different chemical problems, in particular the directional character of the chemical bonds. Obviously, turning to such a hybrid basis will not change any relevant results of the calculations.

According to the previous discussion, one may use for interpretations only quantities that have "rotational–hybridizational invariance", i.e., remain the same if one subjects the orbitals centered on a given atom to any nonsingular linear transformation. (For that reason neither the individual D-matrix elements nor their sum for a given pair of atoms can be assigned any significance—contrary the old π-electron theories in which the off-diagonal elements $D_{\mu\nu}$ were treated as "bond orders.")

We are going to prove here the rotational–hybridizational invariance of Mulliken's net and gross atomic populations of a given atom and of Mulliken's

overlap population of a given pair of atoms. For that reason we consider non-singular linear transformations between the basis orbitals centered on the same atom:

$$\chi'_\mu = \sum_{\nu \in A} T^A_{\nu\mu} \chi_\nu \qquad (\mu \in A) \,. \tag{7.18}$$

One can treat the rotational–hybridizational transformations (7.18) of the basis orbitals of all atoms simultaneously by introducing the block-diagonal transformation matrix $\mathbf{T}$, the diagonal blocks of which are the transformation matrices $\mathbf{T}^A$ of the individual atoms:

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{T}^2 & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{T}^M \end{pmatrix} \tag{7.19}$$

where M is the number of the atoms in the system. Using matrix $\mathbf{T}$, the transformation of the whole basis can compactly be defined as

$$\chi'_\mu = \sum_{\nu=1}^{m} T_{\nu\mu} \chi_\nu \qquad \mu = 1, 2, \dots, m \,. \tag{7.20}$$

As every transformation $\mathbf{T}^A$ has been assumed to be nonsingular, the same holds for matrix $\mathbf{T}$; the inverse matrix $\mathbf{T}^{-1}$ is again block-diagonal and its diagonal blocks are the inverse transformation matrices $(\mathbf{T}^A)^{-1}$, corresponding to the individual atoms. The inverse matrix $\mathbf{T}^{-1}$ permits one to define the inverse of the transformation in (7.20):

$$\chi_\mu = \sum_{\nu=1}^{m} (\mathbf{T}^{-1})_{\nu\mu} \chi'_\nu \,. \tag{7.21}$$

Thus each molecular orbital $a_i = \sum_{\mu=1}^{m} a^i_\mu \chi_\mu$ can also be expressed in terms of the new basis orbitals:

$$\begin{aligned} a_i &= \sum_{\mu=1}^{m} a^i_\mu \chi_\mu = \sum_{\mu=1}^{m} a^i_\mu \sum_{\nu=1}^{m} (\mathbf{T}^{-1})_{\nu\mu} \chi'_\nu \\ &= \sum_{\nu=1}^{m} \left(\sum_{\mu=1}^{m} (\mathbf{T}^{-1})_{\nu\mu} a^i_\mu \right) \chi'_\nu = \sum_{\nu=1}^{m} a^{i\prime}_\nu \chi'_\nu \end{aligned} \tag{7.22}$$

i.e., the vectors of the orbital coefficients transform by the inverse of matrix $\mathbf{T}$:

$$\mathbf{a}^{i\prime} = \mathbf{T}^{-1} \mathbf{a}^i \tag{7.23}$$

and similarly for the orbitals of spin β.

According to these considerations, matrix $\mathbf{D}$ transforms as

$$\begin{aligned}\mathbf{D}' = \mathbf{P}^{a\prime} + \mathbf{P}^{b\prime} &= \sum_{i=1}^{n_a} \mathbf{a}^{i\prime}(\mathbf{a}^{i\prime})^\dagger + \sum_{i=1}^{n_b} \mathbf{b}^{i\prime}(\mathbf{b}^{i\prime})^\dagger \\ &= \sum_{i=1}^{n_a} \mathbf{T}^{-1}\mathbf{a}^i\mathbf{a}^{i\dagger}\mathbf{T}^{-1\dagger} + \sum_{i=1}^{n_b} \mathbf{T}^{-1}\mathbf{b}^i\mathbf{b}^{i\dagger}\mathbf{T}^{-1\dagger} \\ &= \mathbf{T}^{-1}\mathbf{D}\mathbf{T}^{-1\dagger}\,, \end{aligned} \tag{7.24}$$

where the identity $(\mathbf{Ab})^\dagger = \mathbf{b}^\dagger\mathbf{A}^\dagger$ has been utilized.

The overlap matrix transforms as

$$\begin{aligned} S'_{\mu\nu} &= \langle \chi'_\mu | \chi'_\nu \rangle = \langle \sum_{\varrho=1}^{m} T_{\varrho\mu}\chi_\varrho | \sum_{\tau=1}^{m} T_{\tau\nu}\chi_\tau \rangle \\ &= \sum_{\varrho,\tau=1}^{m} T^*_{\varrho\mu} S_{\varrho\tau} T_{\tau\nu} = \sum_{\varrho,\tau=1}^{m} (\mathbf{T}^\dagger)_{\mu\varrho} S_{\varrho\tau} T_{\tau\nu} \end{aligned} \tag{7.25}$$

i.e.,

$$\mathbf{S}' = \mathbf{T}^\dagger\mathbf{S}\mathbf{T}\,. \tag{7.26}$$

The net atomic population d_A and the interatomic overlap populations d_{AB} are invariant with respect to the rotational–hybridizational transformations described by matrix $\mathbf{T}$; consequently Mulliken's gross atomic population representing their sum according to (7.15) is invariant as well.

Proof

We have for the net atomic population in the transformed basis

$$\begin{aligned} d'_A &= \sum_{\mu,\nu\in A} D'_{\mu\nu} S'_{\nu\mu} = \sum_{\mu,\nu\in A} (\mathbf{T}^{-1}\mathbf{D}\mathbf{T}^{-1\dagger})_{\mu\nu} (\mathbf{T}^\dagger\mathbf{S}\mathbf{T})_{\nu\mu} \\ &= \sum_{\kappa,\lambda,\varrho,\tau=1}^{m} \sum_{\mu,\nu\in A} (\mathbf{T}^{-1})_{\mu\kappa} D_{\kappa\lambda} (\mathbf{T}^{-1\dagger})_{\lambda\nu} (\mathbf{T}^\dagger)_{\nu\varrho} S_{\varrho\tau} T_{\tau\mu}\,. \end{aligned} \tag{7.27}$$

Owing to the block-diagonality of matrices $\mathbf{T}$, $\mathbf{T}^{-1}$ and $\mathbf{T}^\dagger$, $\mathbf{T}^{-1\dagger}$, one has by summing over μ and ν,

$$\sum_{\mu\in A} T_{\tau\mu} (\mathbf{T}^{-1})_{\mu\kappa} = \begin{cases} \delta_{\tau\kappa} & \text{if } \tau,\kappa \in A \\ 0 & \text{otherwise} \end{cases} \tag{7.28}$$

and

$$\sum_{\nu\in A} (\mathbf{T}^{-1\dagger})_{\lambda\nu} (\mathbf{T}^\dagger)_{\nu\varrho} = \begin{cases} \delta_{\lambda\varrho} & \text{if } \lambda,\varrho \in A \\ 0 & \text{otherwise} \end{cases} \tag{7.29}$$

respectively. Therefore, the summations over κ, λ, ϱ, and τ can be restricted to the indices belonging to atom A. We get accordingly

$$d'_A = \sum_{\kappa,\lambda,\varrho,\tau\in A} \delta_{\tau\kappa} D_{\kappa\lambda} \delta_{\lambda\varrho} S_{\varrho\tau} = \sum_{\tau,\varrho\in A} D_{\tau\varrho} S_{\varrho\tau} = d_A \ , \tag{7.30}$$

i.e., the net atomic population is invariant. The overlap population between atoms A and B is invariant, as well. We have, similarly to the previous case

$$\begin{aligned} d'_{AB} &= \sum_{\mu\in A} \sum_{\nu\in B} D'_{\mu\nu} S'_{\nu\mu} \qquad (7.31) \\ &= \sum_{\kappa,\lambda,\varrho,\tau=1}^{m} \sum_{\mu\in A} \sum_{\nu\in B} (\mathbf{T}^{-1})_{\mu\kappa} D_{\kappa\lambda} (\mathbf{T}^{-1\dagger})_{\lambda\nu} (\mathbf{T}^{\dagger})_{\nu\varrho} S_{\varrho\tau} T_{\tau\mu} \ . \end{aligned}$$

For the sum over μ we have again (7.28), whereas for the sum over ν we get (7.29) but with A being replaced by B. Thus the summation over κ and τ should be restricted to atom A whereas the sums for λ and ϱ are restricted to atom B:

$$d'_{AB} = \sum_{\kappa,\tau\in A} \sum_{\lambda,\varrho\in B} \delta_{\tau\kappa} D_{\kappa\lambda} \delta_{\lambda\varrho} S_{\varrho\tau} = \sum_{\tau\in A} \sum_{\varrho\in B} D_{\tau\varrho} S_{\varrho\tau} = d_{AB} \ . \tag{7.32}$$

Q.E.D.

One may conclude that Mulliken's overlap population is the simplest quantity that is linear in the interatomic D-matrix elements and has the correct rotational–hybridizational invariance.

2. Bond orders and valences

2.1. Bond order index

Although Mulliken's overlap population is a parameter that permits one to identify chemically bonded atoms, it cannot be put directly into correspondence with such a chemical notion as the bond order (multiplicity) of a chemical bond—it does not have the value of (about) 1, 2, and 3 for a single, double, and triple chemical bond, respectively. Classically the bond order between the atoms of a diatomic molecule can be defined as $\frac{1}{2}(N_{\text{bonding}} - N_{\text{antibonding}})$, where N_{bonding} and $N_{\text{antibonding}}$ are the number of electrons occupying bonding and antibonding orbitals, respectively. Thus the molecular ion H_2^+ is characterized by a bond order of $\frac{1}{2}$ (one electron on a bonding orbital) and molecule H_2 has a bond order of 1 (2 electrons on a bonding MO), whereas the helium dimer has a zero bond order. One can describe He_2 either by using two doubly occupied localized atomic $1s$ orbitals or by using the sum and difference of them, representing delocalized MOs of bonding and antibonding character,

respectively. As both are doubly occupied, $N_{\text{bonding}} = N_{\text{antibonding}} = 2$ and one gets a bond order 0 between the two atoms, in full accord with the chemical expectations. Similarly, one gets, for instance, a bond order 3 for N_2 and 1 for F_2. Obviously, such a counting of electrons is strictly possible for a diatomics only.

In the framework of the simple Hückel theory (and, in general, in the π-electron theories) the off-diagonal elements of the matrix $\mathbf{D}$ were called (Coulson's) bond order. This definition, however, could not be generalized for theories in which a given atom bears more than one basis orbital. The first quantum chemical quantity that could be put in correspondence with the classical chemical notion of the bond multiplicity for a multiatomic molecule and is computable from the wave function was the so-called Wiberg index. It was originally introduced in the framework of the semiempirical CNDO theory and is applicable only if the basis orbitals are orthonormalized.

The Wiberg index between atoms A and B of a closed shell-system is defined as

$$W_{AB} = \sum_{\mu \in A} \sum_{\nu \in B} |D_{\mu\nu}|^2 \,. \tag{7.33}$$

The adequate generalization of this formula for an open-shell system treated at the single-determinant UHF level is

$$\begin{aligned} W_{AB} &= 2 \sum_{\mu \in A} \sum_{\nu \in B} (|P^a_{\mu\nu}|^2 + |P^b_{\mu\nu}|^2) \\ &= \sum_{\mu \in A} \sum_{\nu \in B} (|D_{\mu\nu}|^2 + |P^s_{\mu\nu}|^2) \,, \end{aligned} \tag{7.34}$$

where matrices $\mathbf{D}$, $\mathbf{P}^a$, $\mathbf{P}^b$, and $\mathbf{P}^s$ are defined by (7.8), (6.116) and (7.10). The equivalence of the two forms given in (7.34) can be checked trivially by using the equalities $\mathbf{D} = \mathbf{P}^a + \mathbf{P}^b$; $\mathbf{P}^s = \mathbf{P}^a - \mathbf{P}^b$.

By writing $|D_{\mu\nu}|^2$ as $D_{\mu\nu} D_{\nu\mu}$, etc., and using (7.24), one can easily see that the Wiberg index is invariant with respect to unitary transformations between the orbitals centered on the same atom—as those induced by the rotations of the coordinate frame (then $\mathbf{T}^{-1\dagger} = \mathbf{T}$)—but not with respect to a general (nonsingular but nonunitary) transformation. Therefore the Wiberg index as defined previously is not applicable in the *ab initio* theory, in which invariance with respect to such transformations should be required, as well. (The basis orbitals may be nonorthogonal.) The proper generalization of the Wiberg index for the *ab initio* theory has been introduced by the present author. (An analogous definition was proposed by Giambiagi in the framework of the semiempirical "extended Hückel theory" also applying overlapping basis sets.)

The respective definitions of the bond order index are

$$B_{AB} = \sum_{\mu\in A}\sum_{\nu\in B}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu} \tag{7.35}$$

for the closed-shell (RHF) case and

$$\begin{aligned} B_{AB} &= 2\sum_{\mu\in A}\sum_{\nu\in B}[(\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + (\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu}] \\ &= \sum_{\mu\in A}\sum_{\nu\in B}[(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu} + (\mathbf{P}^s\mathbf{S})_{\mu\nu}(\mathbf{P}^s\mathbf{S})_{\nu\mu}] \end{aligned} \tag{7.36}$$

for the open-shell (UHF) one. Again, the equivalence of the two forms in (7.36) can be checked by using the definitions $\mathbf{D} = \mathbf{P}^a + \mathbf{P}^b$, $\mathbf{P}^s = \mathbf{P}^a - \mathbf{P}^b$. Obviously, definition (7.36) reduces to the previous one if applied to a closed-shell RHF wave function ($\mathbf{P}^s$=**0**). Furthermore, if the basis set is orthonormalized (**S**=**1**), then one has $B_{AB} = W_{AB}$. We shall also note that these definitions remain valid also for correlated wave functions. In that case the definition of matrices $\mathbf{P}^a$ and $\mathbf{P}^b$ or $\mathbf{D}$ and $\mathbf{P}^s$ via the orbital coefficients is not applicable, of course, but these matrices may be identified as the matrices of expansion coefficients of electron density and spin density in (7.7) and (7.9).

Invariance

The bond order index defined by (7.35) for the closed-shell RHF case is the simplest quantity that is quadratic in the interatomic elements of matrix $\mathbf{D}$ and has the correct rotational–hybridizational invariance. Similar considerations hold for definition (7.36) applicable in the UHF case involving two matrices $\mathbf{P}^a$ and $\mathbf{P}^b$ or $\mathbf{D}$ and $\mathbf{P}^s$. We shall prove the invariance of definition (7.35); the matrices in (7.36) behave in exactly the same manner.

Thus we have to study the bond order index in (7.35) and establish its invariance under the transformation in (7.18) involving the block-diagonal matrix in (7.19). We have after the transformation

$$B'_{AB} = \sum_{\mu\in A}\sum_{\nu\in B}(\mathbf{D}'\mathbf{S}')_{\mu\nu}(\mathbf{D}'\mathbf{S}')_{\nu\mu}\ , \tag{7.37}$$

where the transformed matrices $\mathbf{D}'$ and $\mathbf{S}'$ are given by (7.24) and (7.26), respectively. Accordingly, we have for their product

$$\mathbf{D}'\mathbf{S}' = \mathbf{T}^{-1}\mathbf{D}\mathbf{T}^{-1\dagger}\mathbf{T}^{\dagger}\mathbf{S}\mathbf{T} = \mathbf{T}^{-1}\mathbf{D}\mathbf{S}\mathbf{T} \tag{7.38}$$

and

$$B'_{AB} = \sum_{\kappa,\lambda,\varrho,\tau=1}^{m}\sum_{\mu\in A}\sum_{\nu\in B}(\mathbf{T}^{-1})_{\mu\kappa}(\mathbf{DS})_{\kappa\lambda}T_{\lambda\nu}(\mathbf{T}^{-1})_{\nu\varrho}(\mathbf{DS})_{\varrho\tau}T_{\tau\mu}\,. \tag{7.39}$$

Similar to the case of the overlap population d_{AB}, we again should use (7.28) and (7.29), replacing in the latter A with B. The sums over κ and τ should again be restricted to atom A, and those for λ and ϱ to atom B:

$$\begin{aligned} B'_{AB} &= \sum_{\kappa,\tau\in A}\sum_{\lambda,\varrho\in B} \delta_{\tau\kappa}(\mathbf{DS})_{\kappa\lambda}\delta_{\lambda\varrho}(\mathbf{DS})_{\varrho\tau} \\ &= \sum_{\tau\in A}\sum_{\varrho\in B}(\mathbf{DS})_{\tau\varrho}(\mathbf{DS})_{\varrho\tau} = B_{AB} \end{aligned} \tag{7.40}$$

Q.E.D.

2.2. Valence indices

Following Wiberg, it is worthwhile to define the quantity $b_\mu = 2q_\mu - q_\mu^2$, where q_μ is Mulliken's gross orbital population of the basis orbital χ_μ. This quantity can be considered as a measure of the actual bonding ability of orbital χ_μ in the molecule: it has a maximum for $q_\mu = 1$, characterizing an orbital participating in a pure covalent bond and vanishes for both an empty orbital having no role in the molecule ($q_\mu = 0$) and for a doubly filled one ($q_\mu = 2$) having no chemical significance (e.g., a core orbital). If one sums the quantities b_μ for all the basis orbitals centered on a given atom, but subtracts the chemically irrelevant intra-atomic partial bond orders, then one gets the definition of the actual *valence* of an atom in the molecule:

$$V_A = 2Q_a - \sum_{\mu,\nu\in A}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu}\ , \tag{7.41}$$

where Q_A is Mulliken's gross atomic population of atom A.

In the closed-shell RHF case one has for matrix $\mathbf{P}$ the idempotency property $(\mathbf{PS})^2 = \mathbf{PS}$, which leads to the equality $(\mathbf{DS})^2 = 2\mathbf{DS}$ for the matrix $\mathbf{D} = 2\mathbf{P}$. As a consequence, one obtains the equality

$$V_A = \sum_{\substack{B \\ (B\neq A)}} B_{AB}\ , \tag{7.42}$$

i.e., if a closed-shell determinant wave function is used, then the valence of an atom is equal to the sum of its bond orders formed with all the other atoms.

In fact, by substituting $Q_A = \sum_{\mu\in A} q_\mu = \sum_{\mu\in A}(\mathbf{DS})_{\mu\mu}$, one gets

$$V_A = \sum_{\mu\in A} 2(\mathbf{DS})_{\mu\mu} - \sum_{\mu,\nu\in A}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu}\ . \tag{7.43}$$

It follows from the idempotency property mentioned previously that

$$2(\mathbf{DS})_{\mu\mu} = [(\mathbf{DS})^2]_{\mu\mu} = \sum_{\nu=1}^{m} (\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu} \,. \tag{7.44}$$

Substituting this into Eq. (7.43) we have, separating out terms with $\nu \in A$ and grouping the remaining ones by the respective atoms,

$$\begin{aligned} V_A &= \sum_{\mu\in A}\left[\sum_{\nu\in A}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu} + \sum_{\substack{B\\ B\neq A}}\sum_{\nu\in B}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu}\right] \\ &\quad - \sum_{\mu,\nu\in A}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu} = \sum_{\substack{B\\ B\neq A}}\sum_{\mu\in A}\sum_{\nu\in B}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu} \\ &= \sum_{\substack{B\\ (B\neq A)}} B_{AB} \,. \end{aligned} \tag{7.45}$$

Q.E.D.

Equality (7.45) does not hold for an open-shell (UHF) wave function or if a correlated wave function is used. In these cases the difference between the actual total valence of the atom and the sum of its bond orders is defined as the *free valence* of the atom

$$F_A = V_A - \sum_{\substack{B\\ (B\neq A)}} B_{AB} \,. \tag{7.46}$$

The valence index V_A may be considered as a quantum chemical counterpart of the chemist's concept of the actual covalent valence of an atom, and it usually gives values close to the integer numbers corresponding to the classical chemical picture of molecules. In the case of free radicals, biradicals, etc., the free valence index can be put in a close correspondence to the "dots" used by chemists to denote the radical centers on the chemical formulae. (More generally, significant values of the free valence index indicate the atoms where radical attack may be expected in the molecule.)

The previous interpretation of the free valence index is supported by the fact that in the UHF case the free valence index can be expressed through the spin-density matrix $\mathbf{P}^s$ as

$$F_A = \sum_{\mu,\nu\in A}(\mathbf{P}^s\mathbf{S})_{\mu\nu}(\mathbf{P}^s\mathbf{S})_{\nu\mu} \,. \tag{7.47}$$

To derive (7.47), first we shall get a transformed expression for the quantity $2Q_A$ occurring in the definition (7.41) of V_A, by using the equality $\mathbf{D} = \mathbf{P}^a + \mathbf{P}^b$ and the idempotency properties $(\mathbf{P}^a\mathbf{S})^2 = \mathbf{P}^a\mathbf{S}$, $(\mathbf{P}^b\mathbf{S})^2 = \mathbf{P}^b\mathbf{S}$:

$$
\begin{aligned}
2Q_A &= 2\sum_{\mu\in A}(\mathbf{DS})_{\mu\mu} = 2\sum_{\mu\in A}[(\mathbf{P}^a\mathbf{S})_{\mu\mu} + (\mathbf{P}^b\mathbf{S})_{\mu\mu}] \\
&= 2\sum_{\mu\in A}\left\{[(\mathbf{P}^a\mathbf{S})^2]_{\mu\mu} + [(\mathbf{P}^b\mathbf{S})^2]_{\mu\mu}\right\} \\
&= 2\sum_{\mu\in A}\sum_{\nu=1}^{m}[(\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + (\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu}] \qquad (7.48) \\
&= 2\sum_{\mu,\nu\in A}[(\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + (\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu}] \\
&\quad + 2\sum_{\mu\in A}\sum_{\substack{B\\(B\neq A)}}\sum_{\nu\in B}[(\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + (\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu}] \\
&= 2\sum_{\mu,\nu\in A}[(\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + (\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu}] + \sum_{\substack{B\\(B\neq A)}} B_{AB}\,,
\end{aligned}
$$

where the definition in (7.36) of the bond order index B_{AB} has also been utilized. By substituting the definition in (7.41) into (7.46), we have for the free valence index

$$
\begin{aligned}
F_A &= V_A - \sum_{\substack{B\\(B\neq A)}} B_{AB} = 2Q_A - \sum_{\mu,\nu\in A}(\mathbf{DS})_{\mu\nu}(\mathbf{DS})_{\nu\mu} - \sum_{\substack{B\\(B\neq A)}} B_{AB} \\
&= 2Q_A - \sum_{\mu,\nu\in A}[(\mathbf{P}^a + \mathbf{P}^b)\mathbf{S}]_{\mu\nu}[(\mathbf{P}^a + \mathbf{P}^b)\mathbf{S}]_{\nu\mu} - \sum_{\substack{B\\(B\neq A)}} B_{AB}\,.
\end{aligned} \qquad (7.49)
$$

Substituting here expression (7.48) we get (terms containing B_{AB} cancel):

$$
\begin{aligned}
F_A &= \sum_{\mu,\nu\in A}\{2(\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + 2(\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu} \\
&\quad -[(\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + (\mathbf{P}^a\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu} \qquad (7.50) \\
&\quad +(\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^a\mathbf{S})_{\nu\mu} + (\mathbf{P}^b\mathbf{S})_{\mu\nu}(\mathbf{P}^b\mathbf{S})_{\nu\mu}]\} \\
&= \sum_{\mu,\nu\in A}[(\mathbf{P}^a - \mathbf{P}^b)\mathbf{S}]_{\mu\nu}[(\mathbf{P}^a - \mathbf{P}^b)\mathbf{S}]_{\nu\mu} = \sum_{\mu,\nu\in A}(\mathbf{P}^s\mathbf{S})_{\mu\nu}(\mathbf{P}^s\mathbf{S})_{\nu\mu}
\end{aligned}
$$

Q.E.D.

2.3. Exchange density and bond order

The expectation value $\varrho(\vec{r}) = \langle\Psi|\sum_{i=1}^{N}\delta(\vec{r}_i - \vec{r})|\Psi\rangle$ of the operator $\hat{\varrho} = \sum_{i=1}^{N}\delta(\vec{r}_i - \vec{r})$ gives the electron density, i.e., the probability density of finding an electron in the vicinity of the point with the radius-vector $\vec{r}$. Similarly to this, we can define the pair density $\varrho_2(\vec{r},\vec{r}\,')$ giving the probability of finding one electron around the point $\vec{r}$ and another electron around the point $\vec{r}\,'$. It is, obviously, given by the expectation value of the operator

$$\hat{\varrho}_2(\vec{r},\vec{r}\,') = \sum_{\substack{i,j\\(i\neq j)}} \delta(\vec{r}_i - \vec{r})\delta(\vec{r}_j - \vec{r}\,') \ . \tag{7.51}$$

One can easily transform this operator as

$$\begin{aligned}\hat{\varrho}_2(\vec{r},\vec{r}\,') &= \sum_{i<j}\delta(\vec{r}_i - \vec{r})\delta(\vec{r}_j - \vec{r}\,') + \sum_{j<i}\delta(\vec{r}_i - \vec{r})\delta(\vec{r}_j - \vec{r}\,') \\ &= \sum_{i<j}\left[\delta(\vec{r}_i - \vec{r})\delta(\vec{r}_j - \vec{r}\,') + \delta(\vec{r}_j - \vec{r})\delta(\vec{r}_i - \vec{r}\,')\right] \ .\end{aligned} \tag{7.52}$$

(We have interchanged the summation indices i and j in the second term.) Now, operator $\hat{\varrho}_2(\vec{r},\vec{r}\,')$ is a symmetric two-electron operator of the type considered in Section 6.3 of Chapter 5 with the definition

$$\hat{g}(i,j) = \delta(\vec{r}_i - \vec{r})\delta(\vec{r}_j - \vec{r}\,') + \delta(\vec{r}_i - \vec{r}\,')\delta(\vec{r}_j - \vec{r}) \ . \tag{7.53}$$

If Ψ is a single-determinant wave function built up of the *orthonormalized* spin orbitals $\psi_i(\vec{r},\sigma) = \varphi_i(\vec{r})\gamma_i(\sigma)$, then the expectation value of $\hat{\varrho}_2(\vec{r},\vec{r}\,')$ can be computed by using (5.116) as

$$\begin{aligned}&\varrho_2(\vec{r},\vec{r}\,') \\ &= \frac{1}{2}\sum_{i,j=1}^{N}\Big[\langle\varphi_i(1)\varphi_j(2)|\delta(\vec{r}_1-\vec{r})\delta(\vec{r}_2-\vec{r}\,')+\delta(\vec{r}_1-\vec{r}\,')\delta(\vec{r}_2-\vec{r})|\varphi_i(1)\varphi_j(2)\rangle \\ &\quad -\langle\varphi_i(1)\varphi_j(2)|\delta(\vec{r}_1-\vec{r})\delta(\vec{r}_2-\vec{r}\,')+\delta(\vec{r}_1-\vec{r}\,')\delta(\vec{r}_2-\vec{r})|\varphi_j(1)\varphi_i(2)\rangle\delta_{\gamma_i\gamma_j}\Big] \\ &= \frac{1}{2}\sum_{i,j=1}^{N}\Big[|\varphi_i(\vec{r})|^2|\varphi_j(\vec{r}\,')|^2 + |\varphi_i(\vec{r}\,')|^2|\varphi_j(\vec{r})|^2 \\ &\qquad - \Big(\varphi_i^*(\vec{r})\varphi_j(\vec{r})\varphi_j^*(\vec{r}\,')\varphi_i(\vec{r}\,') + \varphi_i^*(\vec{r}\,')\varphi_j(\vec{r}\,')\varphi_j^*(\vec{r})\varphi_i(\vec{r})\Big)\,\delta_{\gamma_i\gamma_j}\Big] \ .\end{aligned} \tag{7.54}$$

After interchanging the summation indices i and j in the second and fourth term, one gets

$$\varrho_2(\vec{r},\vec{r}\,') = \sum_{i,j=1}^{N} \left(|\varphi_i(\vec{r})|^2|\varphi_j(\vec{r}\,')|^2 - \varphi_i^*(\vec{r})\varphi_j(\vec{r})\varphi_j^*(\vec{r}\,')\varphi_i(\vec{r}\,')\delta_{\gamma_i\gamma_j}\right) . \quad (7.55)$$

As $\sum_{i=1}^{N}|\varphi_i(\vec{r})|^2 = \varrho(\vec{r})$, one can write

$$\varrho_2(\vec{r},\vec{r}\,') = \varrho(\vec{r})\varrho(\vec{r}\,') - \varrho_2^x(\vec{r},\vec{r}\,') , \quad (7.56)$$

where the first term (it may be called "direct" or "Coulomb type") is the product of the electron densities in points $\vec{r}$ and $\vec{r}\,'$, whereas the second is the two-electron "exchange pair density." The exchange pair density originates from the antisymmetry of the wave function; it contains the compensation for the "self-interaction" of the electrons (*vide infra*) as well.

According to (7.55) we have

$$\varrho_2^x(\vec{r},\vec{r}\,') = \sum_{i,j=1}^{N} \varphi_i^*(\vec{r})\varphi_j(\vec{r})\varphi_j^*(\vec{r}\,')\varphi_i(\vec{r}\,')\delta_{\gamma_i\gamma_j} . \quad (7.57)$$

If Ψ is a determinant built up of n_a orbitals a_i filled with spin α and n_b orbitals b_i filled with spin β, then

$$\varrho_2^x(\vec{r},\vec{r}\,') = \sum_{i,j=1}^{n_a} a_i^*(\vec{r})a_j(\vec{r})a_j^*(\vec{r}\,')a_i(\vec{r}\,') + \sum_{i,j=1}^{n_b} b_i^*(\vec{r})b_j(\vec{r})b_j^*(\vec{r}\,')b_i(\vec{r}\,') . \quad (7.58)$$

Now we integrate $\varrho_2^x(\vec{r},\vec{r}\,')$ over both variables $\vec{r}$ and $\vec{r}\,'$, by taking into account that both sets of orbitals $\{a_i\}$ and $\{b_i\}$ are orthonormalized, and obtain

$$\begin{aligned}\int \varrho_2^x(\vec{r},\vec{r}\,')dvdv' &= \sum_{i,j=1}^{n_a} \langle a_i|a_j\rangle\langle a_j|a_i\rangle + \sum_{i,j=1}^{n_b} \langle b_i|b_j\rangle\langle b_j|b_i\rangle \\ &= \sum_{i,j=1}^{n_a} \delta_{ij} + \sum_{i,j=1}^{n_b} \delta_{ij} = n_a + n_b = N . \end{aligned} \quad (7.59)$$

We substitute the LCAO expansion in (6.115) of the orbitals into (7.59):

$$\begin{aligned}&\sum_{i,j=1}^{n_a} \langle \sum_{\mu=1}^{m} a_\mu^i\chi_\mu | \sum_{\nu=1}^{m} a_\nu^j\chi_\nu\rangle\langle \sum_{\varrho=1}^{m} a_\varrho^j\chi_\varrho | \sum_{\tau=1}^{m} a_\tau^i\chi_\tau\rangle \\ &\quad + \sum_{i,j=1}^{n_b} \langle \sum_{\mu=1}^{m} b_\mu^i\chi_\mu | \sum_{\nu=1}^{m} b_\nu^j\chi_\nu\rangle\langle \sum_{\varrho=1}^{m} b_\varrho^j\chi_\varrho | \sum_{\tau=1}^{m} b_\tau^i\chi_\tau\rangle = N \end{aligned} \quad (7.60)$$

i.e.,

$$\sum_{i,j=1}^{n_a} \sum_{\mu,\nu,\varrho,\tau=1}^{m} a_\tau^i a_\mu^{i*} S_{\mu\nu} a_\nu^j a_\varrho^{j*} S_{\varrho\tau} + \sum_{i,j=1}^{n_b} \sum_{\mu,\nu,\varrho,\tau=1}^{m} b_\tau^i b_\mu^{i*} S_{\mu\nu} b_\nu^j b_\varrho^{j*} S_{\varrho\tau} = N \ . \tag{7.61}$$

By using the definitions in (6.116) this can be rewritten as

$$\sum_{\nu,\tau=1}^{m} [(\mathbf{P}^a\mathbf{S})_{\tau\nu}(\mathbf{P}^a\mathbf{S})_{\nu\tau} + (\mathbf{P}^b\mathbf{S})_{\tau\nu}(\mathbf{P}^b\mathbf{S})_{\nu\tau}] = N \ . \tag{7.62}$$

We can group the terms on the left-hand side according to the atoms on which the basis orbitals are centered. We get

$$\sum_{A,B} \sum_{\nu\in A} \sum_{\tau\in B} [(\mathbf{P}^a\mathbf{S})_{\tau\nu}(\mathbf{P}^a\mathbf{S})_{\nu\tau} + (\mathbf{P}^b\mathbf{S})_{\tau\nu}(\mathbf{P}^b\mathbf{S})_{\nu\tau}] = N \ . \tag{7.63}$$

Comparing this expression with the definition in (7.36), we see that the bond order B_{AB} between atoms A and B is the diatomic contribution to the integral of the exchange pair density $\varrho^x(\vec{r},\vec{r}\,')$. (The factor 2 present in the definition in (7.36) is also recovered, because the sum for A and B in expansion (7.63) runs over all the atoms independently.)

One may observe that equality (7.62) can also be obtained as a consequence of the idempotency of matrices $\mathbf{P}^a\mathbf{S}$ and $\mathbf{P}^b\mathbf{S}$ and of using the equalities (Appendix VII) $\mathrm{Tr}(\mathbf{P}^a\mathbf{S}) = n_a$ and $\mathrm{Tr}(\mathbf{P}^b\mathbf{S}) = n_b$. Accordingly, by using these idempotency relations one can define also higher order bond indices, considering for instance the three-center contributions present in the quantity $\mathrm{Tr}(\mathbf{P}^a\mathbf{S}\mathbf{P}^a\mathbf{S}\mathbf{P}^a\mathbf{S} + \mathbf{P}^b\mathbf{S}\mathbf{P}^b\mathbf{S}\mathbf{P}^b\mathbf{S}) = N$, and so on.

We have to add a few more remarks concerning the close connection between exchange and chemical bonding. One may observe that the electron–electron repulsion energy $\langle\Psi|\sum_{i<j}\frac{1}{r_{ij}}|\Psi\rangle$ can be expressed by the pair density $\varrho_2(\vec{r},\vec{r}\,')$ as

$$E_{el-el} = \langle\Psi|\sum_{i<j}\frac{1}{r_{ij}}|\Psi\rangle = \iint \frac{\varrho_2(\vec{r},\vec{r}\,')}{\vec{r}-\vec{r}\,'}\, dv dv' \ . \tag{7.64}$$

This follows from the possibility of writing $\frac{1}{r_{ij}}$ in the form of the integral

$$\frac{1}{r_{ij}} = \frac{1}{|\vec{r}_i - \vec{r}_j|} = \frac{1}{2}\iint \frac{\delta(\vec{r}_i-\vec{r})\delta(\vec{r}_j-\vec{r}\,') + \delta(\vec{r}_i-\vec{r}\,')\delta(\vec{r}_j-\vec{r})}{|\vec{r}-\vec{r}\,'|}\, dv dv' \ . \tag{7.65}$$

Then, when computing $\langle\Psi|\sum_{i<j}\frac{1}{r_{ij}}|\Psi\rangle$ one may change the order of integrations and complete all of them except those over $\vec{r}$ and $\vec{r}\,'$. For a single-determinant

wave function, (7.64) can further be written by using (7.56) as

$$E_{el-el} = \iint \frac{\varrho(\vec{r})\varrho(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\, dvdv' - \iint \frac{\varrho_2^x(\vec{r},\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\, dvdv' \,. \tag{7.66}$$

As $\varrho(\vec{r})$ is a density obtained by summing for all the electrons, the first term on the right-hand side contains the electrostatic interaction of each electron with each electron—including the interaction of a given electron in point $\vec{r}$ with its own charge density in point $\vec{r}\,'$. This is the case because we have included the self-repulsion terms $\hat{J}_i - \hat{K}_i$ in the Fock operator and the Hartree–Fock energy. The necessary compensation of this self-interaction is contained in the second term on the right-hand side (the exchange energy).

Obviously, a similar representation of E_{el-el} is also possible for correlated wave functions, but in that case the exchange part $\varrho_2^x(\vec{r},\vec{r}\,')$ of the pair density should be replaced by the "exchange-correlation" one, which may be (formally) defined as the difference $\varrho_2^{xc}(\vec{r},\vec{r}\,') = \varrho_2(\vec{r},\vec{r}\,') - \varrho(\vec{r})\varrho(\vec{r}\,')$. In practice it can be computed by using the actual correlated wave function.

In the single-determinant case, by substituting (7.55) into (7.66) one immediately obtains the two-electron part of the Hartree–Fock energy expression (6.44). Its exchange component is

$$\begin{aligned} E^x &= -\iint \frac{\varrho_2^x(\vec{r},\vec{r}\,')}{\vec{r}-\vec{r}\,'}\, dvdv' = -\sum_{i,j=1}^{N} [\varphi_i\varphi_j|\varphi_j\varphi_i]\delta_{\gamma_i\gamma_j} \\ &= -\sum_{i,j=1}^{n_a} [a_ia_j|a_ja_i] - \sum_{i,j=1}^{n_b} [b_ib_j|b_jb_i] \,. \end{aligned} \tag{7.67}$$

This equality shows a close connection between the exchange pair density $\varrho_2^x(\vec{r},\vec{r}\,')$ and the exchange part E^x of the total Hartree–Fock energy. It is to be pointed out that the exchange energy is negative. In fact, each integral $[\varphi_i\varphi_j|\varphi_j\varphi_i]$ is positive—at least if real orbitals are used—as it represents the electrostatic self-repulsion of the respective density $\varphi_i(\vec{r})\varphi_j(\vec{r})$. (This is in accord with the fact that exchange energy contains the correction for the self-interaction.)

As we have seen previously, the integral $\iint \varrho_2^x(\vec{r},\vec{r}\,')dvdv'$ of the exchange pair density can be decomposed into one- and two-center components if atom-centered one-electron basis functions are used. This cannot be performed for the exchange energy (7.67) because the expansion of the two-electron integrals over the MOs will contain, in general, three- and four-center integrals over the basis orbitals, as well.

Such a decomposition is, however, possible—and rather instructive—in the simplest case of the CNDO-type model theories using the "zero differential

overlap" approximation:

$$[\chi_\mu\chi_\nu|\chi_\varrho\chi_\tau] = \gamma_{AB}\delta_{\mu\varrho}\delta_{\nu\tau} \qquad \mu \in A;\ \nu \in B\ . \tag{7.68}$$

Here A and B are the atoms on which the basis orbitals χ_μ and χ_ν, respectively, are centered. Using the approximation in (7.68), one obtains by substituting the LCAO expansion (6.115) of the orbitals a_i and b_i into the expression (7.67) and using the definitions (6.116) of matrices $\mathbf{P}^a, \mathbf{P}^b$:

$$\begin{aligned} E^x(CNDO) &= -\sum_{i,j=1}^{n_a}\sum_{\mu,\nu,\varrho,\tau=1}^{m} a_\mu^{i*}a_\nu^{j*}a_\varrho^j a_\tau^i[\chi_\mu\chi_\nu|\chi_\varrho\chi_\tau] \\ &\quad -\sum_{i,j=1}^{n_b}\sum_{\mu,\nu,\varrho,\tau=1}^{m} b_\mu^{i*}\,b_\nu^{j*}b_\varrho^j\,b_\tau^i\,[\chi_\mu\chi_\nu|\chi_\varrho\chi_\tau] \qquad (7.69) \\ &= -\sum_{A,B}\sum_{\mu\in A}\sum_{\nu\in B}\sum_{\varrho,\tau=1}^{m}(P_{\tau\mu}^a P_{\varrho\nu}^a + P_{\tau\mu}^b P_{\varrho\nu}^b)\gamma_{AB}\delta_{\mu\varrho}\delta_{\nu\tau} \\ &= -\sum_{A,B}\sum_{\mu\in A}\sum_{\nu\in B}(|P_{\mu\nu}^a|^2 + |P_{\mu\nu}^b|^2)\gamma_{AB}\ . \end{aligned}$$

Comparing this expression with (7.34) and taking into account that the sum for A and B runs over all the atoms independently, we arrive to the conclusion that the diatomic contribution to the exchange energy is proportional to the respective Wiberg index:

$$E_{AB}^x = -\frac{1}{2}\gamma_{AB}W_{AB}\ . \tag{7.70}$$

The author introduced the *ab initio* bond order index in (7.36) by considering an approximate *ab initio* generalization of this important relationship.

2.4. Bond orders in three-center bonds

Most chemical bonds can be considered as being formed by a pair of electrons occupying a two-center localized molecular orbital. (For multiply bonded atoms a few such bonding electron pairs should be assumed.) Conjugation and aromaticity can also be discussed on this basis, assuming that the wave function is a linear combination of several components, describing "resonance" between different "Kekulé-structures," each of which is characterized by two-electron two-center bonds. There are, however, systems in which one has to postulate the existence of *three-center bonds*. The probably most known example is the diborane molecule B_2H_6 in which the two boron atoms are bridged by two symmetrically positioned hydrogen atoms. The electronic structure of this molecule is usually described in terms of two three-center two-electron bonds

each of which is built up in first approximation of appropriate hybrid atomic orbitals of the two boron atom and the $1s$ orbital of the "bridging" hydrogen. This means that the hydrogen is bonded by partial bonds to both borons symmetrically. (There is a conceptual difference with the usual hydrogen bonds in which hydrogen has only a secondary bonding with one of the partners.) The significant distance between the boron atoms seemingly would indicate that there should be no significant interaction between them. This is, however, not the case: as we shall see, the formation of three-center two-electron bonds leads to the appearance of a partial bond order between the boron atoms as well. The attractive exchange interaction corresponding to these partial bond orders can be of significant importance as far as the energetics (stability) of the diborane molecule (or similar systems) is concerned.

Let us consider the simple model in which two electrons occupy a molecular orbital formed as a linear combination of three normalized basis orbitals χ_1 to χ_3. χ_1 and χ_3 are the boron orbitals directed to the hydrogenic orbital χ_2. We assume that the overlap of the boron and hydrogenic orbitals $\langle\chi_1|\chi_2\rangle = \langle\chi_2|\chi_3\rangle = S$ whereas the overlap of the two boron orbitals is negligible. The overlap matrix $\mathbf{S}$ is, accordingly,

$$\mathbf{S} = \begin{pmatrix} 1 & S & 0 \\ S & 1 & S \\ 0 & S & 1 \end{pmatrix} . \tag{7.71}$$

We construct the simplest symmetric unpolarized three-center orbital as

$$\psi = \frac{1}{\sqrt{2(1+\sqrt{2}S)}} \left[\chi_2 + \frac{1}{\sqrt{2}}(\chi_1 + \chi_3)\right] , \tag{7.72}$$

i.e., the vector of the LCAO coefficients is

$$\mathbf{c} = \frac{1}{\sqrt{2(1+\sqrt{2}S)}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 1 \\ \frac{1}{\sqrt{2}} \end{pmatrix} . \tag{7.73}$$

The trivial evaluation of matrices $\mathbf{D} = 2\mathbf{c}\mathbf{c}^\dagger$ and $\mathbf{DS}$ gives

$$\mathbf{D} = \frac{1}{2(1+\sqrt{2}S)} \begin{pmatrix} 1 & \sqrt{2} & 1 \\ \sqrt{2} & 2 & \sqrt{2} \\ 1 & \sqrt{2} & 1 \end{pmatrix} ; \tag{7.74}$$

$$\mathbf{DS} = \frac{1}{2} \begin{pmatrix} 1 & \sqrt{2} & 1 \\ \sqrt{2} & 2 & \sqrt{2} \\ 1 & \sqrt{2} & 1 \end{pmatrix} , \tag{7.75}$$

respectively. This means that one obtains bond orders $B_{12} = B_{23} = \frac{1}{2}$; $B_{13} = \frac{1}{4}$. (When using the definition in (7.35) one has to take into account that each center bears a single basis orbital in our simplified model.)

These results mean that in the diborane molecule there should be bond orders of about $\frac{1}{2}$ between the bridging hydrogens and each boron atom, as well as between the two boron atoms: because the diborane molecule contains two independent two-electron three-center bonds, the boron–boron bond order should be doubled as compared with the value $\frac{1}{4}$ obtained previously. These values are in very good agreement with those obtained in the actual *ab initio* calculations.

Similar derivations can be performed for the case of three-center four-electron bonds, in which a nonbonding orbital is also doubly occupied. A typical example is the axial bonding system of a hypervalent sulphur atom. In that case one has to assume that the central orbital χ_2 located on the sulphur is of p-type, thus $\langle\chi_1|\chi_2\rangle = S$ and $\langle\chi_2|\chi_3\rangle = -S$. The bonding and nonbonding orbitals are in this case

$$\psi_b = \frac{1}{\sqrt{2(1+\sqrt{2}S)}}\left[\chi_2 + \frac{1}{\sqrt{2}}(\chi_1 - \chi_3)\right] \tag{7.76}$$

and

$$\psi_n = \frac{1}{\sqrt{2}}(\chi_1 + \chi_3)\ , \tag{7.77}$$

respectively. As it is trivial to check, this difference, however, does not influence the results obtained for the bond orders: $B_{12} = B_{23} = \frac{1}{2}$ and $B_{13} = \frac{1}{4}$. (It is not surprising that the presence of the electrons on the nonbonding orbital ψ_n does not influence the bond order values.) These relatively small values of the bond orders can be used to explain why the axial bonds of the hypervalent sulphur compounds like SF_4 are longer than the conventional two-electron two-center equatorial bonds.

Notes

The importance of distinguishing between population analyses performed with respect to the three-dimensional physical space and with respect to the Hilbert space of the basis orbitals was stressed by Hall [1].

Section 1.

Mulliken's population analysis [2] is discussed in most textbooks of quantum chemistry (e.g., [3]). The requirement of the rotational-hybridizational invariance was first studied systematically in connection with the parametrization of the semiempirical CNDO method [4]. Despite its fundamental importance, the rotational-hybridizational importance of Mulliken's net and overlap population is not discussed explicitly in any sources known to the author.

A disadvantage of Mulliken's populations is that the gross orbital population q_μ of orbital χ_μ is not strictly limited by the numbers 0 and 2, as would be the case for an orthonormalized basis set. This may cause problems if large basis sets are used, containing several "diffuse" functions which decay slowly and hence are not really of atomic character [5]. (They are significantly populated for negative ions.) In such a case one has to turn to the Löwdin-orthogonalized counterpart (Section 2.3 of Chapter 3) of the basis applied, and perform the population (and bond order) analysis in terms of this auxiliary orthonormalized basis set.

Section 2.1.

Wiberg introduced his index in [6] for closed-shell systems. The proper generalization for the open-shell systems was given by Borisova and Semenov [7]. They also proved [8] that the Wiberg index is equal to $\frac{1}{2}(N_{\text{bonding}} - N_{\text{antibonding}})$ for first row homonuclear diatomics—except C_2 which has a very peculiar electronic structure [9].

The general bond order index Eq. (7.35) was proposed on intuitive grounds by Giambiagi *et al.* [10] for the semiempirical extended Hückel method. Being unaware of that, the present author introduced it for the *ab initio* case [11] on the basis of analyzing the exchange energy contribution obtained in an energy decomposition scheme [12]. The correct form Eq. (7.36) for the open-shell case was given in [13] as an *ab initio* generalization of the results of Borisova and Semenov [7]. (Also see [14,15].)

The invariance of bond order index with respect to rotational-hybridizational transformations has been established in [16].

Section 2.2.

The concept of the valence index was introduced in the CNDO framework independently by Borisova and Semenov [7] and Armstrong *al.* [17]; it was generalized for the *ab initio* case by the present author [11], who also introduced the free valence index. (Also see [13, 14, 15].)

Section 2.3.

The present derivation has a close relationship to the analysis of the connections between the bond order index and the normalization of the exchange part of the "second order density matrix $\rho_2(x_1, x_2; x_1', x_2')$ [13, 14, 15].

Section 2.3.

The material is based on [18].

Suggested readings

The survey papers [19, 20] are recommended.

1 G.G. Hall *Chairman's remarks,* 5th International Congress on Quantum Chemistry, Montreal, 1985.

2 R.S. Mulliken, J. Chem. Phys. **23**, 1833, 1841, 2338, 2343 (1955)

3 T. Veszprémi and M. Fehér, *Quantum Chemistry: Fundamentals to Applications*, Kluwer Academic/Plenum New York, 1999.

4 J. A. Pople and D. L. Beveridge *Approximate Molecular Orbital Theory,* McGraw-Hill, New York 1970.

5 J. Baker: Theor. Chim. Acta **68**, 221 (1985)

6 K.A. Wiberg, Tetrahedron **24**, 1083 (1968)

7 N.P. Borisova and S.G. Semenov, Vestn. Leningrad. Univ. **1976**, No. 16, 98 (1976)

8 N.P. Borisova and S.G. Semenov, Vestn. Leningrad. Univ. **1973**, No. 16, 119 (1973)

9 K. Jug, private communication to the author, 1984.

10 M. Giambiagi, M.D. de Giambiagi, D.R. Grempel and C.D. Heynemann: J. Chim. Phys. **72**, 15 (1975)

11 I. Mayer, Chem. Phys. Letters, **97**, 270 (1983); Addendum: **117**, 396 (1985)

12 I. Mayer, Int. J. Quantum Chem. **23**, 341 (1983)

13 I. Mayer, Int. J. Quantum Chem. **29**, 73 (1986)

14 I. Mayer, Int. J. Quantum Chem. **29**, 477 (1986)

15 I. Mayer, Theor. Chim. Acta **67**, 315 (1985)

16 I. Mayer, DSc. Thesis, Hungarian Academy of Sciences, Budapest 1986 (unpublished).

17 D.R. Armstrong, P.G. Perkins and J.J.P. Stewart, J. Chem. Soc. Dalton Trans. **1973**, 838, 2273 (1973)

18 I. Mayer, J. Mol. Struct. (Theochem)**186**, 43 (1989)

19 A.B. Sannigrahi, Adv. Quantum Chem. **23**, 301 (1992)

20 A.J. Bridgeman, G. Cavigliasso, L.R. Ireland and J. Rothery, J. Chem. Soc Dalton Trans. **2001** 2095 (2001)

Chapter 8

THE ELECTRON CORRELATION

1. The CI expansion

Let us consider an arbitrary normalizable N-electron wave function $\Psi = \Psi(1, 2, ..., N)$ at the fixed values of the space and spin coordinates of the first $N-1$ electrons: $x_k \equiv \{\vec{r}_k, \sigma_k\}$=const; $k = 1, 2, ..., N-1$. Under these conditions Ψ is a normalizable function of the space and spin coordinates of the Nth electron, and as such, it can be expanded in a series in terms of some complete set $\psi_i = \psi_i(x)$, $i = 1, 2, ..., \infty$ of one-electron functions

$$\Psi(1, 2, ..., N) = \sum_{i_N=1}^{\infty} c_{i_N}(x_1, x_2, ..., x_{N-1})\psi_{i_N}(x_N) \ . \tag{8.1}$$

The coefficients of this expansion are functions of the coordinates of the other electrons $i = 1, 2, ..., N-1$. Obviously, they should also be normalizable functions of their arguments. Therefore, for fixed values of x_i, $i = 1, 2, ..., N-2$, we may again expand $c_{i_N} = c_{i_N}(x_1, x_2, ..., x_{N-1})$ into series as a function of x_{N-1}; thus we get

$$\Psi = \sum_{i_{N-1}, i_N=1}^{\infty} c_{i_{N-1} i_N}(x_1, x_2, ..., x_{N-2})\psi_{i_{N-1}}(x_{N-1})\psi_{i_N}(x_N) \tag{8.2}$$

and so on. Finally we get

$$\Psi = \sum_{i_1, i_2, ..., i_N=1}^{\infty} c_{i_1 i_2 ... i_N}\psi_{i_1}(x_1)\psi_{i_2}(x_2)...\psi_{i_N}(x_N) \ , \tag{8.3}$$

where the coefficients $c_{i_1 i_2 ... i_N}$ do not depend any more on the coordinates of either of the electrons.

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

The function Ψ must be antisymmetric with respect to the interchange of electrons; this leads to some restrictions on the values of the coefficients $c_{i_1 i_2 ... i_N}$. For instance, it follows from the antisymmetry of Ψ with respect to the interchange of electrons "1" and "2" that

$$c_{i_1 i_2 i_3 ... i_N} = -c_{i_2 i_1 i_3 ... i_N} \tag{8.4}$$

and so on. One can collect those terms that contain the same spin orbitals and differ only in the order of the electron coordinates and in the order of the subscripts $i_1, i_2, ..., i_N$ of the coefficients $c_{i_1 i_2 ... i_N}$. Then one obtains the Slater determinant built up of the spin orbitals involved, multiplied with a common coefficient. Requesting the subscripts i_j to form an increasing sequence $(i_1 < i_2 < ... < i_N)$, the set of N indices $\{i_j\}$ is called the Ith "ordered configuration." Then the wave function Ψ can be presented as an expansion

$$\Psi = \sum_I c_I \Psi_I \tag{8.5}$$

over the Slater determinants corresponding to the different ordered configurations

$$\Psi_I = \hat{\mathcal{A}}[\psi_{i_1}(1)\psi_{i_2}(2)...\psi_{i_N}(N)] \ . \tag{8.6}$$

If the one-electron functions ψ_i form a complete set in the one-electron Hilbert space, then the determinants corresponding to all the ordered configurations built up by their use form a complete set in the linear space of the antisymmetric N-electron functions. If the one-electron functions are orthonormalized, then the determinants built up of them are orthonormalized, as well.

Similarly, if we start from a finite number of spin orbitals ψ_i, then the determinants corresponding to the ordered configurations of these orbitals form a complete basis in the subspace of antisymmetric wave functions that can be built up by using the given set of one-electron functions.

The expansion of wave functions in term of determinants described previously is usually called a "configuration interaction" (CI) expansion. This is a traditional nomenclature; it would be more fortunate to use a term like "superposition (or linear combination) of configurations." The CI method used in practice consists of variational determination of the coefficients c_I—which requires the solution of a secular equation corresponding to the linear variational problem (c.f. Section 1 of Chapter 3).

A characteristic feature of the Born–Oppenheimer Hamiltonian used in quantum chemistry is that it contains the interelectronic distances in a singular manner $\left(\frac{1}{r_{ij}} \to \infty \text{ if } r_{ij} \to 0\right)$. The CI expansion is written out in terms of one-electron spin orbitals, so it does not contain explicitly the interelectronic distances. Consequently, the behavior of the wave function in the cases when

two electrons are close to each other is described only as a combined effect of many different terms in the CI expansion—one may say, as their interference. As a consequence, the CI expansion converges rather slowly, especially as far as the local properties (c.f. Section 2.2 of Chapter 9) of the wave function are concerned.

2. The correlation energy and Nesbet's theorem

The Hartree–Fock wave functions do not represent exact solutions of the Schrödinger equation because they reflect the interelectronic interactions in an averaged manner only. This means that the Hartree–Fock wave function does not account for the fact that the different electrons do not move independently of each other but instead there is a *correlation* between their motion. (This correlation leads, among other things, to the dependence of the wave function on the interelectronic distance, as mentioned previously.)

One can characterize the electron correlation as follows. The probability of finding one of the electrons in a given point depends on the position of the other electron(s), too. Let $p_1(\vec{r}_1)$ be the probability density of finding the first electron in the point $\vec{r}_1$, independently of the positions of the other electrons (i.e., admitting arbitrary positions of the other electrons). Similarly we may define the probability density $p_2(\vec{r}_2)$ for the second electron. Then the correlation between the motion of the two electrons means that one has for the probability density $p_{12}(\vec{r}_1, \vec{r}_2)$ of finding the first electron in point $\vec{r}_1$ and at the same time the second electron in the point $\vec{r}_2$ the inequality

$$p_{12}(\vec{r}_1, \vec{r}_2) \neq p_1(\vec{r}_1)p_2(\vec{r}_2) \tag{8.7}$$

as a conditional probability of events that are not statistically independent.

Strictly speaking, the motion of all electrons was treated as being fully independent of each other only in the framework of the (completely obsolete) Hartree method that neglected the antisymmetry requirement and used a single Hartree-product wave function. By introducing the antisymmetrization one excludes the possibility of finding two electrons with the same spin in the same point of space. At the level of the Slater determinants this manifests in the following manner: if two electrons—say the ith and jth—have the same coordinates and the same spin, then $\vec{r}_i = \vec{r}_j$, $\sigma_i = \sigma_j$, and the ith and jth rows of the determinant wave function (5.15) become equal and thus the determinant vanishes. Owing to the continuity of the wave function, the probability of finding two electrons of the same spin close to each other is also small. This phenomenon is called a "Fermi hole"; it is exclusively a consequence of the antisymmetry of the wave function and it would appear even if the electrons were non-interacting. Thus a Fermi hole may be considered a kinematical effect.

Due to the Fermi hole, only the motion of the electrons of opposite spin is really uncorrelated if a Hartree–Fock wave function is used. Nevertheless, owing to the particular importance of the conventional restricted Hartree–Fock method, one defines correlation energy, according to the commonly accepted nomenclature introduced by Löwdin, as the error of the energy given by the RHF method, i.e., as the deviation of the RHF energy from the exact nonrelativistic energy value. Accordingly, the notion of electron correlation includes only effects not following from the antisymmetry and, therefore, not accounted for in the RHF method. Thus one calls electron correlation only those effects that represent consequences of the repulsion between electrons, which reduces the probability of finding them close to each other ("Coulomb hole"). Accordingly, electron correlation is a dynamical effect, which is essentially independent of the spin, as opposed to the Fermi hole which is a spin-dependent kinematical effect.

There is, however, another nomenclature in quite general use, too, in which one distinguishes between "static" and "dynamical" correlation. Static correlation refers to the cases in which the description of the system by using a single determinant RHF wave function built up of doubly filled orbitals does not reflect well the physical situation. (A typical example is the homolytical dissociation discussed in Section 4.3 of Chapter 6, but there can be more complex cases, as well.) Dynamical correlation is that observed near the equilibrium configurations of common molecules as the direct consequence of the interelectron repulsion. Although this distinction is often useful and can be even transparent in some cases, it can hardly be made very exact.

In many problems of chemical interest one may assume with some degree of confidence that the correlation energy is nearly constant. (This is, however, not the case when the number and/or character of the chemical bonds present in the system is changing.) Therefore, the results obtained by performing Hartree–Fock–Roothaan calculations with a good basis set are often sufficient for semiquantitative results. However, if one is interested in accurate numbers—and especially in the cases in which small energy differences are considered—then one cannot avoid explicit accounting for electron correlation. There are problems—e.g., the studies of reaction paths—in which one cannot expect even qualitatively correct results without considering electron correlation.

The most direct description of electron correlation can be performed by introducing explicitly the interelectronic distances r_{ij} into the wave functions. Traditionally that was possible only for the smallest systems like the He atom or the H_2 molecule, and it permitted one to obtain practically exact solutions of the Schrödinger equation for these two-electron systems. Nowadays there are (approximate) methods permitting one to include some r_{ij} dependence into refined wave functions of larger systems, as well. However, the absolute ma-

jority of electron correlation calculations performed in practice is based on the CI expansion of the wave function, discussed in the previous section.

Almost all methods start from the (restricted or unrestricted) Hartree–Fock wave function as the "reference state." Then the ordered configurations I correspond to the determinants built up of the occupied and virtual HF orbitals. (These usually—but not necessarily—are the canonical HF orbitals.) Denoting the HF wave function as Ψ_0 and using the intermediate (or "correlational") normalization also applied in Chapter 4, we can write

$$\Psi = \Psi_0 + \varphi = \Psi_0 + \sum_{I=1} c_I \Psi_I \tag{8.8}$$

with

$$\langle \Psi_0 | \Psi_0 \rangle = 1; \quad \langle \Psi_0 | \varphi \rangle = 0 \,. \tag{8.9}$$

(The last equalities follow from the orthonormalized character of the occupied and virtual HF orbitals, leading to $\langle \Psi_0 | \Psi_I \rangle = \delta_{0I}$.)

Substituting the expansion in (8.8) into the Schrödinger equation

$$\hat{H}\Psi = E\Psi \tag{8.10}$$

one has

$$\hat{H}\left(\Psi_0 + \sum_{I=1} c_I \Psi_I\right) = E\left(\Psi_0 + \sum_{I=1} c_I \Psi_I\right) \,. \tag{8.11}$$

By multiplying (8.11) with Ψ_0^* and integrating, we get, by taking into account that $\langle \Psi_0 | \Psi_I \rangle = 0$ if $I \neq 0$,

$$E = \underbrace{\langle \Psi_0 | \hat{H} | \Psi_0 \rangle}_{E_{HF}} + \sum_{I=1} c_I \underbrace{\langle \Psi_0 | \hat{H} | \Psi_I \rangle}_{H_{0I}} \tag{8.12}$$

i.e.,

$$\Delta E = E - E_{HF} = \sum_I c_I H_{0I} \,. \tag{8.13}$$

As the determinants Ψ_I are built up of the occupied and virtual HF orbitals, they can be classified according to the number of spin orbitals that are replaced by virtual ones, as compared with the HF wave function Ψ_0. Thus one has singly, doubly, triply substituted (or "excited") determinants, and so on—up to N-times substituted ones, where N is the number of electrons. If Ψ_I is a singly excited determinant, then the matrix element $\langle \Psi_0 | \hat{H} | \Psi_I \rangle = 0$ owing to the Brillouin theorem (Section 1 of Chapter 6). For determinants with three or more substitutions with respect to Ψ_0, the matrix element vanishes because of

the Slater rules (Section 6 of Chapter 5). Therefore, the only terms contributing to the right-hand side of equality (8.13) are the doubly excited determinants. Introducing for the latter the notation $\Psi_{ij}^{uv} = \Psi_2(\psi_i \to \psi_u; \psi_j \to \psi_v)$ and denoting by c_{ij}^{uv} the respective coefficients in the CI expansion in (8.8) of the exact wave function, we get the expression of the correlation energy as

$$\begin{aligned} \Delta E &= \sum_{i<j}^{(\text{occ.})} \sum_{u<v}^{(\text{virt.})} c_{ij}^{uv} \langle \Psi_0 | \hat{H} | \Psi_{ij}^{uv} \rangle \\ &= \sum_{i<j}^{(\text{occ.})} \sum_{u<v}^{(\text{virt.})} c_{ij}^{uv} \left(\langle \psi_i(1)\psi_j(2) | \frac{1}{r_{12}} | \psi_u(1)\psi_v(2) \rangle \right. \\ &\quad \left. - \langle \psi_i(1)\psi_j(2) | \frac{1}{r_{12}} | \psi_v(1)\psi_u(2) \rangle \right) , \end{aligned} \tag{8.14}$$

where the integrals refer to spin orbitals, i.e., the integrations include summation over the spins.

The results in (8.14) indicates that the calculation of the exact correlation energy requires the knowledge of the exact CI expansion coefficients of the doubly excited configurations only. This result is usually called Nesbet's theorem.

An important consequence of (8.14) is that by introducing some obvious notations, the correlation energy can be presented as a sum of the form

$$\Delta E = \sum_{i<j}^{(\text{occ.})} \varepsilon_{ij} , \tag{8.15}$$

where the summation is over the occupied HF *spin orbitals*. This result indicates that the exact correlation energy can always be presented as the sum of contributions

$$\begin{aligned} \varepsilon_{ij} = \sum_{u<v}^{(\text{virt.})} c_{ij}^{uv} \Big(&\langle \psi_i(1)\psi_j(2) | \frac{1}{r_{12}} | \psi_u(1)\psi_v(2) \rangle \\ &- \langle \psi_i(1)\psi_j(2) | \frac{1}{r_{12}} | \psi_v(1)\psi_u(2) \rangle \Big) \end{aligned} \tag{8.16}$$

usually called "pair correlation energies," corresponding to the individual pairs of occupied spin orbitals. Each such pair-correlation-energy contribution depends, strictly speaking, on all the other orbitals, too, because the values of the coefficients c_{ij}^{uv} are influenced—directly or indirectly—by all the configurations of the CI expansion.

If the spatial part of the spin orbitals is the same—i.e., they correspond to a doubly filled spatial orbital filled with spins α and β—then one speaks of "intrapair correlation," otherwise one speaks of "interpair correlation." Usually

the intrapair correlation contributions are the largest ones, as they correspond to a pair of electrons that are forced to occupy the same spatial orbital at the HF level, which involves a relatively large probability that the two electrons to be close to each other. Analogously to the distinction between intrapair and interpair correlations, one can also define "intrashell" and "intershell" correlations for atoms, and so on.

The considerations leading to equalities (8.14) and (8.15) are equally valid independently of whether or not the one-electron HF spin orbitals are the canonical ones. Whereas the resulting correlation energy ΔE is invariant of unitary transformations of occupied orbitals, the individual pair correlation energies are, of course, not. Both ΔE and every ε_{ij} are invariant with respect to unitary transformations of the virtual orbitals. (The invariance of ΔE follows from the fact that the solution of the Schrödinger equation cannot depend on the selection of the basis in the linear space of the normalizable antisymmetric N-electron wave functions. Also, it follows from the properties of the determinants that the *subspace* of determinants that can be obtained by double excitations from a *given pair* of occupied spin orbitals remains invariant under the unitary transformations of the virtual orbitals, leading to the invariance of the pair correlation energies.)

One expects that the dominance of the intrapair correlation will be even more pronounced if localized orbitals are used and that only those pair correlation energies will be of considerable value that correspond to orbitals located not very far from each other in physical space. (Both the integrals and the coefficients in (8.16) decay quickly with increases in the distance.) Also, for similar orbitals (e.g., atomic $1s$ orbitals) one gets quite similar values in different molecules. This often makes it possible to exclude from consideration the correlation of the core orbitals without changing significantly the energy differences in which we are interested.

From the point of view of the intrapair correlation corresponding to a given doubly occupied orbital, the virtual orbitals having the greatest importance are the ones that are mainly located in the same part of space as the occupied orbital considered but have more nodal surfaces than the latter. For instance, in the case of a chemical bond that is described at the SCF level by a localized bonding orbital, a significant part of the intrapair correlation will be accounted for by considering the excitations to the corresponding antibonding orbital. In this manner one describes the so-called "left–right" correlation, which means that if one electron of the pair is near to the one nucleus then another electron will be found with a greater probability near the another one. (In that way, accounting for electron correlation reduces the relative weight of the ionic terms that is exaggerated at the SCF level—c.f., Section 4.3 of Chapter 6.) Similarly, one speaks—especially in the case of free atoms—of the radial (or "in–out") correlation; for the description of which one needs radial functions with different

orbital exponents and different numbers of nodal surfaces. (For instance, in order to describe the radial correlation of a $1s$ shell, one needs orbitals of type $2s$, $3s$, etc.) Another important class of correlation is the "angular correlation," which means that the probability of finding two electrons in different directions with respect to the nucleus is larger than finding them in the same direction. This effect can be described by using orbitals with higher angular quantum numbers. (One needs, for instance, the use of p, d, f, etc., orbitals to describe the angular correlation of a pair of $1s$ electrons.) In general, one has to use in the CI expansion such orbitals that are strongly oscillating and have a number of separated domains in which the occupied orbitals significantly differ from zero. One has to keep in mind that a very large number of different one-electron functions is needed to describe well the Coulomb hole: as it has already been noted, the r_{ij} dependence of the wave function appears only as an interference of many terms of the CI expansion.

We should mention that there is an interesting approach of trying to take a shortcut to inclusion of electron correlation by considering every quantity of interest as a functional of a single variable, the electron density. It is called "Density functional theory" (DFT). It is *formally* proved, indeed, that there exists a "universal functional" of the electron density giving the sum of the exchange and correlational energies. The actual form of this exchange–correlational functional is, however, unknown, and there exists no methods that would permit one to start from an approximate form of the functional and improve it in any systematic fashion. For that reason it is the present author's opinion that DFT should be considered as a useful semiempirical scheme of including approximately some electron correlation in the calculations. We are not going to consider DFT in any detail here.

Example

Let us consider the simplest example of the H_2 molecule treated by using the minimal basis set consisting of two atomic orbitals χ_A and χ_B. The spatial symmetry of the system determines the MOs uniquely. There is a symmetric bonding orbital

$$\varphi = \mathcal{N}(\chi_A + \chi_B) \tag{8.17}$$

and an antisymmetric antibonding one:

$$\psi = \mathcal{N}'(\chi_A - \chi_B) \ . \tag{8.18}$$

The values of the normalization coefficients $\mathcal{N}$ and $\mathcal{N}'$ are, as is easy to check:

$$\mathcal{N} = \frac{1}{\sqrt{2(1+S)}} \ ; \quad \mathcal{N}' = \frac{1}{\sqrt{2(1-S)}} \ , \tag{8.19}$$

where $S = \langle \chi_1 | \chi_2 \rangle$. Owing to their different symmetry, the orbitals φ and ψ are orthogonal.

In this simple model problem there are only two determinant wave functions $\Psi_0 = \hat{\mathcal{A}}[\varphi(1)\alpha(1)\varphi(2)\beta(2)]$ and $\Psi_d = \hat{\mathcal{A}}[\psi(1)\alpha(1)\psi(2)\beta(2)]$, which are symmetric with respect to the inversion in the center of the molecule; both are built up of doubly filled orbitals. The ground-state wave function using the correlated normalization is

$$\Psi = \Psi_0 + c\Psi_d \,. \tag{8.20}$$

(We note that the singly excited determinants $\hat{\mathcal{A}}[\varphi(1)\alpha(1)\psi(2)\beta(2)]$ and $\hat{\mathcal{A}}[\psi(1)\alpha(1)\varphi(2)\beta(2)]$ are antisymmetric with respect to the inversion and are absent from the expansion of the ground state CI wave function.)

Expanding Ψ in (8.20) we get

$$\begin{aligned}
\Psi &= \hat{\mathcal{A}}[\varphi(1)\alpha(1)\varphi(2)\beta(2)] + c\hat{\mathcal{A}}[\psi(1)\alpha(1)\psi(2)\beta(2)] \\
&= \mathcal{N}^2\hat{\mathcal{A}}\,\{[\chi_A(1)+\chi_B(1)]\alpha(1)[\chi_A(2)+\chi_B(2)]\beta(2)\} \\
&\quad + c\mathcal{N}'^2\hat{\mathcal{A}}\,\{[\chi_A(1)-\chi_B(1)]\alpha(1)[\chi_A(2)-\chi_B(2)]\beta(2)\} \\
&= \mathcal{N}^2\Big\{\hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_A(2)\beta(2)] + \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_A(2)\beta(2)] \\
&\quad + \hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_B(2)\beta(2)] + \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_B(2)\beta(2)]\Big\} \qquad (8.21) \\
&\quad + c\mathcal{N}'^2\Big\{\hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_A(2)\beta(2)] - \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_A(2)\beta(2)] \\
&\quad - \hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_B(2)\beta(2)] + \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_B(2)\beta(2)]\Big\} \\
&= (\mathcal{N}^2 + \mathcal{N}'^2 c)\underbrace{\Big\{\hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_A(2)\beta(2)] + \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_B(2)\beta(2)]\Big\}}_{\text{ionic terms } (\mathrm{H^-H^+ \text{ and } H^+H^-})} \\
&\quad + (\mathcal{N}^2 - \mathcal{N}'^2 c)\underbrace{\Big\{\hat{\mathcal{A}}[\chi_A(1)\alpha(1)\chi_B(2)\beta(2)] + \hat{\mathcal{A}}[\chi_B(1)\alpha(1)\chi_A(2)\beta(2)]\Big\}}_{\text{covalent terms}}.
\end{aligned}$$

The covalent terms alone represent the so-called Heitler–London wave function. The birth of quantum chemistry as a branch of science may be dated by the work of Heitler and London (1927), which represented the first example in which the existence of a chemical bond was really explained, and described semiquantitatively, on the basis of the underlying laws of physics (quantum mechanics).

The single-determinant MO wave function Ψ_0 contains too large an amount of ionic terms—it does not account for the "left–right" electron correlation.

(This causes the "dissociation catastrophe" discussed in Section 4.3 of Chapter 6.) At the same time the Heitler–London wave function containing only covalent terms may be considered "overcorrelated." The optimal amount of left–right correlation is given by the solution of the CI problem (optimization of the coefficient c); it is also equivalent to the optimization of the weight of ionic terms, as it was first performed by Weinbaum (1933).

3. Methods of electron correlation calculations

A comprehensive treatment of the techniques used for calculating electron correlation is out of the scope of the present book. We shall only discuss very briefly the most widely used methods by pointing out those aspects which seem important from a general theoretical point of view.

3.1. The CI method

The CI method is conceptually (but, of course, not practically) very simple. One writes out the CI expansion (8.5) corresponding to the given set of one-electron basis functions and optimizes the coefficients c_I variationally by solving the respective secular equation (see Section 1 of Chapter 3). Often the many-electron functions used as basis functions of the linear variational problem are not the individual determinants but are combinations of them that represent pure spin functions (eigenvectors of operators $\hat{S}_z$ and $\hat{S}^2$) and/or correspond to the spatial symmetry of the system. The use of such spin- and symmetry-adapted configurations can reduce the dimension of the secular equation considerably.

The method in which one solves the eigenvalue equation in terms of all the determinants (or spin-adapted configurations) that can be constructed by using a given one-electron basis is called "full CI" (FCI) method—it represents the exact solution of the variational problem for the given one-electron basis set. One can consider that the basis set used defines a model problem with the respective model Hamiltonian, and the full CI method gives the *exact solution(s)* of the Schrödinger equation written out for this *model Hamiltonian*. This may be formulated as follows: instead of solving the Schrödinger equation $\hat{H}\Psi = E\Psi$, one solves the Schrödinger equation with the projected Hamiltonian

$$\hat{\mathcal{P}}\hat{H}\hat{\mathcal{P}}\Psi = E\Psi \ , \tag{8.22}$$

where $\hat{\mathcal{P}}$ is the projection operator on the subspace of all antisymmetric N-electron wave functions that can be expanded by using the given one-electron basis set. (As the basis set increases and becomes complete, $\hat{\mathcal{P}} \to \hat{1}$, and the solutions of the FCI problem tend to the exact solutions of the original Schrödinger equation.) Obviously, when working in a given finite basis, one need not use explicitly the projector $\hat{\mathcal{P}}$, as one considers only wave functions for which $\hat{\mathcal{P}}\Psi \equiv \Psi$.

In accord with the previous discussion, the results of FCI calculations depend only on the subspace of the one-electron functions applied but are independent of the one-electron basis that is selected in that subspace and is used to construct the determinants (or their spin- and/or symmetry-adapted combinations) serving as the basis functions for treating the N-electron linear variational problem. The use of determinants (configurations) built up directly of nonorthogonal one-electron basis orbitals usually applied in quantum chemistry would be, obviously, not practical. Any orthogonalized set of orbitals spanning the same one-electron subspace is appropriate; usually one applies the occupied and virtual Hartree–Fock orbitals, which permits a fast convergence to be achieved in the iterative solution of the eigenvalue equation.

As discussed in the previous section, the number of determinants increases extremely quickly with the increase of the basis set and of the number of electrons, and one can not perform in practice FCI calculations except for rather small systems treated with very small basis sets—even if the most powerful computers are used. Nevertheless, these calculations are very important from the methodological point of view, because they provide us the benchmarks to which all other (approximate) results should be compared.

Whereas full CI calculations are prohibitively expensive in essentially all problems of practical chemical interest, at the same time there is no real need in doing them (except of reference purposes mentioned) because the absolute majority of configurations (determinants) does not contribute to the wave function in any appreciable extent and does not practically influence the results, especially as far as the energetic properties (which are usually of largest importance for applications) are concerned.

If one uses the expansion in (8.5) applying the Hartree–Fock orbitals, then it follows from Nesbet's theorem discussed in the previous section that—besides the HF ground-state wave function Ψ_0—the by far most important determinants are the doubly excited ones: if we knew their coefficients in the exact CI expansion then we would be able to compute the exact FCI energy.

According to the previous point, the simplest CI method for treating ground states of closed-shell molecules would be that in which one considers double excitations only. In general, this has not been used, however, because the singly excited determinants (configurations) do not interact with the Hartree–Fock ground state because of the Brillouin theorem, but each singly excited determinant (configuration) interacts with several doubly excited ones, and thus they have a nonnegligible indirect influence on the calculated correlation energy. At the same time the number of singly excited configurations is much smaller than the number of doubly excited ones; therefore their inclusion does not increase appreciably the necessary computational work. (The number of singly excited states is proportional to $n_{\mathrm{occ}} \times n_{\mathrm{virt}}$, where n_{occ} and n_{virt} are the

number of occupied and virtual HF orbitals, respectively, whereas the number of doubly excited ones, obviously, is proportional to $n_{\rm occ}^2 \times n_{\rm virt}^2$.)

Inclusion of all singly and doubly excited determinants in the CI calculation (the so-called CI–SD method) permits one to account for a significant part of the full CI correlation energy of small molecules; this method has, however, the serious disadvantage that it is not size consistent (c.f. Section 4.1 of Chapter 4). If we consider two noninteracting systems, each treated at the CI–SD level as subsystems of a formal supersystem, then the wave function ought to be the antisymmetrized product of the two subsystems' wave functions. However, this product of two CI–SD wave function is not a CI–SD wave function: it also contains terms that formally represent quadruple excitations and are not included in the CI–SD description of the supersystem. For that reason the CI–SD method is not widely used any more—except, maybe, if approximate corrections for the size consistency are introduced. One should keep in mind that a "truncated" CI calculation (i.e., a CI not including all the excitations) can never be strictly size consistent. A CI scheme including excitations up to quadruple ones would probably exhibit a sufficiently small size-consistency error; such CI calculations would be, however, prohibitively expensive for problems of chemical interest, owing to the very large number of different triple and quadruple excitations. At the same time, strictly size-consistent results of similar quality can be obtained by using perturbational or "coupled cluster" methods (Sections 3.3 and 3.4) with a much smaller computational effort.

When doing practical CI calculations, one does not build up explicitly the Hamiltonian (CI) matrix $\mathbf{H}$ with the elements $H_{IJ} = \langle \Psi_I | \hat{H} | \Psi_J \rangle$, where Ψ_I and Ψ_J are the N-electron determinants (configurations), but uses the so-called "direct" method, in which one calculates the vectors $\mathbf{y} = \mathbf{Hx}$ for vectors $\mathbf{x}$ containing the CI expansion coefficients using directly the one- and two-electron integrals over the orthonormalized orbitals φ_i. For that reason one has to go through the individual integrals and find those elements of the matrix-vector product $\mathbf{Hx}$ to which a given integral can contribute and calculate these contributions without calculating explicitly the matrix elements H_{IJ}.

3.2. MC–SCF methods

Instead of performing CI calculations with a very large number of configurations, most of which may finally appear unimportant, one may try to use a limited number of appropriately selected configurations that involve only a part of the available one-electron orbitals but optimize not only the CI expansion coefficients but also the selected one-electron orbitals. This approach is called "multiconfiguration SCF" (MC–SCF) method; it combines the concepts of the SCF and CI schemes. Accordingly, an MC–SCF calculation consists of two separate parts, both of which should be repeated cyclically until convergence is achieved: optimization of the CI coefficients for the current one-electron

orbitals, which is a standard linear variational problem, and optimization of the one-electron orbitals for the current CI coefficients, which is a procedure similar to the ROHF method discussed in Section 6 of Chapter 6.

The previous definition covers a wide variety of possible MC–SCF schemes. One limit is the use of only a few—maybe only two—determinants that are inevitable to exclude the static correlation (Section 2) present in the problem considered. Another limit is the so-called "complete active space SCF" (CAS–SCF) method, in which one selects a few electrons and a small number of orbitals for which essentially a full CI problem is solved, while the remaining electrons are assigned to doubly filled orbitals.

One may also consider the so-called APSG method ("antisymmetrized product of strongly orthogonal geminals") as a special MC–SCF one. In the APSG method one utilizes the fact that many chemical systems can be viewed as containing singlet-coupled electron pairs. The respective wave function Ansatz (trial wave function) is then

$$\Psi(1,2,...,N) = \hat{\mathcal{A}} \prod_{i=1}^{N/2} g_i(2i-1,2i) \frac{1}{\sqrt{2}} [\alpha(2i-1)\beta(2i) - \beta(2i-1)\alpha(2i)] \tag{8.23}$$

and one has to optimize the *two-electron functions* ("geminals") g_i, $i = 1, 2, \ldots N/2$. (Each geminal is a symmetric function of its arguments.)

Practically one can treat wave functions of this type only by introducing the requirement (limitation) called "strong orthogonality":

$$\int g_i(1,2) g_j(1,2) dv_1 = \int g_i(1,2) g_j(1,2) dv_2 = 0 \text{ if } i \neq j \tag{8.24}$$

—i.e., integrating the product of different geminals over coordinates of either electrons gives 0. The geminals are usually expanded in terms of one-electron functions φ_k

$$g_i(1,2) = \sum_{k,l} c^i_{kl} \varphi_k(1) \varphi_l(2) \tag{8.25}$$

with $c^i_{kl} = c^i_{lk}$; thus function (8.23) is essentially a special case of the expansion in (8.5). Strong orthogonality in (8.24) can be ensured trivially if one assigns to the individual geminals disjunct subsets of the orthonormalized one-electron basis $\{\varphi_k\}$.

The possibility of considering the APSG wave function as a special MC–SCF one does not mean that one performs its optimization by expanding it as a sum of determinants. But in the APSG case, too, one has to optimize both the orbitals φ_k, φ_l and the linear coefficients c^i_{kl}.

We may also mention here the "generalized valence bond" (GVB) method of Goddard, which is a simple case of APSG: only one pair of one-electron

orbitals is used for each geminal, but these orbitals need not be orthogonal to each other. (They still are required to be orthogonal to each orbital used for other geminals.) After orthogonalizing these orbitals, one can express this wave function in such a manner that for each pair of electrons one uses four terms, quite similar to those discussed above in Section 2 for the H_2 molecule. (Such a wave function written out for the single electron pair of the H_2 molecule is the classical Coulson–Fischer wave function of 1949. In the minimal basis case it is equivalent to the CI wave function given in Section 2.)

3.3. The MBPT or Møller–Plesset method

One can account for (a part of) electron correlation by applying perturbational theory (Chapter 4). The most widely used version is often called—for historical reasons—"many body perturbation theory" (MBPT), referring to the diagrammatic technique of derivations adopted from theoretical physics. MBPT is none other than the application of the Rayleigh–Schrödinger perturbation theory discussed in detail in Chapter 4 to the electron correlation problem by using the so-called Møller–Plesset partitioning of the Born–Oppenheimer Hamiltonian. Accordingly, the nth order perturbational results obtained in this framework are usually denoted by the acronym MPn (MP2, MP3 and so on), MP2 and MP4 being the most commonly used variants.

In the Møller–Plesset partitioning the unperturbed Hamiltonian is selected as the sum of the Fock operators for all the electrons in the system:

$$\hat{H}^0 = \sum_{i=1}^{N} \hat{F}(i) \; . \tag{8.26}$$

Accordingly, any determinant wave function built up of the (occupied and/or virtual) *canonical* Hartree–Fock orbitals is an eigenvector of this unperturbed Hamiltonian with an eigenvalue equal to the sum of the orbital energies corresponding to the orbitals of which the determinant in question is built up. (Note that $\hat{H}^0$ commutes with the antisymmetrizer—c.f. Section 4.2 of Chapter 5.)

Owing to the invariance properties of the Fock operator (Section 2.1 of Chapter 6), one can construct *the same* $\hat{H}^0$ not only by using the canonical orbitals but also applying any set of orthogonal orbitals (e.g., localized ones) spanning the occupied subspace. However, only the canonical orbitals provide that all the "ordered configurations" built up of the Hartree–Fock orbitals (c.f. Section 1) are eigenfunctions of *this* unperturbed Hamiltonian. Nevertheless, when treating really large systems it may be meaningful to use localized orbitals and utilize the pronounced local character of the pair correlation energies—c.f. Section 2—even at the expense that one has to solve some more involved equations for calculating the perturbational corrections.

Strictly speaking, (8.26) is valid only in the RHF case when doubly filled orbitals are used and one has to deal with a single Fock operator—this is the

"RMPn" ("restricted" Møller-Plesset) theory, or the MPn theory in a narrow sense. If one uses the UHF theory for the single-determinant wave functions, as it is usually the case for odd-electron or other open-shell species, then one has separate Fockians for the orbitals of spin α and β. Then the respective unperturbed Hamiltonian can be written as

$$\hat{H}^0 = \sum_{i=1}^{N} \left[\hat{F}^\alpha(i)\delta_{\alpha\gamma_i} + \hat{F}^\beta(i)\delta_{\beta\gamma_i} \right] , \tag{8.27}$$

where the Kronecker symbols indicate that the operators $\hat{F}^\alpha(i)$ and $\hat{F}^\beta(i)$ act on the spin orbitals of the respective spin only. (The definition in (8.27) remains applicable in the RHF case, too, for which $\hat{F}^\alpha = \hat{F}^\beta = \hat{F}$.)

According to the definition in (8.27), the zero-order energy can be written as

$$E_0 = \sum_{i}^{(\text{occ.})} \varepsilon_i^{\gamma_i} , \tag{8.28}$$

where "(occ.)" indicates that the summation should be performed for all orbitals occupied in the Hartree–Fock ground-state determinant. (Somewhat later the analogous notation "(virt.)" will also be used for summations over all the virtual orbitals.)

In accord with the usual practice in the perturbation theory—c.f. (4.2)—we write $\hat{H} = \hat{H}^0 + \hat{V}$ i.e., the perturbation operator is

$$\hat{V} = \hat{H} - \hat{H}^0 . \tag{8.29}$$

(We have not introduced a formal perturbational parameter λ here.)

The first-order energy correction is the expectation value of the perturbation operator over the unperturbed wave function—c.f. Eq. (4.23):

$$E^{(1)} = \langle \Psi_0 | \hat{V} | \Psi_0 \rangle = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle - \sum_{i}^{(\text{occ.})} \varepsilon_i^{\gamma_i} . \tag{8.30}$$

(The Hartree–Fock orbitals are assumed orthonormalized, thus $\langle \Psi_0 | \Psi_0 \rangle = 1$.)

According to (8.28) and (8.30), the sum of the zero- and first-order energies equals the Hartree–Fock energy, i.e., the expectation value of the total Hamiltonian over the unperturbed (Hartree–Fock) wave function.

The second-order energy correction (4.30) can be written in terms of the ordered configurations as

$$E^{(2)} = \sum_{\substack{K \\ (K \neq 0)}} \frac{-|\langle \Psi_K | \hat{H} | \Psi_0 \rangle|^2}{E_K - E_0} = \sum_{i<j}^{(\text{occ.})} \sum_{u<v}^{(\text{virt.})} \frac{-|\,[uv||ij]'\,|^2}{\varepsilon_u^{\gamma_u} + \varepsilon_v^{\gamma_v} - \varepsilon_i^{\gamma_i} - \varepsilon_j^{\gamma_j}} , \tag{8.31}$$

where the sum runs over the doubly excited configurations, because—as discussed in Section 2—only the doubly excited determinants interact with the Hartree–Fock ground state. (We have utilized the fact that $\langle\Psi_K|\hat{H}^0|\Psi_0\rangle = 0$ if $K \neq 0$, thus $\langle\Psi_K|\hat{V}|\Psi_0\rangle = \langle\Psi_K|\hat{H}|\Psi_0\rangle$.)

In (8.31) we have introduced the simplified notation "with double bars"

$$[ab||cd]' = \langle\psi_a(1)\psi_b(2)|\frac{1}{r_{12}}|\psi_c(1)\psi_d(2)\rangle - \langle\psi_a(1)\psi_b(2)|\frac{1}{r_{12}}|\psi_d(1)\psi_c(2)\rangle \ . \tag{8.32}$$

(The prime is used to stress the fact that the integrations include summations over the spins and to distinguish from the analogous integrals written out for pure spatial orbitals.)

The result in (8.31) is completely in line with Nesbet's theorem, because the first-order wave function is a linear combination of the doubly excited determinants with the coefficients that are equal, according to Eq. (4.28), to

$$c_{ij}^{uv} = -\frac{[uv||ij]'}{\varepsilon_u^{\gamma_u} + \varepsilon_v^{\gamma_v} - \varepsilon_i^{\gamma_i} - \varepsilon_j^{\gamma_j}} \ . \tag{8.33}$$

In order to get the second-order energy expression in a form more suitable for the practical calculations, one has to rewrite (8.31) in terms of the spatial orbitals. To do this, let us assume that the Hartree–Fock wave function is built up of n_α orbitals a_i filled with spin α and n_β orbitals b_i filled with spin β. (Here we consider the UHF case; it includes the RHF one as a special case with $n_\alpha = n_\beta = n$; $a_i = b_i = \varphi_i$.) Let us assume that the spin orbitals are ordered in such a manner that the first n_α occupied orbitals bear spin α and then follow the occupied spin orbitals of spin β. A similar numbering may be introduced for the virtual orbitals, as well. In this case the sum over the spin orbitals ψ_i and ψ_j with the restriction $i < j$ used in (8.31) will be decomposed into three sums—two for the case when ψ_i and ψ_j are of the same spin and one for the case when ψ_i is of spin α and ψ_j is of spin β. The first two sums should be performed for the pairs of orbitals a_i, a_j and b_i, b_j, both with the restriction $i < j$, whereas the third sum is for the pairs of orbitals a_i, b_j with no restriction on the values of i and j, because—according to the convention applied—every orbital with spin α has a number lower than any orbital with spin β. Furthermore, one has to take into account that if both occupied orbitals occurring in a given terms are of spin α (β), then virtual orbitals should bear spin α (β), too; otherwise the integrals will not survive the summation over the spins. Similarly, if one occupied orbital is of spin α and another is of spin β, then the same must hold for the virtual orbitals. In this case the second term on the right-hand side of (8.32) vanishes as a consequence of the orthogonality of the spin functions. The first integral on the right-hand side of (8.32) will be

denoted as

$$[ab|cd]' = \langle \psi_a(1)\psi_b(2)|\frac{1}{r_{12}}|\psi_c(1)\psi_d(2)\rangle \tag{8.34}$$

—i.e., without the double vertical bar.

In the following we shall omit the primes because we conserve only terms for which the summation over the spins gives a factor 1 and the integrals are assumed to be over the spatial orbitals. Thus we obtain the general MP2 correlation energy formula, corresponding to the UHF unperturbed wave function and usually denoted UMP2:

$$\begin{aligned}\Delta E_{\mathrm{UMP2}} = & -\left\{\sum_{i<j}^{(\mathrm{occ.}\alpha)}\sum_{u<v}^{(\mathrm{virt.}\alpha)} \frac{|[a_i a_j||a_u a_v]|^2}{\varepsilon_u^a + \varepsilon_v^a - \varepsilon_i^a - \varepsilon_j^a}\right. \\ & + \sum_{i<j}^{(\mathrm{occ.}\beta)}\sum_{u<v}^{(\mathrm{virt.}\beta)} \frac{|[b_i b_j||b_u b_v]|^2}{\varepsilon_u^b + \varepsilon_v^b - \varepsilon_i^b - \varepsilon_j^b} \\ & \left. + \sum_{i}^{(\mathrm{occ.}\alpha)}\sum_{j}^{(\mathrm{occ.}\beta)}\sum_{u}^{(\mathrm{virt.}\alpha)}\sum_{v}^{(\mathrm{virt.}\beta)} \frac{|[a_i b_j|a_u b_v]|^2}{\varepsilon_u^a + \varepsilon_v^b - \varepsilon_i^a - \varepsilon_j^b}\right\}.\end{aligned} \tag{8.35}$$

As the RHF wave function is a special case of the UHF one, this formula can be (and often is) used in the RHF case, too, by substituting in it $a_i = b_i = \varphi_i$. In the RHF case some further simplifications are also possible, however.

In the RHF case the first two sums in (8.35) become identical. One can combine them into a single sum with summation over i, j; $i \neq j$, by interchanging i and j in the second sum and utilizing the identity $[\varphi_j\varphi_i||\varphi_u\varphi_v]^2 \equiv (-[\varphi_i\varphi_j||\varphi_u\varphi_v])^2 = [\varphi_i\varphi_j||\varphi_u\varphi_v]^2$. (Now we assume that the orbitals are real and there is no need to indicate the absolute values.) The third sum in (8.35) can be decomposed into three, corresponding to the cases $u < v, u = v$, and $u > v$. We interchange the indices u and v in the last case, and separate out the terms with $i = j$. ($[\varphi_i\varphi_i|\varphi_v\varphi_u] \equiv [\varphi_i\varphi_i|\varphi_u\varphi_v]$.) Then we expand the square $[\varphi_i\varphi_j||\varphi_u\varphi_v]^2 = ([\varphi_i\varphi_j|\varphi_u\varphi_v] - [\varphi_i\varphi_j|\varphi_v\varphi_u])^2$ and combine the sums over i, j; $i \neq j$ into a single sum, by taking into account the identity $[\varphi_i\varphi_j|\varphi_v\varphi_u] \equiv [\varphi_j\varphi_i|\varphi_u\varphi_v]$ and the fact that i and j are independent summation indices that may be interchanged. We get

$$\begin{aligned}\Delta E_{\mathrm{RMP2}} = & -\left\{\sum_{\substack{i,j\\(i\neq j)}}^{(occ.)}\sum_{u<v}^{(virt.)} \frac{2[\varphi_i\varphi_j|\varphi_u\varphi_v](2[\varphi_i\varphi_j|\varphi_u\varphi_v] - [\varphi_i\varphi_j|\varphi_v\varphi_u])}{\varepsilon_u + \varepsilon_v - \varepsilon_i - \varepsilon_j}\right. \\ & \left. + \sum_{i}^{(occ.)}\sum_{u<v}^{(virt.)} \frac{2[\varphi_i\varphi_i|\varphi_u\varphi_v]^2}{\varepsilon_u + \varepsilon_v - 2\varepsilon_i} + \sum_{i,j}^{(occ.)}\sum_{u}^{(virt.)} \frac{[\varphi_i\varphi_j|\varphi_u\varphi_u]^2}{2\varepsilon_u - \varepsilon_i - \varepsilon_j}\right\}.\end{aligned} \tag{8.36}$$

This is one of the equivalent forms suitable for programming the RMP2 correlation energy correction.

By using the expressions obtained in the framework of the Rayleigh–Schrödinger perturbation theory (Chapter 4), one can, in principle, obtain the expressions of the higher order MP energy corrections, as well. As the number of terms increases very rapidly, these derivations are usually performed by using a diagrammatic technique permitting an appropriate "bookkeeping." We shall not go into a detailed discussion of these higher order terms but will only note that the third-order energy formula contains only doubly excited configurations ("doubles") similar to the second order energy correction. (It is the usual opinion that the MP3 results are not well balanced.)

The fourth-order energy correction, however, requires the knowledge (at least, c.f. Section 1.4.2 of Chapter 4) of the second order wave function that involves all configurations that interact with the Hartree–Fock ground state and the doubly excited configurations entering the first-order wave function correction, i.e., all the configurations up to quadruply excited ones. Although the role of the quadruply excited configurations ("quadruples") is mostly to ensure size consistency and their treatment does not represent a very large practical problem, the treating of the numerous triply excited configurations ("triples") appears more difficult (requires more computer time). As the effect of the triples is usually quite moderate, one sometimes omits them from consideration and gets the approximate MP4 scheme usually denoted MP4(SPQ), as opposed to the "full" MP4 that may, accordingly, be denoted as MP4(SPTQ) to stress that all excitations are included.

We note that the $1s$ shells of atoms are often "frozen" i.e., omitted from consideration in the MP theory and other correlation energy calculations. Although the correlation energy that one can attribute to the $1s$ orbitals is considerable, it does not depend practically on the chemical environment of the atoms and does not influence the energy *differences* in which we are interested.

Owing to the presence of the self-repulsion terms in the Fock operator, the virtual orbital energies are shifted upward (Section 3.1 of Chapter 6) and this leads to relatively large energy denominators in the MP formulas. As a consequence, in the usual MP calculations of molecules near their equilibrium configurations, one gets only a part of the correlation energy. (Contrary to this, the so-called Epstein–Nesbet partitioning, in which the off-diagonal CI matrix elements constitute the perturbation, can sometimes overshoot and give energy values that are lower than the exact ones.)

We shall stress once again that the perturbational methods are not variational and do not provide, therefore, upper bounds of the energy. This can lead to completely wrong MP potential curves far from the equilibrium conformations —e.g., when a homolytical dissociation is considered: the MP methods are not well suited to deal with the static correlation (c.f. Section 2).

3.4. The Coupled Cluster Approach

If one has several noninteracting subsystems, then—as we have discussed several times—the overall wave function is an (antisymmetrized) product of the subsystems' wave functions. This means that in order to ensure size consistency, one has to include in the wave function such excitations that are—at least formally—of higher type than those used for the description of the individual subsystems. From the point of view of the supersystem, the appearance of those higher excitations that are combinations of independent lower order ones can be considered as a kinematical effect; they should be distinguished from those higher order excitations that reflect some true dynamical effects in the system. However, these alternatives represent only two limiting cases, and in practice these two types of excitations are usually intermixed. If two subsystems approach each other, then increasingly intensive interactions appear between them and, as a consequence, the coefficients of terms corresponding to higher-order excitations will deviate more and more from the products of the respective lower-order intrasystem excitation coefficients. In some cases one may expect that these coefficients remain close to that product even at smaller distances, because the true (dynamical) higher-order effects are small; in other cases significant changes will take place. For instance, singly excited configurations contribute to the full CI wave functions of isolated subsystems only to a small degree because they appear only as indirect consequences of their interactions with the doubly (and triply) excited configurations. The wave function of the supersystem consisting of two noninteracting subsystems will contain the double excitations corresponding to simultaneous single excitations in the two subsystems with coefficients equal to the product of the coefficients of the respective single excitations in the individual subsystems. (Thus these coefficients will be very small.) In accord with the size–consistency requirement, these double excitations will not contribute at all to the resulting correlation energy because the corresponding integrals in (8.14) vanish owing to the noninteracting character of the subsystems. At a large but finite distance of the subsystems the integrals in question do not vanish any more and this results in a rapid deviation of the coefficients at the doubly excited configuration from the simple product value; also a contribution to the correlation energy appears. (The correlation energy component connected with these double excitations accounts for the important phenomenon of dispersion.) At even smaller distances it becomes meaningless to speak about two intra-system excitations and one has a gradual transition to a conventional double excitation that is the result of the electron correlation.

It is desirable, therefore, to apply a formalism in which one can separate from each other the kinematical and dynamical contributions in the higher- order excitations of the actual wave function. This permits one to perform calculations by including all kinematical higher-order excitations in an automated fashion

(providing in this manner the size consistency) and to compute the genuine, dynamical higher-order effects to the extent as much as they are required by the physics of the problem (and are affordable computationally).

In order to develop the overall frame of an appropriate formalism, let us consider at first the simple model of M noninteracting subsystems. Let us assume that the wave function of an individual subsystem can be given, by using the intermediate normalization, as

$$\Psi^{(i)} = \Psi_0^{(i)} + \varphi^{(i)} = \Psi_0^{(i)} + q_i \Psi_1^{(i)} = (1 + \hat{X}_i)\Psi_0^{(i)} , \tag{8.37}$$

where the excitation operator $\hat{X}_i$ can most easily be defined by using the "bra-ket" formalism:

$$\hat{X}_i = q_i |\Psi_1^{(i)}\rangle\langle\Psi_0^{(i)}| . \tag{8.38}$$

($\Psi_0^{(i)}$ and $\Psi_1^{(i)}$ are assumed orthonormalized, $q_i = \text{const.}$)

The wave function of the supersystem is

$$\begin{aligned} \Psi &= \Psi^{(1)}\Psi^{(2)}\ldots\Psi^{(M)} = (1+\hat{X}_1)\Psi_0^{(1)}(1+\hat{X}_2)\Psi_0^{(2)}\ldots(1+\hat{X}_M)\Psi_0^{(M)} \\ &= (1+\hat{X}_1)(1+\hat{X}_2)\ldots(1+\hat{X}_M)\Psi_0 = \prod_{i=1}^{M}(1+\hat{X}_i)\Psi_0 \end{aligned} \tag{8.39}$$

with

$$\Psi_0 = \Psi_0^{(1)}\Psi_0^{(2)}\ldots\Psi_0^{(M)} . \tag{8.40}$$

In (8.39) we have utilized the fact that $\hat{X}_i\Psi_0^{(j)} = \Psi_0^{(j)}\hat{X}_i$ and $\hat{X}_i\hat{X}_j = \hat{X}_j\hat{X}_i$ if $i \neq j$, as being related to different variables. (Also, we recall in this connection that the antisymmetrization between noninteracting subsystems is immaterial—c.f. Section 4.3 of Chapter 5.)

Expanding the product in (8.39), one easily gets

$$\Psi = \left(1 + \sum_{i=1}^{M}\hat{X}_i + \sum_{i<j}\hat{X}_i\hat{X}_j + \sum_{i<j<k}\hat{X}_i\hat{X}_j\hat{X}_k + \ldots\right)\Psi_0 . \tag{8.41}$$

This can be transformed, utilizing the commutation property of the excitation operators $\hat{X}_i$, as

$$\begin{aligned} \Psi &= \left(1 + \sum_{i=1}^{M}\hat{X}_i + \frac{1}{2!}\sum_{i,j=1}\hat{X}_i\hat{X}_j + \frac{1}{3!}\sum_{i,j,k=1}\hat{X}_i\hat{X}_j\hat{X}_k + \ldots\right)\Psi_0 \qquad (8.42) \\ &= \left(1 + \sum_{i=1}^{M}\hat{X}_i + \frac{1}{2!}\sum_{i=1}\hat{X}_i\sum_{j=1}\hat{X}_j + \frac{1}{3!}\sum_{i=1}\hat{X}_i\sum_{j=1}\hat{X}_j\sum_{k=1}\hat{X}_k + \ldots\right)\Psi_0 . \end{aligned}$$

The previous transformation is possible because the terms with $i = j$, etc., need not to be omitted from these sums as they vanish anyway:

$$\hat{X}_i \hat{X}_i = q_i^2 |\Psi_1^{(i)}\rangle\langle\Psi_0^{(i)}|\Psi_1^{(i)}\rangle\langle\Psi_0^{(i)}| = 0 \tag{8.43}$$

according to the definition in (8.38) and the orthogonality of states $\Psi_0^{(i)}$ and $\Psi_1^{(i)}$.

It is known that functions of operators are defined via their power series expansions; thus for any "well-behaved" linear operator $\hat{L}$ one has

$$e^{\hat{L}} = 1 + \hat{L} + \frac{1}{2!}\hat{L}^2 + \frac{1}{3!}\hat{L}^3 + \dots \,. \tag{8.44}$$

Accordingly, (8.42) can be rewritten as

$$\Psi = e^{\hat{X}} \Psi_0 \tag{8.45}$$

with the definition

$$\hat{X} = \sum_{i=1}^{M} \hat{X}_i \,. \tag{8.46}$$

Powers $\hat{X}^k$ with $k > M$ vanish as a consequence of (8.43); therefore the series expansion of $e^{\hat{T}}$ truncates after the term $\hat{X}^M$. In particular, if there are only one and two subsystems ($M = 1$, and $M = 2$) we recover, respectively

$$e^{\hat{X}} = 1 + \hat{X}_1 \tag{8.47}$$

and

$$\begin{aligned} e^{\hat{X}} &= 1 + \hat{X} + \frac{1}{2}\hat{X}^2 = 1 + \hat{X}_1 + \hat{X}_2 + \frac{1}{2}(\hat{X}_1 + \hat{X}_2)^2 \\ &= 1 + \hat{X}_1 + \hat{X}_2 + \hat{X}_1\hat{X}_2 = (1 + \hat{X}_1)(1 + \hat{X}_2) \end{aligned} \tag{8.48}$$

as expected. (When expanding the square in (8.48), we have utilized the fact that $\hat{X}_i^2 = 0$.)

Whereas the wave function of independent subsystems is a "multiplicatively separable" quantity, the use of the exponential formulation ("cluster expansion") permits one to deal with the excitation operators $\hat{T}_i$ representing "additively separable" objects. It is important to stress once again that all these transformations are possible because the different operators $\hat{X}_i$ commute with each other.

The "coupled cluster approach" (CCA) consists in applying the previous scheme to the electron correlation problem. In CCA one uses the cluster expansion of the wave function in the form

$$\Psi = e^{\hat{T}} \Psi_0 \,. \tag{8.49}$$

Here Ψ_0 is the "reference" wave function—in the simplest and most common case the RHF ground state wave function of the system—and the global excitation operator $\hat{T}$ is the sum

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots \ . \tag{8.50}$$

The different terms of this expansion are defined as follows. Operator $\hat{T}_1$ generates the different singly excited configurations. Similarly, $\hat{T}_2$ generates the doubly excited configurations but without those terms that are the result of pairs of simultaneous single excitations, and are already included in $e^{\hat{T}_1}$. (The different operators $\hat{T}_i$ commute—*vide infra*—, thus the exponent in (8.49) can be written as $e^{\hat{T}_1}e^{\hat{T}_2}e^{\hat{T}_3}\dots$.) Analogously, $\hat{T}_3$ generates the "genuine" triple excitations, i.e., those that cannot be presented as combinations of independent single and double excitations. The expansion can be continued in principle but in practice one usually does not go beyond $\hat{T}_3$. A detailed form of these operators is

$$\begin{aligned}
\hat{T}_1 &= \sum_{i}^{\text{(occ.)}} \sum_{u}^{\text{(virt.)}} t_i^u \hat{T}_i^u \\
\hat{T}_2 &= \sum_{i<j}^{\text{(occ.)}} \sum_{u<v}^{\text{(virt.)}} t_{ij}^{uv} \hat{T}_{ij}^{uv} \\
\hat{T}_3 &= \sum_{i<j<k}^{\text{(occ.)}} \sum_{u<v<w}^{\text{(virt.)}} t_{ijk}^{uvw} \hat{T}_{ijk}^{uvw} \ .
\end{aligned} \tag{8.51}$$

In (8.51) operator $\hat{T}_i^u$ is an operator exciting one electron from the occupied spin orbital ψ_i to the virtual spin orbital ψ_u. It performs this substitution for any determinant wave function in which ψ_i is occupied and ψ_u is not; if these conditions are not met, the result of acting with $\hat{T}_i^u$ on a determinant is 0. Usually, operators of this type are written out in term of "creation" and "annihilation" operators of the so-called "second quantization" formalism; we will not need here their explicit form. Similarly, T_{ij}^{uv} $(i \neq j; u \neq v)$ is an excitation operator replacing the occupied spin orbitals ψ_i and ψ_j by the virtual spin orbitals ψ_u and ψ_v, respectively. Obviously, when acting on a general determinant, this again leads to nonzero result only if both ψ_i and ψ_j are present in it and both ψ_u and ψ_v are absent. Operator $\hat{T}_{ijk}^{uvw}$ is defined analogously. The coefficients t_i^u, t_{ij}^{uv}, and t_{ijk}^{uvw} are called "cluster amplitudes."

The product of two (or more) excitation operators $\hat{T}_i^u$, $\hat{T}_{ij}^{uv}$, etc., indicates that the respective substitutions are to be performed consecutively, and is, therefore, a higher-order substitution operator—or it equals 0 if the substitutions are not possible (e.g., $t_i^u t_i^v = t_i^u t_j^u = 0$, because one cannot excite two electrons

from the same spin orbital ψ_i or to the same spin orbital ψ_u). Therefore, $\hat{T}_1^2$ consists of two-electron excitations, $\hat{T}_1\hat{T}_2$ and $\hat{T}_1^3$ of three-electron ones, etc. Obviously, the excitation operators commute, as the order in which the different substitutions are performed is immaterial. Consequently, the operators $\hat{T}_i$ representing linear combinations of excitation operators commute, as well.

Expanding the exponent in (8.49) into series, one gets

$$\Psi = \left(1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \dots\right)\Psi_0 \,. \tag{8.52}$$

The different powers in (8.52) should be evaluated by using the expansion (8.50), i.e.,

$$\hat{T}^2 = \hat{T}_1^2 + 2\hat{T}_1\hat{T}_2 + \hat{T}_2^2 + 2\hat{T}_1\hat{T}_3 + 2\hat{T}_2\hat{T}_3 + \hat{T}_3^2 + \dots \tag{8.53}$$

and so on. One can group the terms of the expansion in (8.52) according to the number of electron excitations involved:

$$\Psi = \left(1 + \hat{T}_1 + \frac{1}{2!}\hat{T}_1^2 + \hat{T}_2 + \frac{1}{3!}\hat{T}_1^3 + \hat{T}_1\hat{T}_2 + \hat{T}_3 + \dots\right)\Psi_0 \,. \tag{8.54}$$

Then, substituting the expansions in (8.51) one can collect the terms containing the same excitations. For instance, the two-electron excitations are

$$\begin{aligned}
\frac{1}{2}\hat{T}_1^2 + \hat{T}_2 &= \frac{1}{2}\sum_{i}^{(\mathrm{occ.})}\sum_{u}^{(\mathrm{virt.})} t_i^u\hat{T}_i^u \sum_{j}^{(\mathrm{occ.})}\sum_{v}^{(\mathrm{virt.})} t_j^v\hat{T}_j^v + \sum_{i<j}^{(\mathrm{occ.})}\sum_{u<v}^{(\mathrm{virt.})} t_{ij}^{uv}\hat{T}_{ij}^{uv} \\
&= \frac{1}{2}\sum_{i,j}^{(\mathrm{occ.})}\sum_{u,v}^{(\mathrm{virt.})} t_i^u t_j^v\hat{T}_i^u\hat{T}_j^v + \sum_{i<j}^{(\mathrm{occ.})}\sum_{u<v}^{(\mathrm{virt.})} t_{ij}^{uv}\hat{T}_{ij}^{uv} \\
&= \frac{1}{2}\sum_{i<j}^{(\mathrm{occ.})}\sum_{u<v}^{(\mathrm{virt.})} t_i^u t_j^v\hat{T}_i^u\hat{T}_j^v + \frac{1}{2}\sum_{i<j}^{(\mathrm{occ.})}\sum_{u>v}^{(\mathrm{virt.})} t_i^u t_j^v\hat{T}_i^u\hat{T}_j^v \\
&\quad + \frac{1}{2}\sum_{i>j}^{(\mathrm{occ.})}\sum_{u<v}^{(\mathrm{virt.})} t_i^u t_j^v\hat{T}_i^u\hat{T}_j^v + \frac{1}{2}\sum_{i>j}^{(\mathrm{occ.})}\sum_{u>v}^{(\mathrm{virt.})} t_i^u t_j^v\hat{T}_i^u\hat{T}_j^v \\
&\quad + \sum_{i<j}^{(\mathrm{occ.})}\sum_{u<v}^{(\mathrm{virt.})} t_{ij}^{uv}\hat{T}_{ij}^{uv} \\
&= \sum_{i<j}^{(\mathrm{occ.})}\sum_{u<v}^{(\mathrm{virt.})} (t_i^u t_j^v - t_i^v t_j^u + t_{ij}^{uv})\hat{T}_{ij}^{uv} \,,
\end{aligned} \tag{8.55}$$

where we have interchanged some summation indices to get $i < j$ and $u < v$ in each term and have taken into account that the operator–product $\hat{T}_i^u\hat{T}_j^v$ produces

the same substitutions as $\hat{T}_{ij}^{uv}$, but the product $\hat{T}_i^v \hat{T}_j^u$ results in a determinant with spin orbitals ψ_u and ψ_v interchanged—hence the minus sign at the second term in the parentheses.

Comparing these results with the CI expansion (8.8) that can be written as

$$\Psi = \Psi_0 + \sum_i^{(\text{occ.})} \sum_u^{(\text{virt.})} c_i^u \Psi_i^u + \sum_{i<j}^{(\text{occ.})} \sum_{u<v}^{(\text{virt.})} c_{ij}^{uv} \Psi_{ij}^{uv} + \sum_{i<j<k}^{(\text{occ.})} \sum_{u<v<w}^{(\text{virt.})} c_{ijk}^{uvw} \Psi_{ijk}^{uvw} + \ldots \tag{8.56}$$

one gets

$$c_i^u = t_i^u \tag{8.57}$$

and

$$c_{ij}^{uv} = t_i^u t_j^v - t_i^v t_j^u + t_{ij}^{uv} \,. \tag{8.58}$$

Similarly, the coefficients c_{ijk}^{uvw} can be expressed through the amplitudes t_i^u, t_{ij}^{uv}, and t_{ijk}^{uvw}, and so on for the higher CI coefficients. Each CI coefficient contains different combinations of the lower-order cluster amplitudes describing the kinematical effects of lower-order excitations and a single genuine cluster amplitude giving the dynamical component of the excitation considered.

Let us consider the example in which the spin orbitals ψ_i and ψ_u are on the same subsystem, whereas spin orbitals ψ_j and ψ_v are on another one and the two subsystems do not interact. Then (8.58) gives $c_{ij}^{uv} = t_i^u t_j^v$ for the kinematically coupled double excitation. (The other terms on the right-hand side of (8.58) vanish as the subsystems are assumed to be noninteracting.) If the numbering of orbitals happens to be such that the two noninteracting subsystems contain spin orbitals ψ_i, ψ_v and ψ_j, ψ_u, respectively, then we get $c_{ij}^{uv} = -t_i^v t_j^u$. (For this numbering of the orbitals the determinant Ψ_{ij}^{vu} containing intrasystem excitations $\psi_i \to \psi_v$, $\psi_j \to \psi_u$ equals the ordered configuration Ψ_{ij}^{uv} with a minus sign.) If all four orbitals are within the interaction range, then both previous terms will appear as kinematical coupling of single excitations. Furthermore, the dynamical effects will cause the coefficient c_{ij}^{uv} to deviate from the kinematical value $t_i^u t_j^v - t_i^v t_j^u$; this is accounted for by the genuine double excitation amplitude t_{ij}^{uv}. (We shall note that in electron correlation problems these genuine double excitations are the most important terms; the role of singly excited configurations is rather small.)

By substituting (8.58) into the expression in (8.14) of Nesbet's theorem, one gets the correlation energy in terms of the cluster amplitudes

$$\begin{aligned} \Delta E = & \sum_{i<j}^{(\text{occ.})} \sum_{u<v}^{(\text{virt.})} (t_i^u t_j^v - t_i^v t_j^u + t_{ij}^{uv}) \\ & \times \left(\langle \psi_i(1)\psi_j(2) | \frac{1}{r_{12}} | \psi_u(1)\psi_v(2) \rangle - \langle \psi_i(1)\psi_j(2) | \frac{1}{r_{12}} | \psi_v(1)\psi_u(2) \rangle \right) \end{aligned} \tag{8.59}$$

Equation (8.59) represents the exact correlation energy for the given basis if the exact cluster amplitudes (i.e., those corresponding to a full CI calculation) are used. In this case the wave function satisfies the (projected) Schrödinger equation in (8.22). It can be reformulated as

$$\langle\Phi|\hat{H}|\Psi\rangle = E\langle\Phi|\Psi\rangle \tag{8.60}$$

for arbitrary N-electron wave functions Φ that can be expressed by using the given one-electron basis. (As Ψ is also sought in that basis, one has $\langle\Phi|\hat{\mathcal{P}} \equiv \langle\Phi|$; $\hat{\mathcal{P}}|\Psi\rangle \equiv |\Psi\rangle$.) One usually speaks of "projecting the Schrödinger equation on different functions Φ."

Equation (8.60) can also be written as

$$\langle\Phi|\hat{H}'|\Psi\rangle = \Delta E\langle\Phi|\Psi\rangle \ , \tag{8.61}$$

where $\hat{H}'$ is the Hamiltonian with an energy shift, defined as

$$\hat{H}' = \hat{H} - E_0 \tag{8.62}$$

with E_0 being the Hartree–Fock energy $E_0 = \langle\Psi_0|\hat{H}|\Psi_0\rangle$.

By choosing $\Phi = \Psi_0$ one recovers (8.59), whereas choosing Φ as the different excited configurations Ψ_i^u, Ψ_{ij}^{uv}, etc., one gets as many equations as the number of ordered configurations (c.f. Section 1) is. That number is equal to the number of the independent CI expansion coefficients c_i^u, c_{ij}^{uv} etc., or that of the independent cluster amplitudes t_i^u, t_{ij}^{uv} etc. The equations obtained in that manner are none other than the respective equations of the full CI problem. Each equation obtained by projecting on Ψ_i^u, Ψ_{ij}^{uv} etc., contains the CI coefficients of all determinants that interact with the given one (i.e., of those, that differ from it by no more than two spin orbitals). As was discussed previously, the CI coefficients are complicated functions of the cluster amplitudes and, therefore, the full CI equations become nonlinear if expressed in terms of the cluster amplitudes.

In practical CCA calculations one cannot use the exact cluster expansion containing excitation operators up to $\hat{T}_N$ for an N-electron system but has to truncate it after the terms $\hat{T}_2$ or $\hat{T}_3$ (in exceptional cases, maybe, $\hat{T}_4$.) Then one needs as many equations as the number of independent amplitudes that are contained in the truncated cluster expansion. For that reason one has to project the Schrödinger equation on all functions Ψ_i^u and Ψ_{ij}^{uv}. (If $\hat{T}_3$ is also retained in the expansion, then projection on the functions Ψ_{ijk}^{uvw} is also necessary.) Projecting on Ψ_0 gives the correlation energy

$$\langle\Psi_0|\hat{H}'|\Psi\rangle = \Delta E \ , \tag{8.63}$$

which is equivalent to (8.59).

The coupled cluster equations obtained in the manner described—we shall not discuss them explicitly—are complicated nonlinear equations containing all the amplitudes retained. The higher-order cluster amplitudes are small numbers; thus omitting them from the equations does not change by very much the values of those that are of interest from the point of view of correlation energy i.e., (8.59). Nevertheless, if one solves the truncated CCA equations, then the resulting correlation energy will not be strictly variational. This causes no serious problems as far as the correlation energy values themselves are concerned, but it makes much more complicated the calculation of the energy gradients and the values of different physical quantities. (Conceptually one could use the exponential Ansatz $\Psi = e^T \Psi_0$ in a variational procedure, but that would be hopelessly involved.)

The CCA scheme, however, has the very important advantage that it is size consistent independently of the level of truncation applied. To see this, let us consider two noninteracting subsystems A and B with the wave functions $\Psi^A = e^{\hat{T}^A}\Psi_0^A$ and $\Psi^B = e^{\hat{T}^B}\Psi_0^B$, respectively. Then we can form the wave function of the supersystem as $\Psi^{AB} = \Psi^A\Psi^B = e^{\hat{T}^A+\hat{T}^B}\Psi_0^A\Psi_0^B = e^{\hat{T}^{AB}}\Psi_0^{AB}$. (We recall again that the antisymmetrization between noninteracting subsystems is immaterial—c.f. Section 4.3 of Chapter 5. By the way, in CCA one usually applies second quantized formalism incorporating antisymmetry in an automatical fashion.)

We assume that Ψ^A and Ψ^B represent the solutions of the CCA equations for the subsystems and we are going to show that the wave function $\Psi^{AB} = \Psi^A\Psi^B$ will satisfy the CCA equations written out for the supersystem at the same level of truncation as used for the subsystems.

The CCA equations (8.61) and (8.63) for the subsystem A can be written as

$$\begin{aligned} \langle\Phi^A|\hat{H}^{A\prime}|\Psi^A\rangle &= \Delta E^A\langle\Phi^A|\Psi^A\rangle\ ; \\ \langle\Psi_0^A|\hat{H}^{A\prime}|\Psi^A\rangle &= \Delta E^A\ , \end{aligned} \tag{8.64}$$

where $\hat{H}^{A\prime}$ is the energy-shifted Hamiltonian of subsystem A, Φ^A is an arbitrary determinant of subsystem A within the excitation range considered and ΔE^A is the corresponding correlation energy. We have an analogous set of equations for subsystem B, as well.

The wave function of the supersystem, which contains the same excitations as Φ^A does, is the wave function $\Phi^A\Psi_0^B$. Thus we have to show that the fulfillment of the equations in (8.64) for the separate subsystems A and B is sufficient to provide the fulfillment of the respective supersystem equations:

$$\langle\Phi^A\Psi_0^B|\hat{H}^{A\prime} + \hat{H}^{B\prime}|\Psi^A\Psi^B\rangle = (\Delta E^A + \Delta E^B)\langle\Phi^A\Psi_0^B|\Psi^A\Psi^B\rangle\ , \tag{8.65}$$

and of those with A and B interchanged. In fact, (8.65) can be rewritten as

$$\begin{aligned}\langle\Phi^A|\hat{H}^{A\prime}|\Psi^A\rangle\langle\Psi_0^B|\Psi^B\rangle + \langle\Psi_0^B|\hat{H}^{B\prime}|\Psi^B\rangle\langle\Phi^A|\Psi^A\rangle \\ = (\Delta E^A + \Delta E^B)\langle\Phi^A|\Psi^A\rangle\langle\Psi_0^B|\Psi^B\rangle\end{aligned} \tag{8.66}$$

Because $\langle\Psi_0^B|\Psi^B\rangle = 1$ according to the intermediate normalization, and $\langle\Psi_0^B|\hat{H}^{B\prime}|\Psi^B\rangle = \Delta E^B$ according to the second equation (8.64) written out for subsystem B, (8.66) reduces to the first equation in (8.64).

Furthermore, (8.63) reads for the supersystem as

$$\langle\Psi_0^{AB}|\hat{H}^{A\prime} + \hat{H}^{B\prime}|\Psi^{AB}\rangle = \Delta E^{AB} . \tag{8.67}$$

Substituting here the definitions $\Psi^{AB} = \Psi^A\Psi^B$, $\Psi_0^{AB} = \Psi_0^A\Psi_0^B$, we get

$$\langle\Psi_0^A|\hat{H}^{A\prime}|\Psi^A\rangle\langle\Psi_0^B|\Psi^B\rangle + \langle\Psi_0^B|\hat{H}^{B\prime}|\Psi^B\rangle\langle\Psi_0^A|\Psi^A\rangle = \Delta E^{AB} . \tag{8.68}$$

Using the second equation in (8.64) written out for both A and B and taking into account the intermediate normalization, we get

$$\Delta E^{AB} = \Delta E^A + \Delta E^B \tag{8.69}$$

as expected.

To complete the proof of size consistency, one has to write out the equivalent of (8.61) for the supersystem with an excited determinant containing one or more intersystem excitations. One can see that it is trivially fulfilled: every integral vanishes on both sides as the number of electrons corresponding to subsystem A (B) is different in the "bra" and in the "ket," whereas the subsystems are assumed noninteracting.

Q.E.D.

We recall that the product of two truncated CI wave functions would satisfy analogous equations; however, it contains higher-order excitations, too, so it does not represent a truncated supermolecule CI wave function *with the same level of truncation* as used for the individual subsystems. Accordingly, a truncated CI wave function is variational but not size consistent. (It is not possible to eliminate the terms containing ΔE^B from (8.66) if the kinematically coupled quadruple excitations are not included in the wave function. Thus one cannot decouple the equations describing the excitations in the noninteracting subsystems.)

Depending on what excitations are included in the wave function, one speaks of "coupled cluster doubles" (CCD), "coupled cluster singles and doubles" (CCSD), or "coupled cluster singles, doubles and triples" (CCSDT) etc. schemes. The simplest CCD is not in practical use, whereas the full account of the triple excitations, the number of which is huge, is often not affordable—but it is not really imperative because of their small importance. For that reason the

amplitudes t_{ijk}^{uvw} are often computed only approximately (e.g., in a perturbational manner). depending on the approximations introduced, one gets somewhat different schemes; the most widely used one is usually denoted CCSD(T), where the parentheses indicate the approximate character of including the triple excitations.

Notes

Section 1.

The considerations presented here are not rigorous from a formal mathematical point of view. According to [1], Fadeev [2] rigorously proved the possibility of expanding the exact solution of the Schrödinger equation in the form of Eq.(8.5) with no more than three electrons, but no strict proof exists for the general N-electron case. (McWeeny [3] calls this possibility a "belief" for which the basis is given by the the considerations described.) This—not too serious in the author's opinion—mathematical short-fall is generally disregarded and all methods of accounting for electron correlation are based on the CI expansion of the wave function, except the "explicitly correlated" methods, in which the interelectronic distance(s) are among the arguments of the wave function. In practice one uses a finite set of one-electron basis functions and in that case the expansion Eq.(8.5) covers, of course, any possible antisymmetric wave functions which can be built up by using a given basis.

Section 2.

The generally accepted definition of the correlation energy has been given by Löwdin [4].

Nesbet's theorem was obtained in [5]; its present formulation was given in [6], where the concept of pair correlation energies has also been formulated.

The classical calculations of the H_2 molecule, mentioned were: Heitler and London [7, 8], MO [9], Weinbaum [10].

Section 3.1.

For a very effective realization of the full CI scheme we refer to [11, 12]. Large scale CI calculations became possible due to Roos's introducing the "direct" method of calculations [13]. Also see his survey article [14].

The most important methods of approximate size consistency correction of truncated CI results are those of Davidson and Pople. The Davidson correction [15] consists in estimating the effect of the omitted higher (triple and, especially, quadruple) excitations in a CISD calculation as $\Delta E = (1 - C_0^2)E_{CISD}^{corr.}$, where $E_{CISD}^{corr.}$ is the correlation energy obtained in a CISD calculation and C_0^2 is the coefficient of the RHF determinant in the CISD wave function. Pople's [16] "quadratic CI" (QCI) may be considered [17, 18] as an (approximate) variant of the "coupled cluster approach" (Section 3.4).

For methods of solving the large CI eigenvalue problems we refer to [19].

Section 3.2.

The MC–SCF method was rediscovered several times (e.g. [20–22]); for the technical aspects see [23].

Some classical references for the methods mentioned are CAS-SCF: [24]; APSG: [25–27]; GVB: [30]; Coulson-Fischer: [31].

Section 3.3.

As noted in the text, the Møllet–Plesset perturbation theory is a particular case of the general Rayleigh-Schrödinger perturbational theory, usually treated by diagrammatic techniques. Some relevant references are: [32–35].

Section 3.4.
It was interesting for the author to observe how in the last decades coupled cluster theory grew from a subject cultivated by a handful of colleagues having enhanced mathematical interest, to a routine method of calculations, first used for producing reference data and then also for many applications. Obviously this happened because CCA is an approach in many respects very close to the variational CI method but has *"built in"* size consistency.

After some pioneering work on the cluster expansion of the wave function and density matrices [36, 37], the CCA theory was adapted from nuclear physics [38, 39] to quantum chemistry by Čižek [40]. Then, as a result of a steady development due mainly to the Florida and Czechoslovak groups (e.g., [41, 42]) and their cooperation [43], (as well as to the development of Pople's "quadratically convergent CI" method [16], which is also, as already noted, a CCA variant) in the last decade the CCA method superseded the so called pair correlation methods [44, 45] like the "coupled electron pair approximation" (CECA) which essentially became obsolete.

The coupled cluster theory usually is considered by heavily using second quantization formalism and diagrammatic techniques not used in the present book. For that reason our discussion concentrated on the concept of the exponential Ansatz and the size consistency property of the CCA method.

Recently the CCA method has been combined (e.g., [46, 47]) with the (approximate) inclusion of the interelectronic distances in the wave function, leading to an extremely promising method.

Suggested readings

The textbooks such as [3,34,48] contain detailed discussions of many problems of electron correlation. There are numerous survey articles on the individual methods, some of them were already quoted. The monograph [49] is also recommended.

References

1 G. Biczó and J. Ladik, Számológép **1**, No. 3, 5 (1971)
2 L. D. Fadeev, Trud. Mat. Inst. Steklov **69**, 1 (1963)
3 R. McWeeny, *Methods of Molecular Quantum Mechanics*, Academic Press, London 1992.
4 P.-O. Löwdin, Adv. Chem. Phys. **2**, 207 (1959)
5 R.K. Nesbet, Phys. Rev. **109**, 1632 (1958)
6 R.K. Nesbet, Adv. Chem. Phys. **9**, 321 (1965)
7 W. Heitler and F. London, Z. Physik **44**, 455 (1927)
8 Y. Sugiura, Z. Physik **45**, 484 (1927)
9 R. Mulliken, Phys. Rev. **32** 186, 761 (1928), **33**, 730 (1929)
10 S. Weinbaum, J. Chem. Phys. **1**, 593 (1933)
11 P.J. Knowles and N. C. Handy, Chem. Phys. Letters. **111**, 315 (1984)
12 P.J. Knowles and N. C. Handy, Computer Physics Communications **54**, 75 (1989)
13 B.O. Roos, Chem. Phys. Letters, **15**, 153 (1972)
14 B. Roos, *The Configuration Interaction Method* in *Computational Techniques in Quantum Chemistry and Molecular Physics* (ed. G.H.F. Dierksen, B.T. Suttcliffe and A. Veillard) Reidel, Dordrecht 1974.
15 S. R. Langhoff and E.R. Davidson, Int. J. Quantum Chem. **8**, 61 (1974)
16 J.A. Pople, M. Head-Gordon and K. Raghavachari, J. Chem. Phys. **87**, 5968 (1987)
17 J. Paldus, J. Čižek and B. Jeziorski, J. Chem. Phys. **90**, 4356 (1989)
18 J. Noga, unpublished work; also see J. Noga, Acta Univ. Debreceniensis Ser. Physica et Chimica XXX/2, 67 (1995)

19 E. R. Davidson, *Matrix Eigenvector Methods* in *Methods in Computational Molecular Physics* (ed. G.H.F. Dierksen and S. Wilson) Reidel, Dordrecht 1983.
20 R. McWeeny, Proc. Roy. Soc (London) **A232**, 114 (1955)
21 T.L. Gilbert, J. Chem. Phys. **43**, S245, (1965)
22 E. Clementi, J. Chem. Phys. **46**, 3842 (1967)
23 B. O. Roos, *The Multiconfigurational (MC) SCF Method* in *Methods in Computational Molecular Physics* (ed. G.H.F. Dierksen and S. Wilson) Reidel, Dordrecht 1983.
24 B.O. Roos, P. R. Taylor, Chem. Phys. **48** 157 (1980)
25 A.C. Hurley, J. Lenard-Jones,and J.A. Pople, Proc Roy. Soc. (London) **A220**, 446 (1953)
26 E. Kapuy, Acta Phys. Hung. **11**, 409 (1960); **12**, 185 (1960)
27 R. Ahlrichs and W. Kutzelnigg J. Chem. Phys. **48**, 1819 (1967)
28 W. Kutzelnigg, Theor. Chim. Acta **1**, 327 (1963)
29 W. Kutzelnigg, J. Chem. Phys. **40**, 3640 (1964)
30 W.A. Goddard III, T.H. Dunning Jr., W. J. Hunt and P.J. Hay, Acc. Chem. res. **6**, 368 (1973)
31 C.A. Coulson and I. Fischer, Philos. Mag. [7] **40**, 386 (1949)
32 C. Møller and M.S. Plesset, Phys. Rev. **46**, 618 (1934)
33 J. Paldus and J. Čižek, Adv. Quantum Chem. **9**, 10 (1975)
34 P. Jørgensen and J. Simons, *Second Quantization-Based Methods in Quantum Chemistry*, Academic Press, New York 1981.
35 S. Wilson *Diagrammatic Many-Body Perturbation Theory* in *Methods in Computational Molecular Physics* (ed. G.H.F. Dierksen and S. Wilson) Reidel, Dordrecht 1983.
36 O. Sinanoğlu, Adv. Chem. Phys. **6**, 315 (1964)
37 H. Primas, in *Modern Quantum Chemistry (Istanbul Lectures)*, ed. O. Sinanoğlu, Academic Press, New York 1965.
38 F. Coester, Nucl. Phys. **7**, 421 (1958)
39 F. Coester and H. Kümmel, Nucl. Phys. **17**, 477 (1960)
40 J. Čižek, Adv. Chem. Phys. **14**, 35 (1969)
41 M. Urban, I. Černušák, V. Kellö, J. Noga, *Electron correlation in molecules*, in *Computational Methods in Chemistry,* ed. S. Wilson, Plenum Press, New York 1987
42 R.J. Bartlett, *Coupled-Cluster Theory: an Overview and Recent Developments* in *Modern Electronic Structure Theory* ed. D.R. Yarkony, World Scientific, Singapore 1995.
43 R.J. Bartlett, S.A. Kucharski, J. Noga, J.D. Watts, G. W. Trucks, *Some Consideration of Alternative Ansätz in Coupled-Cluster Theory,* in *Lecture Notes in Chemistry,* Vol.**52** Springer Verlag, Heidelberg 1989
44 M.A. Robb, *Pair Functions and Diagrammatic Perturbation theory* in *Computational Techniques in Quantum Chemistry and Molecular Physics*, (ed. G.H.F. Dierksen, B.T. Suttcliffe and A. Veillard) Reidel, Dordrecht 1974.
45 R. Ahlrichs, *Pair Correlation theories* in *Methods in Computational Molecular Physics* (ed. G.H.F. Dierksen and S. Wilson) Reidel, Dordrecht 1983.
46 J. Noga, W. Klopper, W. Kutzelnigg *CC-R12: an explicitly correlated coupled-cluster theory.* in *Recent Advances in Computational Chemistry - Vol 3.: Recent Advances in Coupled-Cluster Methods,* ed. R. J. Bartlett, World Scientific, Singapore 1997
47 J. Noga and P. Valiron, *Explicitly correlated coupled cluster R12 calculations.* in *Computational Chemistry: Reviews of Current Trends. Vol. 7* ed. J. Leszcynski, World Scientific, Singapore 2002
48 A. Szabo and N.S. Ostlund *Modern Quantum Chemistry*, McGraw Hill, New York 1989.
49 A.C. Hurley, *Electron Correlation in Small Molecules* Academic Press, New York 1976.

Chapter 9

MISCELLANEOUS

In this chapter we consider some miscellaneous results that are often referred to but are not so easy to find in the literature and could not be integrated in the framework of the previous chapters.

1. The 1/Z expansion

We are going to discuss briefly here a scheme permitting one to treat different atoms (ions) on a common energy scale, which may be useful to get a better qualitative understanding of the periodic table of elements. There is some formal similarity between this derivation and the scaling procedure discussed in Section 3.3 of Chapter 3, but here we do not consider a transformation of the wave function; the Hamiltonian is rewritten in terms of some new coordinates.

The Born–Oppenheimer Hamiltonian for an N-electron atom (ion) with the nuclear charge Z is, in atomic units,

$$\hat{H} = \sum_{i=1}^{N} \left(-\frac{1}{2}\Delta_i \right) - \sum_{i=1}^{N} \frac{Z}{r_i} + \sum_{i<j} \frac{1}{r_{ij}} \,. \tag{9.1}$$

We introduce the new coordinates $\vec{\varrho}_i$ as

$$\vec{\varrho}_i = Z\vec{r}_i \,, \tag{9.2}$$

which involves $\vec{r}_i = \frac{1}{Z}\vec{\varrho}_i$; $\frac{1}{r_i} = \frac{Z}{\varrho_i}$; $\frac{1}{r_{ij}} = \frac{Z}{\varrho_{ij}}$. Furthermore, one has, according to the chain rule

$$\frac{\partial}{\partial x_i} = \frac{\partial}{\partial \varrho_{x_i}} \frac{\partial \varrho_{x_i}}{\partial x_i} = Z \frac{\partial}{\partial \varrho_{x_i}} \,; \quad \frac{\partial^2}{\partial x_i^2} = Z^2 \frac{\partial^2}{\partial \varrho_{x_i}^2} \tag{9.3}$$

I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*
© Springer Science+Business Media New York 2003

i.e.,

$$\Delta_i = Z^2 \Delta_{\varrho_i} \,. \tag{9.4}$$

Thus the Hamiltonian can be written in the new coordinates as

$$\begin{aligned} \hat{H} &= \sum_{i=1}^{N} \left(-\frac{1}{2} Z^2 \Delta_{\varrho_i} \right) - \sum_{i=1}^{N} \frac{Z^2}{\varrho_i} + \sum_{i<j} \frac{Z}{\varrho_{ij}} \\ &= Z^2 \left[\sum_{i=1}^{N} \left(-\frac{1}{2} \Delta_{\varrho_i} \right) - \sum_{i=1}^{N} \frac{1}{\varrho_i} + \frac{1}{Z} \sum_{i<j} \frac{1}{\varrho_{ij}} \right] \\ &= Z^2 \left[\sum_{i=1}^{N} \hat{h}_H(i_\varrho) + \frac{1}{Z} \sum_{i<j} \frac{1}{\varrho_{ij}} \right] , \end{aligned} \tag{9.5}$$

where $\hat{h}_H$ is the Hamiltonian of a free hydrogen atom, i_ϱ stands for all the coordinates of the ith electron in the transformed system of coordinates, and the factor Z^2 indicates a change in the energy scale.

The first sum in (9.5) is that of N formally noninteracting hydrogen atoms, and the electron–electron interaction may be viewed as a perturbation, multiplied by a factor $\frac{1}{Z}$, decreasing with the increase of the nuclear charge Z.

Equation (9.5) expresses the obvious fact that near the nucleus the relative importance of the electron–electron repulsion decreases with the increase of the nuclear charge Z as compared with the electron–nuclear attraction. This should be particularly the case for the internal shells that are not influenced too much by the outer electrons. Thus the Hamiltonian in (9.5) predicts that the energy of the characteristic X-ray quanta (the inner-shell ionization potentials) of the different elements should be proportional to Z^2 to good accuracy, which one can indeed observe as a linear relationship between $\sqrt{\frac{1}{\lambda}}$ and Z, where λ is the wavelength of the X-ray line (Moseley's law). This was the first experimental proof in 1913 for the proposition of van den Broek that the atomic numbers of the elements in the periodic table are equal to their nuclear charge (and the number of electrons in the neutral atom).

In the particular case of the He-like ions (ions with two electrons) the accurate ionization potential for different values of Z can be well approximated by a series

$$I(Z) = Z^2 \left(\frac{1}{2} - \frac{5}{8} \cdot \frac{1}{Z} + 0.1576556 \frac{1}{Z^2} + 0.008535 \frac{1}{Z^3} + 0.00034 \frac{1}{Z^4} + \ldots \right) . \tag{9.6}$$

Here the first term comes from the energy of the respective noninteracting H-atom scaled with the factor Z^2. The second term is the first-order energy

correction that equals (c.f. Chapter 4) the expectation value of the perturbation operator over the unperturbed wave function. [The value $\frac{5}{8}$ equals the integral $\iint |\varphi_{1s}(\vec{r}_1)|^2 \frac{1}{r_{12}} |\varphi_{1s}(\vec{r}_2)|^2 dv_1 dv_2$ for the $1s$ hydrogenic orbital $\varphi_{1s} = \sqrt{\frac{1}{\pi}} e^{-r}$; the factor $\frac{1}{Z}$ comes from the coefficient $\frac{1}{Z}$ at the last term on the right-hand side of (9.5).]

2. The cusp condition

2.1. The nuclear cusp

We have seen in the previous section that the internal electron shells of an atom are mostly determined by the nuclear charge Z. This is even more pronounced if one considers the behavior of the electronic wave function in the *immediate vicinity* of the nucleus, i.e., in the limiting case when one of the electron–nuclear distances tend to 0.

In the limit when an electron approaches the nucleus very closely ($r \to 0$), one can disregard all the other electrons and nuclei of the system and consider the motion of the given electron only in the field of the nucleus. That is, very close to the nucleus one has essentially to consider the problem of the hydrogenlike ion, but only in the interval $r \ll 1$. The respective Schrödinger equation then reads

$$-\frac{1}{2}\Delta\psi - \frac{Z}{r}\psi = \varepsilon\psi \ . \tag{9.7}$$

It is known from mathematics that the Laplacian $\Delta \equiv \nabla^2$ can be written by using spherical coordinates r, ϑ, φ as

$$\Delta = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2}\left[\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial}{\partial\vartheta}\right) + \frac{1}{\sin^2\vartheta}\frac{\partial^2}{\partial\vartheta^2}\right] \ . \tag{9.8}$$

Furthermore, it is known from basic quantum mechanics that for the angular part of this operator one has

$$\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\left(\sin\vartheta\frac{\partial}{\partial\vartheta}\right) + \frac{1}{\sin^2\vartheta}\frac{\partial^2}{\partial\vartheta^2} = -\hat{l}^2 \ , \tag{9.9}$$

i.e., it equals, up to a minus sign, the operator of the square of the angular momentum. Thus we may write (9.7) as

$$-\frac{1}{2}\left[\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right) - \frac{\hat{l}^2}{r^2}\psi\right] - \frac{Z}{r}\psi = E\psi \ . \tag{9.10}$$

The wave function can be written in the form $\psi(r, \vartheta, \varphi) = R(r)Y_l^m(\vartheta, \varphi)$, where $R(r)$ is the radial part and $Y_l^m(\vartheta, \varphi)$ is a spherical harmonics. $Y_l^m(\vartheta, \varphi)$

is an eigenfunction of operator $\hat{l}^2$ with an eigenvalue $l(l+1)$: $\hat{l}^2 Y_l^m(\vartheta,\varphi) = l(l+1)Y_l^m(\vartheta,\varphi)$. If one is interested in an s-state that is spherically symmetric, then the angular derivatives vanish; this is in accord with the fact that $l = 0$ in that case. Strictly speaking, if a many-electron system is considered, the state ψ of the electron approaching the nucleus need not be a pure angular momentum eigenfunction. However, it can always be expanded as a linear combination of such ones, and the considerations that follow will be valid for each component separately. (Obviously, the $1s$ states are of the highest interest as far as the vicinity of the nucleus is concerned.)

Thus one obtains for the radial function $R = R(r)$:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) - \frac{l(l+1)}{r^2}R + \frac{2Z}{r}R = -2ER\,. \tag{9.11}$$

Obviously, the term on the right-hand side can be neglected in the limit of $r \to 0$, as it contains a finite coefficient, whereas the coefficients of the other terms diverge.

Let us consider the leading term of $R(r)$ to be proportional to r^q $(q > 0)$ and take into account that in the limit $r \to 0$ any term proportional to r^p with $p > q$ is negligible as compared with r^q. Then $\frac{dR}{dr} \sim qr^{q-1}$; $r^2\frac{dR}{dr} \sim qr^{q+1}$; and $\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) \sim q(q+1)r^{q-2}$. The second term on the left-hand side varies with the same power r^{q-2}, whereas the third term behaves as r^{q-1}, i.e., it can be neglected as compared with the first two terms when $r \to 0$. Hence one gets $q = l$.

This result shows that for any nonzero values of l one has $\lim_{r\to 0} R(r) = 0$, whereas for $l = 0$ (s-states) $\lim_{r\to 0} R(r) \neq 0$ and there is a nonzero probability of finding the electron on the nucleus.

In order to get a more detailed insight into the behavior near the nucleus, let us consider one more term of the series expansion of $R(r)$ and write

$$R(r) \cong Cr^l(1+\alpha r) = C(r^l + \alpha r^{l+1})\,, \tag{9.12}$$

where $C =$ const. Then we have

$$\frac{dR}{dr} \cong C[lr^{l-1} + \alpha(l+1)r^l] \tag{9.13}$$

and

$$\frac{d}{dr}\left(r^2\frac{dR}{dr}\right) \cong C[l(l+1)r^l + \alpha(l+1)(l+2)r^{l+1}]\,. \tag{9.14}$$

Substituting (9.14) into (9.11) and neglecting again the term on the right-hand side, we get

$$\begin{aligned} C[l(l+1)r^{l-2} + \alpha(l+1)(l+2)r^{l-1}] - l(l+1)C(r^{l-2} + \alpha r^{l-1}) \\ +2ZC(r^{l-1} + \alpha r^l) = 0\,. \end{aligned} \tag{9.15}$$

The terms with r^{l-2} cancel, and the term with r^l can be neglected as compared to the terms $\sim r^{l-1}$ when $r \to 0$. Thus we get

$$C\alpha(l+1)(l+2) - \alpha Cl(l+1) + 2ZC = 0 \,, \tag{9.16}$$

i.e.,

$$\alpha = -\frac{Z}{l+1} \,. \tag{9.17}$$

Substituting this result into (9.12) and (9.13), we have

$$\begin{aligned} R(r) &\cong C\left(r^l - \frac{Z}{l+1} r^{l+1}\right) \,; \\ \frac{dR(r)}{dr} &\cong C(lr^{l-1} - Zr^l) \,. \end{aligned} \tag{9.18}$$

This indicates that only s-states ($l = 0$) have $\lim_{r\to 0} R(r) \neq 0$ and that they satisfy the "cusp condition"

$$\lim_{r\to 0} \frac{\frac{dR}{dr}}{R(r)} = -Z \tag{9.19}$$

or

$$\lim_{r\to 0} \frac{dR}{dr} = -ZR(0) \,. \tag{9.20}$$

If no spherical symmetry around the nucleus can be assumed, then one has to consider the general expansion of the form

$$\psi(r, \vartheta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} C_{lm} R_l(r) Y_l^m(\vartheta, \varphi) \,. \tag{9.21}$$

In this case, according to (9.18) only the component with $l = 0$ contributes to the value of ψ at the nucleus and only components with $l = 0$ and 1 have nonzero derivatives there. The state $l = 1$ is odd, i.e., the corresponding component of the function changes its sign when the nucleus is approached from two opposite directions $\vec{r}$ and $-\vec{r}$. (The angular coordinates are ϑ and φ in one case and $\pi - \vartheta$ and $\varphi + \pi$ in the other.) This means that the inversion leads to $\psi \to -\psi$ if an $l = 1$ state is considered, which involves also $\frac{d\psi}{dr} \to -\frac{d\psi}{dr}$ for that state. Therefore, if one averages $\frac{d\psi}{dr}$ over all the angles, then only the contribution from the s-state survives. Consequently, the cusp condition can be formulated for the general expansion (9.21) as

$$\lim_{r\to 0} \langle \frac{d\psi}{dr} \rangle_{\text{aver.}} = -Z\psi(0) \,. \tag{9.22}$$

2.2. The electronic cusp

When two electrons approach each other very closely, then one can again neglect all the other particles in the system in order to study the limiting behavior of the given pair of electrons. The equations that one has to consider in that case are analogous to those for the nuclear cusp, but there are important differences, as well.

First, the nucleus is considered fixed in the Born–Oppenheimer framework; thus the mass entering the kinetic energy part of the Hamiltonian $\frac{-\hbar^2}{2m}\Delta$ is the electronic mass m_e. (Note that equation (9.7) is written out in atomic units in which $\hbar = m_e = 1$.) When the relative motion of two electrons is considered, we have to use the reduced mass (c.f. Section 1.1 of Chapter 1) $\mu = \frac{1}{2}m_e$. Therefore the factor $\frac{1}{2}$ in the term containing the Laplacian cancels. Furthermore, two electrons exhibit repulsion and not attraction, which means that one has to replace the nuclear charge $+Z$ by the electronic charge -1. Accordingly, the Schrödinger equation for the radial function that is to be considered if two electrons are very close to each other becomes

$$\frac{1}{r_{12}^2}\frac{d}{dr_{12}}\left(r_{12}^2\frac{dR}{dr_{12}}\right) - \frac{l(l+1)}{r_{12}^2}R - \frac{1}{r_{12}}R = -ER\,. \tag{9.23}$$

All the considerations given for the nuclear cusp can be applied here, as well, provided that one replaces Z with $-\frac{1}{2}$. Thus instead of (9.22) we obtain for the electron–electron cusp the condition

$$\lim_{r_{12}\to 0}\langle\frac{d\psi}{dr_{12}}\rangle_{\text{aver.}} = \frac{1}{2}\psi\Big|_{r_{12}=0}\,. \tag{9.24}$$

This formula is valid both for pairs of electrons having the same spin and for those with different spins. However, if the two electrons have the same spin, then the wave function should be antisymmetric, which means that only odd values of l are admitted. In this case the condition in (9.24) is not really informative as both sides vanish trivially. For electrons of opposite spin the cusp condition in (9.24) gives the important result that if they approach each other very closely, then the wave function, in average, behaves as $\sim C(1+\frac{1}{2}r_{12}+\ldots)$. This may be utilized for setting up explicitly correlated approximate wave functions.

3. The asymptotical behavior of the wave function at very large distances

If one of the electrons is very far from the rest of the system, then one has again to deal with an effective one-electron problem described by (9.10) but must replace in it $-\frac{Z}{r}$ with the effective potential $U(r)$. In this case one

essentially considers the limiting behavior of the Hartree-Fock equation for the ith orbital. For that reason we shall write ψ_i and ε_i instead of ψ and E. Expanding the derivative of the product

$$\frac{d}{dr}\left(r^2\frac{d\psi_i}{dr}\right) = 2r\frac{d\psi_i}{dr} + r^2\frac{d^2\psi_i}{dr^2} \tag{9.25}$$

we get

$$\frac{2}{r}\frac{d\psi_i}{dr} + \frac{d^2\psi_i}{dr^2} - \frac{\hat{l}^2}{r^2}\psi_i - 2U(r)\psi_i = -2\varepsilon_i\psi_i \ . \tag{9.26}$$

As we consider the case $r \to \infty$, any term $\sim\frac{1}{r}$ or $\frac{1}{r^2}$ becomes negligible. Far from the system the overall behavior of the potential $U(r)$ is also Coulombic $\left(\sim\frac{1}{r}\right)$. Thus we are left with

$$\frac{d^2\psi_i}{dr^2} = -2\varepsilon_i\psi_i \ . \tag{9.27}$$

As it is known from mathematics, the general solution of the differential equation in (9.27) is $\psi_i = Ae^{\zeta r} + Be^{-\zeta r}$ with $\zeta = \sqrt{-2\varepsilon_i}$. (Note that $\varepsilon_i < 0$.) However, the first term diverges when $r \to \infty$, and it should be omitted. Thus we obtain the asymptotic expression

$$\psi \sim Be^{-\sqrt{-2\varepsilon_i}r} \ . \tag{9.28}$$

If one is interested in how the overall wave function of a molecule behaves at very large distances, then one has to consider that orbital that decays most slowly. Then, taking into account Koopmans theorem (Section 3 of Chapter 6) connecting the orbital energies and the ionization potentials, one may write for the most slowly decaying component of the wave function the asymptotic expression

$$\psi \sim Be^{-\sqrt{2I_{\min}}\,r} \ , \tag{9.29}$$

where $I_{\min} = \min|\varepsilon_i|$ is the lowest ionization potential of the molecule (atom).

4. Basis functions: The Gaussian product theorem

We have seen in the previous section that the adequate basis orbitals for electronic structure calculation would decay exponentially at large distances. The orbitals of this type are usually written in the form

$$\chi_{nlm}(\vec{r}_a) = \mathcal{N}r_a^{n-1}e^{-\zeta r_a}Y_l^m(\vartheta_a, \varphi_a) \ , \tag{9.30}$$

where $\vec{r}_a = \vec{r} - \vec{R}_A$ is the radius vector of the electron with respect to the center (atom) A, r_a, ϑ_a, and φ_a are the spherical coordinates referred to the center A, $Y_l^m(\vartheta_a, \varphi_a)$ is a spherical harmonics, and $\mathcal{N}$ is a normalization coefficient.

Orbitals of type (9.30) are called Slater-type orbitals (STOs). The exact orbitals of the hydrogen atom can also be presented as linear combinations of a few STOs (of same exponent), so STOs can be considered as an adequate generalization (and at the same time a simplification) of the hydrogenic orbitals for the use in the many-electron problems. Accordingly, the use of STOs was the standard practice in the early period of quantum chemistry.

The use of STOs permitted one to perform excellent atomic Hartree–Fock calculations and some quite useful diatomic calculations as well. One, however, encountered the "bottleneck" that no treatable expressions could be derived for the three- and four-center integrals that are needed to calculate general multi-atom molecules. Although the research aiming to develop practicable methods of calculating the integrals over the STOs is continuing, the absolute majority of actual quantum chemical calculations use "Gaussian-type orbitals" (GTOs). Originally Boys (1950) proposed to calculate numerically the integrals over the STOs by approximating them as linear combinations of Gaussian functions, i.e., such in that the radial dependence is $\sim e^{-\alpha r^2}$ instead of $e^{-\zeta r}$. Essentially this idea is still utilized when one applies the simplest "STO–nG" basis sets in which a Slater-type orbital is approximated as a sum of n Gaussians. (n is usually between 3 and 6). However, in most cases one uses the Gaussian functions themselves as the basis orbitals; a basis function can either be a single Gaussian or a linear combination ("contraction") of several Gaussians with fixed coefficients.

Owing to their functional form, the Gaussian functions are not well suited to the problem; their advantage is merely in the relative simplicity of calculating all the necessary one- and two-electrons integrals. The possibility of calculating the integrals over the GTOs is provided by the so-called "Gaussian product theorem," permitting a product of two Gaussians centered on different points of space to be presented as a third Gaussian that is centered on a point that lies on the straight line connecting the original centers. Thus the three- and four-center integrals are reduced to two-center ones that can be calculated relative easily.

One usually writes the primitive Gaussians in the "Cartesian" form

$$\chi_{uvw}(\vec{r}_a) = \mathcal{N} x_a^u y_a^v z_a^w e^{-\alpha r_a^2} , \tag{9.31}$$

where u, v, and w are nonnegative integers, and

$$\vec{r}_a = \vec{i}x_a + \vec{j}y_a + \vec{k}z_a = \vec{i}(x - X_A) + \vec{j}(y - Y_A) + \vec{k}(z - Z_A) \tag{9.32}$$

is the radius vector with respect to the center A, X_A, Y_A, and Z_A being the coordinates of center A. (The "pure" d, f, etc., functions, the angular behavior

of which is described by a spherical harmonics $Y_l^m(\vartheta_a, \varphi_a)$, can be expressed as linear combinations of Cartesian terms.)

We are interested in the product

$$\begin{aligned} e^{-\alpha r_a^2} e^{-\beta r_b^2} &= \exp\Big\{-\alpha[(x-X_A)^2 + (y-Y_A)^2 + (z-Z_A)^2] \\ &\quad -\beta[(x-X_B)^2 + (y-Y_B)^2 + (z-Z_B)^2]\Big\} \ . \end{aligned} \tag{9.33}$$

This can be written as a product of three similar factors. One of them is

$$\begin{aligned} &\exp\Big\{-\alpha(x-X_A)^2 - \beta(x-X_B)^2\Big\} \\ &= \exp\Big\{-\alpha(x^2 - 2xX_A + X_A^2) - \beta(x - 2xX_B + X_B^2)\Big\} \\ &= \exp\Big\{-(\alpha+\beta)x^2 + 2x(\alpha X_A + \beta X_B) - (\alpha X_A^2 + \beta X_B^2)\Big\} \\ &= \exp\left\{-(\alpha+\beta)\left[x^2 - 2x\frac{\alpha X_A + \beta X_B}{\alpha+\beta} + \frac{\alpha X_A^2 + \beta X_B^2}{\alpha+\beta}\right]\right\} \qquad (9.34) \\ &= \exp\left\{-(\alpha+\beta)\left[\left(x - \frac{\alpha X_A + \beta X_B}{\alpha+\beta}\right)^2 - \left(\frac{\alpha X_A + \beta X_B}{\alpha+\beta}\right)^2 \right.\right. \\ &\qquad\qquad \left.\left. + \frac{\alpha X_A^2 + \beta X_B^2}{\alpha+\beta}\right]\right\} \\ &= \exp\left\{-(\alpha+\beta)\left[\left(x - \frac{\alpha X_A + \beta X_B}{\alpha+\beta}\right)^2 - \frac{\alpha\beta}{(\alpha+\beta)^2}(X_A - X_B)^2\right]\right\} \\ &= \text{const} \times \exp[-(\alpha+\beta)(x-X_0)^2] \ , \end{aligned}$$

where

$$X_0 = \frac{\alpha X_A + \beta X_B}{\alpha+\beta} \ . \tag{9.35}$$

Therefore

$$e^{-\alpha r_a^2} e^{-\beta r_b^2} = \text{const} \times e^{-(\alpha+\beta)(\vec{r}-\vec{R}_0)^2} \ , \tag{9.36}$$

where

$$\vec{R}_0 = \frac{\alpha \vec{R}_A + \beta \vec{R}_B}{\alpha+\beta} \ . \tag{9.37}$$

Q.E.D.

5. The integral transformation problem

The correlation energy in almost all cases is described by expressions containing integrals over the molecular orbitals—see, for instance (8.14) describing Nesbet's theorem or the UMP2 and RMP2 correlation energy expressions (8.35) and (8.36), respectively. The MOs, in turn, are usually expressed as linear combinations of some basis (atomic) orbitals (c.f. Section 5 of Chapter 6). Therefore, the integrals over the MOs represent some sums of those for the basis functions. If

$$\varphi_i = \sum_{\mu=1}^{m} c_\mu^i \chi_\mu \tag{9.38}$$

then, as it is easy to see, one has

$$\langle \varphi_i | \hat{h} | \varphi_j \rangle = \sum_{\mu,\nu=1}^{m} c_\mu^i c_\nu^j \langle \chi_\mu | \hat{h} | \chi_\nu \rangle \tag{9.39}$$

and

$$[\varphi_i \varphi_j | \varphi_k \varphi_l] = \sum_{\mu,\nu,\varrho,\tau=1}^{m} c_\mu^i c_\nu^j c_\varrho^k c_\tau^l [\chi_\mu \chi_\nu | \chi_\varrho \chi_\tau] \,. \tag{9.40}$$

(We have assumed for simplicity the orbitals to be real—as it is usually the case.)

The calculation of the one-electron integrals in (9.39)—if required at all—does not cause any problems. It correspond to turning from one finite basis representation of the one-electron Hamiltonian $\hat{h}$ to another one remaining within the same m-dimensional one-electron subspace. Thus, by denoting the integrals $\langle \varphi_i | \hat{h} | \varphi_j \rangle = h_{ij}^{MO}$ and $\langle \chi_\mu | \hat{h} | \chi_\nu \rangle = h_{\mu\nu}^{AO}$ and considering them as elements of the $m \times m$ matrices $\mathbf{h}^{MO}$ and $\mathbf{h}^{AO}$, respectively, and denoting by $\mathbf{C}$ the matrix formed of the LCAO coefficients, one has

$$\begin{aligned} h_{ij}^{MO} &= \sum_{\mu,\nu=1}^{m} c_\mu^i h_{\mu\nu} c_\nu^j = \sum_{\mu,\nu=1}^{m} C_{\mu i} h_{\mu\nu}^{AO} C_{\nu j} \\ &= \sum_{\mu,\nu=1}^{m} (\mathbf{C}^\dagger)_{i\mu} h_{\mu\nu}^{AO} C_{\nu j} = (\mathbf{C}^\dagger \mathbf{h}^{AO} \mathbf{C})_{ij} \end{aligned} \tag{9.41}$$

i.e., $\mathbf{h}^{MO} = \mathbf{C}^\dagger \mathbf{h}^{AO} \mathbf{C}$. This means that $\mathbf{h}^{MO}$ can be computed by two consecutive matrix multiplications, which requires computational effort scaling only with $\sim m^3$.

The situation is much more difficult with the two-electron integrals in (9.40). The number of two-electron integrals over the MOs is proportional to the fourth

power of the dimension of the system—one has to calculate $\sim n_{\text{occ.}}^2 \times n_{\text{virt}}^2$ different integrals if MP2 is used, and $\sim m^4$ for applying the higher-order theories requiring all the possible integrals over the MOs. (Here $n_{\text{occ.}}$ and $n_{\text{virt.}}$ are the number of occupied and virtual orbitals, respectively; $n_{\text{occ.}} + n_{\text{virt.}} = m$. As for any reasonably good basis set $n_{\text{occ.}} \ll m$, this comparison explains why MP2 is much cheaper than any other method of calculating electron correlation.) Every integral on the right-hand side of (9.40) requires $\sim m^4$ arithmetical operations, thus a "brute force" calculation of all integrals would scale as the eighth power of the dimensions of the system, which would be impossible in every case of practical interest. However, this difficulty can be circumvented by performing the transformation in a stepwise manner. (This is quite analogous to reducing the transformations in (9.39) that would scale as $\sim m^4$ to two matrix multiplications scaling as $\sim m^3$.)

First one has to calculate the intermediate quantities

$$[\chi_\mu \chi_\nu | \chi_\varrho \varphi_l] = \sum_{\tau=1}^{m} c_\tau^l [\chi_\mu \chi_\nu | \chi_\varrho \chi_\tau] \,, \tag{9.42}$$

then the quantities

$$[\chi_\mu \chi_\nu | \varphi_k \varphi_l] = \sum_{\varrho=1}^{m} c_\varrho^k [\chi_\mu \chi_\nu | \chi_\varrho \varphi_l] \,, \tag{9.43}$$

and so on. Each such step requires "only" $\sim m^5$ operations if a full integral transformation is done. This small trick is of central importance as far as the possibility of electron correlation calculations for systems of chemical interest is concerned. (Of course, further simplifications can be introduced in practical calculations, but they do not influence the conceptual aspects of the problem.)

Notes

Section 1.

The 1/Z expansion goes back to Hylleraas [1,2] (also see e.g., [3,4]). The references to Moseley and van den Broek are [5,6]

Section 2.

The behavior $R(r) \sim r^l$ near the nucleus is demonstrated in textbooks of quantum mechanics (e.g., [7]). The cusp conditions Eq. (9.22), (9.24) are due to Kato [8].

Section 3.

The derivation represents a slight generalization of that in [7].

Section 4.

Slater originally introduced [9] his orbitals for an approximate description of free atoms; he also gave an empirical rule for determining the exponents and for estimating the atomic total energies as sums of one-electron contributions—c.f. Section 3.1 of Chapter 6. (These estimations are

unexpectedly good.) For near Hartree-Fock calculations of free atoms and ions by using STOs we refer to [10]. Slater's book [11] discusses some classical diatomic calculations also performed by using STOs.

The GTOs were introduced in quantum chemistry by Boys [12], to whom the "Gaussian product theorem" also belongs. The use of Gaussians as independent basis functions—and not only means for estimating integrals over the STOs—was probably first proposed by Shavitt [13].

Section 5.

The methods of integral transformation are discussed in detail e.g., in [14].

Suggested readings

A compilation of classical formulae for calculating integrals over the STOs and GTOs is given in [15]. For detailed discussion of the mathematical methods we refer to [16–18]. The paper [17] contains also a discussion of the Rys-polynomial method introduced in [19] and now generally used for calculating two-electron integrals for Gaussian basis sets.

References

1 E.A. Hylleraas, Z. Phys. **65**, 209 (1930)

2 E.A. Hylleraas, Phys. Rev. **103**, 829 (1956)

3 P.-O. Löwdin, J. Mol. Spectry, **3**, 46 (1959)

4 J.N. Silverman, Phys. Rev. **A23**, 441 (1981)

5 H. Moseley, Phil. Mag. [6] **26**, 1024 (1013); **27**, 703 (1914)

6 A. van den Broek, Nature **92**, 372 (1913)

7 L.D. Landau and I. M. Lifshitz, *Kvantovaya Mekhanika*, Nauka, Moscow 1974.

8 T. Kato, Commun. Pure Appl. Math. **10**, 15 (1957)

9 J.C. Slater, Phys. Rev. **36**, 57 (1930)

10 E. Clementi and C. Roetti, *Roothaan-Hartree-Fock Atomic Wavefunctions*, Academic Press, New York 1974. (Atomic Data and Nuclear Data Tables **14**, 177, 1974)

11 J.C. Slater, *Quantum Theory of Molecules and Solids vol. 1. Electronic Structure of Molecules*, McGraw-Hill, New York 1963.

12 S.F. Boys, Proc. Roy. Soc. (London) **A200**, 542 (1950)

13 I. Shavitt, in *Methods in Computational Physics,* Academic Press, New York 1963.

14 B. Roos, *The Configuration Interaction Method* in *Computational Techniques in Quantum Chemistry and Molecular Physics* (ed. G.H.F. Dierksen, B.T. Suttcliffe and A. Veillard) Reidel, Dordrecht 1974.

15 D.A. Zhogolev and V.B. Volkov, *Metody, Algoritmy i Programmi dlya Kvantovo-Khimicheskikh Raschetov Molekul*, Kiev, Naukova Dumka 1976.

16 V.R. Saunders *An Introduction to Molecular Integral Evaluation*, in *Computational Techniques in Quantum Chemistry and Molecular Physics* (ed. G.H.F. Dierksen, B.T. Suttcliffe and A. Veillard) Reidel, Dordrecht 1974.

17 V. R. Saunders in *Methods in Computational Molecular Physics* (ed. G.H.F. Dierksen and S. Wilson) Reidel, Dordrecht 1983.

18 R.O. Steinborn in *Methods in Computational Molecular Physics* (ed. G.H.F. Dierksen and S. Wilson) Reidel, Dordrecht 1983.

19 M. Dupuis, J. Rys and H.F. King, J. Chem. Phys. **65**, 111 (1976)

APPENDICES

Appendix I
Separating the motion of the center of mass in classical mechanics

Newton's equations for N pointlike masses:

$$m_i \ddot{\vec{r}}_i = \vec{f}_i \qquad i = 1, 2, \ldots, N \tag{I.1}$$

where

m_i is the mass of the ith particle;

$\ddot{\vec{r}}_i$ is the acceleration of the ith particle; ($\vec{r}_i$ is its radius vector and "$\cdot$" denotes differentiation with respect to the time)

$\vec{f}_i$ is the force acting on the ith particle.

Summing up the equations:

$$\sum_i m_i \ddot{\vec{r}}_i = \sum_i \vec{f}_i = \vec{F} \,. \tag{I.2}$$

$\vec{F}$ is the vectorial sum of the *external* forces acting on the particles. (According to Newton's third law the vectorial sum of the internal forces is 0.)

We introduce the total mass M of the system:

$$M = \sum_i m_i \,. \tag{I.3}$$

Multiplying and dividing by M:

$$M \sum_i \frac{m_i}{M} \ddot{\vec{r}}_i = \vec{F} \quad \text{i.e.,} \qquad M\ddot{\vec{R}} = \vec{F} \tag{I.4}$$

where

$$\vec{R} = \sum_i \frac{m_i}{M} \vec{r}_i \tag{I.5}$$

is *the radius-vector of the center of mass:* the mass-weighted average of the radius-vectors of the individual particles.

If the system is closed, $\vec{F} = 0$:

$$M\ddot{\vec{R}} = 0 \quad \longrightarrow \quad \dot{\vec{R}} = const \tag{I.6}$$

—the center of mass of a closed system is at rest or moves with a constant velocity along a constant direction.

To study the internal motion of the system, let us introduce the relative coordinates $\vec{r}_i'$, defined with respect to the center of the mass:

$$\vec{r}_i' = \vec{r}_i - \vec{R} \tag{I.7}$$

then

$$\ddot{\vec{r}}_i' = \ddot{\vec{r}}_i - \ddot{\vec{R}} \ . \tag{I.8}$$

For a closed system $\ddot{\vec{R}} = 0$, therefore

$$\ddot{\vec{r}}_i' = \ddot{\vec{r}}_i = \frac{1}{m_i}\vec{f}_i \ , \tag{I.9}$$

i.e.,—as is well known—in the inertial system of the center of the mass the same Newton equations hold as in the laboratory system. (The difference is only in the initial conditions: in the center of the mass system the total linear momentum of the system must be zero, so only initial velocities satisfying this condition may be selected.)

The kinetic energy becomes the sum of the kinetic energy of the internal motion and that of the motion of the center of the mass:

Let us denote $\vec{v}_i = \dot{\vec{r}}_i$; $\vec{v}_i' = \dot{\vec{r}}_i'$; $\vec{V}_R = \dot{\vec{R}}$; then it follows from the definition $\vec{r}_i' = \vec{r}_i - \vec{R}$ that $\vec{r}_i = \vec{r}_i' + \vec{R}$; $\vec{v}_i = \dot{\vec{r}}_i = \dot{\vec{r}}_i' + \dot{\vec{R}} = \vec{v}_i' + \vec{V}_R$. We may write, therefore,

$$\tfrac{1}{2}\sum_i m_i \vec{v}_i^{\,2} = \tfrac{1}{2}\sum_i m_i \vec{v}_i'^{\,2} + \tfrac{1}{2}\underbrace{\sum_i m_i}_{M}\vec{V}_R^2 + \sum_i m_i \vec{v}_i' \vec{V}_R \ . \tag{I.10}$$

The first two terms on the right-hand-side represent the kinetic energy of the internal (relative) motion and that of the center of the mass (overall mass M) whereas the last term vanishes. In fact:

$$\sum_i m_i \vec{v}_i' \vec{V}_R \equiv M\vec{V}_R \sum_i \frac{m_i}{M}\vec{v}_i' \ . \tag{I.11}$$

Factor $\sum_i \frac{m_i}{M}\vec{v}_i{}'$ is the velocity of the center of the mass in the system of the center of the mass—it is obviously 0:

$$\sum_i \frac{m_i}{M}\vec{v}_i{}' = \sum_i \frac{m_i}{M}(\dot{\vec{r}}_i - \dot{\vec{R}}) = \underbrace{\sum_i \frac{m_i}{M}\dot{\vec{r}}_i}_{\dot{\vec{R}}} - \dot{\vec{R}}\underbrace{\sum_i \frac{m_i}{M}}_{1} = \dot{\vec{R}} - \dot{\vec{R}} = 0\,. \quad \text{(I.12)}$$

The same separation of the motion of the center of mass is also possible in quantum mechanics, but is much more complicated—see Section 1.2 of Chapter 1.

Appendix II
Reducing the two-body problem to two one-body ones in classical mechanics

Newton equations:

$$\begin{aligned} m_1\ddot{\vec{r}}_1 &= \vec{f}_1 \\ m_2\ddot{\vec{r}}_2 &= \vec{f}_2 \end{aligned} \tag{II.1}$$

Closed system (Newton's third law):

$$\vec{f}_2 = -\vec{f}_1 \tag{II.2}$$

The mass of the system:

$$M = m_1 + m_2 \tag{II.3}$$

Motion of the center of mass:

Multiplying both equations of motion by $\dfrac{1}{M} = \dfrac{1}{m_1 + m_2}$ and summing:

$$\ddot{\vec{R}} = \frac{m_1\ddot{\vec{r}}_1 + m_2\ddot{\vec{r}}_2}{m_1 + m_2} = 0 \quad \longrightarrow \quad \dot{\vec{R}} = \frac{m_1\dot{\vec{r}}_1 + m_2\dot{\vec{r}}_2}{m_1 + m_2} = \text{const} \tag{II.4}$$

(as in Appendix I—the center of mass behaves like a single particle).

The relative motion:

Dividing the equations of motion by m_1 and m_2, respectively:

$$\begin{aligned} \ddot{\vec{r}}_1 &= \frac{1}{m_1}\vec{f}_1 \\ \ddot{\vec{r}}_2 &= \frac{1}{m_2}\vec{f}_2 = -\frac{1}{m_2}\vec{f}_1 \qquad (\text{closed system, } \vec{\mathrm{f}}_2 = -\vec{\mathrm{f}}_1) \end{aligned} \tag{II.5}$$

Subtracting:

$$\ddot{\vec{r}}_1 - \ddot{\vec{r}}_2 = \frac{1}{m_1}\vec{f}_1 + \frac{1}{m_2}\vec{f}_1 = \underbrace{\left(\frac{1}{m_1} + \frac{1}{m_2}\right)}_{\frac{1}{\mu}} \vec{f}_1 \tag{II.6}$$

We define:

$\vec{r} = \vec{r}_1 - \vec{r}_2$ – the coordinate of the *relative motion;*

μ – the *"reduced mass."* We have

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} = \frac{m_1 + m_2}{m_1 m_2} \tag{II.7}$$

i.e.,

$$\mu = \frac{m_1 m_2}{m_1 + m_2} = \frac{m_2}{1 + \frac{m_2}{m_1}} \tag{II.8}$$

Important special cases: *a)* if $m_1 = m_2$, $\mu = \frac{1}{2}m_1$; *b)* $\lim\limits_{m_1 \to \infty} \mu = m_2$.

Using the definitions of $\vec{r}$ and μ one has from (II.6):

$$\mu \ddot{\vec{r}} = \vec{f}_1 \tag{II.9}$$

—the equation of the relative motion—formally a one-particle problem.

The total kinetic energy is also the sum of the kinetic energy of the center of mass and that of relative motion corresponding to the reduced mass μ. In fact, according to equations (I.10)–(I.12) of Appendix I, we have

$$T = \frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2} = \frac{M V_R^2}{2} + \frac{m_1 v_1'^2}{2} + \frac{m_2 v_2'^2}{2}\,, \tag{II.10}$$

where v_1', v_2' are the velocities of the particles with respect to the C.O.M. system.

Furthermore, the overall linear momentum with respect to the C.O.M. system vanishes, so one can express

$$\vec{v}_2' = -\frac{m_1}{m_2}\vec{v}_1' \tag{II.11}$$

and then

$$T = \frac{M V_R^2}{2} + \frac{1}{2}\left(m_1 + \frac{m_1^2}{m_2}\right) v_1'^2 = \frac{M V_R^2}{2} + \frac{1}{2} m_1 \left(1 + \frac{m_1}{m_2}\right) v_1'^2\,. \tag{II.12}$$

At the same time, the relative velocity $\vec{v} = \dot{\vec{r}} = \vec{v}_1 - \vec{v}_2 = \vec{v}_1' - \vec{v}_2'$ is, according to (II.11),

$$\vec{v} = \vec{v}_1' - \left(-\frac{m_1}{m_2}\vec{v}_1'\right) = \left(1 + \frac{m_1}{m_2}\right)\vec{v}_1'\,. \tag{II.13}$$

That means in turn that

$$\vec{v}_1' = \left(1 + \frac{m_1}{m_2}\right)^{-1} \vec{v} \,. \tag{II.14}$$

By substituting this into (II.12) we get

$$T = \frac{MV_R^2}{2} + \frac{1}{2}\,\frac{m_1}{1 + \dfrac{m_1}{m_2}} v^2 = \frac{MV_R^2}{2} + \frac{\mu v^2}{2} \,. \tag{II.15}$$

Q.E.D.

Appendix III
Analogy between differentials and variations

dy	δJ
the (first) differential of the *function* $y = f(x)$.	the (first order) variation of the *functional* $J = J(\Psi)$.
If Δx $(= dx)$ is an arbitrary (small) change of the independent variable x, then	If $\delta\Psi$ is an arbitrary (small) variation (change within the class of admitted functions) of Ψ, e.g., $\delta\Psi = \eta\Phi$; $\eta \to 0$ and Φ is a finite function satisfying the boundary conditions of the problem, then
$y(x+\Delta x) - y(x) = dy + \mathcal{O}(\Delta x^2)$.	$J(\Psi+\delta\Psi) - J(\Psi) = \delta J + \mathcal{O}(\delta\Psi^2)$.
One can write: $$dy = \frac{\partial}{\partial\alpha} f(x+\alpha\Delta x)\Big\vert_{\alpha=0}$$	One can consider as a definition: $$\delta J = \frac{\partial}{\partial\alpha}[J(\Psi+\alpha\delta\Psi)]\Big\vert_{\alpha=0}$$
because the r.h.s. (as derivative of an indirect function of α): $= f'(x+\alpha\Delta x)\big\vert_{\alpha=0} \times \Delta x$ $= f'(x)\Delta x = dy$	This analogy indicates that the rules of differentiation of products, fractions, etc., also apply for calculating variations.

Appendix IV
Euler's theorem for homogeneous functions

The function $f = f(x_1, x_2, \ldots, x_n)$ is a homogeneous function of kth order of the variables $x_1, x_2, \ldots, x_n$ if

$$f = f(\alpha x_1, \alpha x_2, \ldots, \alpha x_n) = \alpha^k f(x_1, x_2, \ldots, x_n) \ . \tag{IV.1}$$

For homogeneous functions one has *Euler's theorem:*

$$\sum_{i=1}^{n} x_i \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} = k f(x_1, x_2, \ldots, x_n) \tag{IV.2}$$

Proof.

Let us define the indirect function $u = u(t)$ of the variable t as

$$u(t) = (tx_1^0, tx_2^0, \ldots, tx_n^0) \tag{IV.3}$$

where the point $M(x_1^0, x_2^0, \ldots, x_n^0)$ is an arbitrary internal point of the domain of definition of the function f; this means that all the variables x_i in f are considered as functions $x_i = tx_i^0$ of the variable t.

Differentiating the indirect function:

$$\left.\frac{du}{dt}\right|_{t=1} = \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x_i} \frac{dx_i}{dt}\right|_{t=1} , \tag{IV.4}$$

but

$$\left.\frac{\partial f}{\partial x_i}\right|_{t=1} = \frac{\partial f}{\partial x_i^0}; \qquad \left.\frac{dx_i}{dt}\right|_{t=1} = \left.\frac{d}{dt}(tx_i^0)\right|_{t=1} = x_i^0 \ . \tag{IV.5}$$

therefore

$$\left.\frac{du}{dt}\right|_{t=1} = \sum_{i=1}^{n} x_i^0 \frac{\partial f}{\partial x_i^0} \ . \tag{IV.6}$$

From the other side, because f is a homogeneous function,

$$u(t) = (tx_1^0, tx_2^0, \ldots, tx_n^0) = t^k f(x_1^0, x_2^0, \ldots, x_n^0) \ . \tag{IV.7}$$

Differentiating this

$$\left.\frac{du}{dt}\right|_{t=1} = kt^{k-1} f(x_1^0, x_2^0, \ldots, x_n^0)\Big|_{t=1} = kf(x_1^0, x_2^0, \ldots, x_n^0) \ . \tag{IV.8}$$

Comparing with the previous formula:

$$\sum_{i=1}^{n} x_i^0 \frac{\partial f}{\partial x_i^0} = kf(x_1^0, x_2^0, \ldots, x_n^0) \ . \tag{IV.9}$$

As $M(x_1^0, x_2^0, \ldots, x_n^0)$ is an arbitrary internal point of the domain of definition of f, the superscripts "0" may be omitted. Q.E.D.

Appendix V
The virial theorem in classical mechanics

Time average of a full derivative for finite motion

If the system performs *finite motion* (i.e., all coordinates and momenta remain finite) then the time average (denoted by horizontal bar) of any physical quantity, which can be presented as a *full derivative* over the time, is zero; i.e., if $f(t) = \frac{d}{dt}F$ and $F(t)$ is finite, then $\overline{f} = 0$:

$$\overline{f} = \lim_{T\to\infty} \frac{1}{T}\int_0^T \frac{dF}{dt} = \lim_{T\to\infty} \frac{F(T) - F(0)}{T} = 0 \,. \tag{V.1}$$

The virial theorem

Let us consider the time average of the quantity $\frac{d}{dt}\sum_i \vec{r}_i\vec{p}_i$, where the summation is over all the particles in the system, $\vec{r}_i$ is the radius vector of the ith particle, and $\vec{p}_i$ is its (linear) momentum. If all $\vec{r}_i$, $\vec{p}_i$ remain finite, then

$$\overline{\frac{d}{dt}\sum_i \vec{r}_i\vec{p}_i} = 0 \,. \tag{V.2}$$

From the other side

$$\frac{d}{dt}\sum_i \vec{r}_i\vec{p}_i = \sum_i \vec{v}_i\vec{p}_i + \sum_i \vec{r}_i\dot{\vec{p}}_i \,. \tag{V.3}$$

As $\vec{v}_i\vec{p}_i = m_i v_i^2$, the first sum equals $2T$, T being the kinetic energy of the system. Therefore, one has

$$2\overline{T} = -\overline{\sum_i \vec{r}_i\dot{\vec{p}}_i} \,. \tag{V.4}$$

The last sum can be transformed by using Newton's equation $m_i\ddot{\vec{r}}_i = \vec{F}_i$ written in the form $\dot{\vec{p}}_i = \vec{F}_i$:

$$2\overline{T} = -\overline{\sum_i \vec{r}_i\vec{F}_i} \ . \tag{V.5}$$

The quantity $\sum_i \vec{r}_i\vec{F}_i$ on the right-hand side is called the "virial" of the system. Each term of this sum depends on the origin of the coordinate system; the whole sum is, however, invariant, provided that the system is closed (all the forces are internal). In fact, if we move the origin O to the point O' having the radius vector $\vec{r}_0$, then the radius vectors in the new (primed) system will be $\vec{r}_i' = \vec{r}_i - \vec{r}_0$. The virial in the new system becomes

$$\sum_i \vec{r}_i'\vec{F}_i = \sum_i (\vec{r}_i - \vec{r}_0)\vec{F}_i = \sum_i \vec{r}_i\vec{F}_i - \vec{r}_0 \sum_i \vec{F}_i = \sum_i \vec{r}_i\vec{F}_i \tag{V.6}$$

because $\sum_i \vec{F}_i = 0$ for the sum of internal forces. This proves the invariance of the virial $\sum_i \vec{r}_i\vec{F}_i$.

For a conservative system $\vec{F}_i = -\nabla_i U \equiv \left(\vec{i}\,\frac{\partial U}{\partial x} + \vec{j}\,\frac{\partial U}{\partial y} + \vec{k}\,\frac{\partial U}{\partial z}\right) \equiv -\frac{\partial U}{\partial \vec{r}_i}$, therefore one has from (V.5)

$$2\overline{T} = \overline{\sum_i \vec{r}_i \frac{\partial U}{\partial \vec{r}_i}} \ . \tag{V.7}$$

If the potential U is a kth order homogeneous function of the radius vectors $\vec{r}_i$, then the r.h.s. can be rewritten as $k\overline{U}$, therefore

$$2\overline{T} = k\overline{U} \ . \tag{V.8}$$

Special cases:

Harmonic oscillator: $k = 2,\ \overline{T} = \overline{U},\ E = 2\overline{U}$
Coulomb interaction: $k = -1,\ \overline{T} = -\frac{1}{2}\overline{U},\ E = \frac{1}{2}\overline{U} < 0$
(bounded motion with negative total energy)

Appendix VI
The electronic Schrödinger equation in atomic units

In quantum chemistry we deal mostly with the electronic Schrödinger equation. For simplicity, we shall omit the subscripts "e" and usually shall not explicitly indicate the arguments:

$$\hat{H}\Psi = E\Psi \tag{VI.1}$$

It is worthwhile to introduce the system of "*atomic units,*" in which by definition $\hbar = 1, e_0 = 1$, and $m_e = 1$. Then the unit of length is the Bohr radius (the "radius" of the H-atom) usually called "Bohr" or "atomic unit of the length"(or simply "a.u.") $\cong$ 0.52917726 Å; the velocity of the light is $c \approx 137$ a.u., and the atomic unit of energy ("Hartree"; a.u.; E_h) $\cong$ 27.2113957 e.V. $\cong$ 627.509541 kcal/mol $\cong$ 2625.49992 kJ/mol $\cong$ 219474.625 cm^{-1} $\cong$ 315773.213 K (Kelvin—used as an energy unit that is equal to the thermal energy kT corresponding to the temperature 1K) . In these units the energy of the H atom in the fixed nucleus treatment is *exactly* $-\frac{1}{2}$ a.u.

If atomic units are used, the electronic Hamiltonian acquires a particularly simple form:

$$\hat{H} = -\tfrac{1}{2}\sum_{i=1}^{N_e}\Delta_i - \sum_{\alpha=1}^{N_N}\sum_{i=1}^{N_e}\frac{Z_\alpha}{r_{\alpha i}} + \sum_{i<j}^{N_e}\frac{1}{r_{ij}} + \sum_{\alpha<\beta}^{N_N}\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}\,. \tag{VI.2}$$

The last term describes nuclear–nuclear repulsion. It does not depend on the electronic wave function, so it represents an additive term in the energy. Therefore it is often not included in the electronic Hamiltonian but is added to the energy of the system of the electrons only afterward. (Usual names of the energy values obtained without and with the nuclear repulsion being added are "electronic energy" and "total energy," respectively.)

Appendix VII
The "bra-ket" formalism

1. Dirac's "bra" and "ket" vectors

It is useful to consider the wave functions φ, ψ, etc., as the "ket-vectors" (state vectors) $|\varphi\rangle$, $|\psi\rangle$, etc., and introduce for each of them the respective adjoint "bra-vector" $\langle\varphi|$, $\langle\psi|$ etc. If a bra-vector $\langle\varphi|$ is followed by a ket-vector $|\psi\rangle$, then one automatically forms the scalar product (overlap integral) $\langle\varphi|\psi\rangle$. (The names "bra" and "ket" originate from the English word "bracket" and were introduced by Dirac). This system of notations is extremely flexible and convenient, as it permits one to use very compact notations for projection operators etc., and facilitates their practical handling.

For instance, the projection of some wave function $\psi(x)$ on the normalized wave function $\varphi(x)$ is given by the integral-operator

$$\hat{P}_\varphi \psi(x) = \varphi(x) \int \varphi^*(x')\psi(x')dx' \tag{VII.1}$$

with a kernel $\varphi^*(x')$, where x, x' stand for all the variables on which φ and ψ depend, and the integration is defined accordingly. But

$$\int \varphi^*(x')\psi(x')dx' = \langle\varphi|\psi\rangle \tag{VII.2}$$

in standard notation, therefore, introducing the ket $|\varphi\rangle$ for the wave function φ, one can write

$$\hat{P}_\varphi|\psi\rangle = |\varphi\rangle\langle\varphi|\psi\rangle \ . \tag{VII.3}$$

This means that one has for the projector simply

$$\hat{P}_\varphi = |\varphi\rangle\langle\varphi| \ . \tag{VII.4}$$

In this manner one gets a very flexible formalism in which some simple formal rules should be obeyed:

1. Operators act on kets from the left, on bras from the right, e.g.,

$$\begin{aligned} |\psi_1\rangle &= \hat{A}|\varphi_1\rangle \\ \langle\psi_2| &= \langle\varphi_2|\hat{B} \end{aligned} \tag{VII.5}$$

($\hat{A}$ and $\hat{B}$ are some linear operators).

2. One has a scalar "between two points" and an operator "between two vertical bars":

$$\begin{aligned} \langle u|\ldots|v\rangle &\quad - \quad \text{a scalar} \\ |u\rangle\ldots\langle v| &\quad - \quad \text{an operator .} \end{aligned} \tag{VII.6}$$

Here the dots "... " represent operators and/or scalars. (In the second expression the part denoted by "... " may, of course, contain expressions with operators but, as a whole, it should represent a scalar.)

Similarly, expressions "open from left" are kets; those "open from right" are bras:

$$\begin{aligned} \ldots|u\rangle &\quad - \quad \text{a ket} \\ \langle v|\ldots &\quad - \quad \text{a bra .} \end{aligned} \tag{VII.7}$$

3. One can change between the bra and ket forms and get *equivalent* expressions by forming adjoints:

$$\begin{array}{rccll} \text{vectors} & |\varphi\rangle & \leftrightarrow & \langle\varphi| & (\text{ket} \leftrightarrow \text{bra}) \\ \text{scalars (numbers)} & a & \leftrightarrow & a^* & (\text{complex conjugation}) \\ \text{operators} & A & \leftrightarrow & A^\dagger & (\text{forming adjoints}) . \end{array} \tag{VII.8}$$

(As a special case of the scalars, scalar products transform as $\langle u|v\rangle \leftrightarrow \langle v|u\rangle = \langle u|v\rangle^*$.)

For instance, the eigenvalue equation $\hat{A}|\varphi\rangle = \lambda|\varphi\rangle$ is fully equivalent with the equation $\langle\varphi|\hat{A}^\dagger = \lambda^*\langle\varphi|$; if operator $\hat{A}$ is Hermitian, the latter equation reduces to $\langle\varphi|\hat{A} = \lambda\langle\varphi|$.

It is important to observe that—in accord with the rule $(\hat{A}\hat{B})^\dagger = B^\dagger A^\dagger$—one has

$$(|u\rangle\langle v|)^\dagger = |v\rangle\langle u| \tag{VII.9}$$

and, in general, when forming adjoints of more complex expressions, one has to proceed "from the back to the front." For instance, assuming $\hat{A}$ to be Hermitian but $\hat{B}$ not to be:

$$(\hat{B}|u\rangle\langle\varphi|\hat{A}|\psi\rangle\langle a|b\rangle\langle v|)^\dagger = |v\rangle\langle b|a\rangle\langle\psi|\hat{A}|\varphi\rangle\langle u|\hat{B}^\dagger . \tag{VII.10}$$

One can see the flexibility of this formalism by observing that the same expression in (VII.10) can also be written in the form$Q|u\rangle\langle v|B^\dagger$; where

$Q = \langle b|a\rangle\langle\psi|\hat{A}|\varphi\rangle$ is a scalar, or in the form $P\hat{X}\hat{Y}\hat{B}^\dagger$, where $P = \langle\psi|\hat{A}|\varphi\rangle$ is a scalar and $\hat{X} = |v\rangle\langle b|$, $\hat{Y} = |a\rangle\langle u|$ are operators, and so on.

2. Analogy with the matrix formalism

The rules of using bras and kets are completely similar to the matrix formalism that one uses in the case when the different functions and operators are represented in a complete *orthonormalized* basis set.

It is rather easy to rewrite different expressions from one formalism to another. The ket-vector $|\varphi\rangle$ corresponds to the column-vector $\mathbf{c}$ formed of the expansion coefficients of $|\varphi\rangle$ in the given orthonormalized basis $\{|\chi_i\rangle\}$: $|\varphi\rangle = \sum_i c_i|\chi_i\rangle$. Accordingly, the bra $\langle\varphi|$ corresponds to the row-vector $\mathbf{c}^\dagger$, adjoint to $\mathbf{c}$. The scalar product $\langle\varphi|\psi\rangle$ corresponds to the scalar product $\mathbf{c}^\dagger\mathbf{d}$ (assuming that $\mathbf{c}$ and $\mathbf{d}$ consist of the expansion coefficients of $|\varphi\rangle$ and $|\psi\rangle$, respectively); the operators are mapped to the matrices representing them in the given basis, e.g., if $|\varphi\rangle \sim \mathbf{c}$, $|\psi\rangle \sim \mathbf{d}$, (we denote here the correspondence by "$\sim$"), then

$$|\varphi\rangle\langle\psi| \sim \mathbf{c}\mathbf{d}^\dagger = \begin{pmatrix} \vdots \\ c \\ \vdots \end{pmatrix} (\ldots d^\dagger \ldots) \tag{VII.11}$$

—a matrix (dyadic product). At the same time

$$\langle\psi|\varphi\rangle = \mathbf{d}^\dagger\mathbf{c} = (\ldots d^\dagger \ldots) \begin{pmatrix} \vdots \\ c \\ \vdots \end{pmatrix} \tag{VII.12}$$

—$\langle\psi|\varphi\rangle$ and $\mathbf{d}^\dagger\mathbf{c}$ are scalars that are equal to each other if $\langle\psi| \sim \mathbf{d}^\dagger$ and $|\varphi\rangle \sim \mathbf{c}$.

One can obtain the matrix equivalent of any expression written out in the bra-ket formalism by introducing explicitly the expansions in the selected basis and collecting the respective coefficients, e.g.,

$$\begin{aligned}
|\varphi\rangle &= |\sum_i c_i|\chi_i\rangle = \sum_i c_i|\chi_i\rangle \sim \mathbf{c} \\
\langle\psi| &= \langle\sum_i d_i\chi_i| = \sum_i d_i^*\langle\chi_i| \sim \mathbf{d}^\dagger \\
\langle\psi|\varphi\rangle &= \langle\sum_i d_i\chi_i|\sum_j c_j\chi_j\rangle = \sum_{i,j} d_i^* \underbrace{\langle\chi_i|\chi_j\rangle}_{\delta_{ij}} c_j = \sum_i d_i^* c_i = \mathbf{d}^\dagger\mathbf{c} \\
|\varphi\rangle\langle\psi| &= |\sum_i c_i\chi_i\rangle\langle\sum_j d_j\chi_j| = \sum_{i,j} c_i d_j^* |\chi_i\rangle\langle\chi_j| \sim \mathbf{c}\mathbf{d}^\dagger\,.
\end{aligned} \tag{VII.13}$$

As noted previously, the projector on the state $|\varphi\rangle$ is $P_\varphi = |\varphi\rangle\langle\varphi|$; its mapping in terms of expansion coefficients is accordingly

$$\hat{P}_\varphi = |\varphi\rangle\langle\varphi| = |\sum_i c_i\chi_i\rangle\langle\sum_j c_j\chi_j| = \sum_{i,j} c_i c_j^* |\chi_i\rangle\langle\chi_j| \sim \mathbf{c}\mathbf{c}^\dagger = \mathbf{P}_\varphi \,. \tag{VII.14}$$

In fact, the projection of an arbitrary function described by ket $|\psi\rangle = \sum_j d_j|\chi_j\rangle$ on the one-dimensional subspace of $|\varphi\rangle \sim \mathbf{c}$ is, according to (VII.4)

$$\begin{aligned} \hat{P}_\varphi|\psi\rangle &= |\varphi\rangle\langle\varphi|\psi\rangle = |\sum_i c_i\chi_i\rangle\langle\sum_j c_j\chi_j|\sum_k d_k\chi_k\rangle \\ &= \sum_{i,j,k} c_i c_j^* d_k |\chi_i\rangle\langle\chi_j|\chi_k\rangle = \sum_{i,j,k} c_i c_j^* d_k \delta_{jk}|\chi_i\rangle \\ &= \sum_{i,j} c_i c_j^* d_j |\chi_i\rangle = \sum_i (\mathbf{c}\mathbf{c}^\dagger\mathbf{d})_i|\chi_i\rangle \,, \end{aligned} \tag{VII.15}$$

in accord with the known fact that the dyadic product $\mathbf{c}\mathbf{c}^\dagger$ is the projection matrix for the subspace spanned by vector $\mathbf{c} \sim |\varphi\rangle$. Also, the matrix elements $\langle\chi_k|\hat{P}_\varphi|\chi_l\rangle$ of the operator $\hat{P}_\varphi$ in the orthonormalized basis $\{|\chi_k\rangle\}$ are given by the elements of this dyadic:

$$\begin{aligned} (\mathbf{P}_\varphi)_{kl} &= \langle\chi_k|\hat{P}_\varphi|\chi_l\rangle = \langle\chi_k|\varphi\rangle\langle\varphi|\chi_l\rangle \\ &= \langle\chi_k|\sum_i c_i\chi_i\rangle\langle\sum_j c_j\chi_j|\chi_l\rangle = \sum_{i,j} c_i c_j^* \delta_{ki}\delta_{jl} = c_k c_l^* = (\mathbf{c}\mathbf{c}^\dagger)_{kl} \,. \end{aligned} \tag{VII.16}$$

3. The use of an overlapping basis

One can also generalize these mappings for the case when the basis $\{|\chi_i\rangle\}$ is not orthonormalized, but the expressions become more complex by the appearance of the overlap matrix $\mathbf{S}$ having the elements $S_{ij} = \langle\chi_i|\chi_j\rangle$. For instance,

$$\begin{aligned} \langle\varphi|\psi\rangle &= \langle\sum_i c_i\chi_i|\sum_j d_j\chi_j\rangle = \sum_{i,j} c_i^*\langle\chi_i|\chi_j\rangle d_j \\ &= \sum_{i,j} c_i^* S_{ij} d_j = \mathbf{c}^\dagger\mathbf{S}\mathbf{d} \,. \end{aligned} \tag{VII.17}$$

In the case of overlapping basis sets one has to distinguish between the matrix of the integrals of a linear operator $\hat{A}$ with the elements

$$A_{ij} = \langle\chi_i|\hat{A}|\chi_j\rangle \tag{VII.18}$$

and the matrix describing the linear transformation induced by that operator.

Let $|\varphi\rangle \sim \mathbf{c}$ and $|\psi\rangle \sim \mathbf{d}$ and let us assume that $|\varphi\rangle = \hat{A}|\psi\rangle$. We have therefore

$$|\varphi\rangle = |\sum_i c_i\chi_i\rangle = \hat{A}|\sum_j d_j\chi_j\rangle \ . \tag{VII.19}$$

By multiplying (VII.19) with $\langle\chi_l|$ ($l = 1, 2, \ldots$), one gets

$$\langle\chi_l|\sum_i c_i\chi_i\rangle = \langle\chi_l|\hat{A}|\sum_j d_j\chi_j\rangle \ , \tag{VII.20}$$

i.e.,

$$\sum_i S_{li}c_i = \sum_j A_{lj}d_j \tag{VII.21}$$

or, in matrix notations,

$$\mathbf{Sc} = \mathbf{Ad} \ . \tag{VII.22}$$

Multiplying (VII.22) by the inverse overlap matrix $\mathbf{S}^{-1}$ from left, we get

$$\mathbf{c} = \mathbf{S}^{-1}\mathbf{Ad} \ . \tag{VII.23}$$

This result indicates that in the case of an overlapping basis ($\mathbf{S} \neq \mathbf{1}$), the matrix of the linear transformation induced by operator $\hat{A}$ is $\mathbf{S}^{-1}\mathbf{A}$, where the elements of matrix $\mathbf{A}$ are defined according to (VII.18). For a projector $\hat{P}_\varphi$ this result means that the matrix of the integrals of $\hat{P}_\varphi$ is, as it is easy to see, $\mathbf{Scc}^\dagger\mathbf{S}$, whereas the projection matrix is $\mathbf{cc}^\dagger\mathbf{S}$. (This latter matrix has the correct idempotency, etc., properties discussed bellow for the projectors.)

4. Example of using the bra-ket formalism: The hypervirial theorem

Let us form the integral $\langle\Psi_i|\hat{H}\hat{Q} - \hat{Q}\hat{H}|\Psi_j\rangle$ containing the commutator of the Hamiltonian $\hat{H}$ and of an arbitrary operator $\hat{Q}$ for the exact solutions Ψ_i and Ψ_j of the time-independent Schrödinger equation. The Schrödinger equation for the exact state (ket-vector) $|\Psi_j\rangle$ is

$$\hat{H}|\Psi_j\rangle = E_j|\Psi_j\rangle \tag{VII.24}$$

and the adjoint Schrödinger equation for the exact bra-vector $\langle\Psi_i|$ is

$$\langle\Psi_i|\hat{H} = E_i\langle\Psi_i| \ . \tag{VII.25}$$

Using these relationships, we get (the arrows indicate the directions in which the Hamiltonian is acting):

$$\begin{aligned}\langle\Psi_i|\hat{H}\hat{Q} - \hat{Q}\hat{H}|\Psi_j\rangle &= \langle\Psi_i|\overleftarrow{\hat{H}}\,\hat{Q}|\Psi_j\rangle - \langle\Psi_i|\hat{Q}\,\overrightarrow{\hat{H}}|\Psi_j\rangle \\ &= E_i\langle\Psi_i|\hat{Q}|\Psi_j\rangle - E_j\langle\Psi_i|\hat{Q}|\Psi_j\rangle = (E_i - E_j)Q_{ik} \ .\end{aligned} \tag{VII.26}$$

This is the so-called ("off-diagonal", if $i \neq j$) "hypervirial theorem"; it is used in the theory of intensity of spectral lines.

In the particular case of $i = j$ ($\Psi_i = \Psi_j = \Psi$) we have the "diagonal" hypervirial theorem, according to which, for arbitrary operators and the *exact* wave function Ψ, one has for the commutator of the Hamiltonian

$$\langle \Psi | \hat{H}\hat{Q} - \hat{Q}\hat{H} | \Psi \rangle = 0 \,. \tag{VII.27}$$

The derivation of the virial theorem in Section 3.2 of Chapter 2 could directly be based on this result.

5. Projection operators

As discussed previously, the operator of the projection on the one-dimensional subspace spanned by vector $|\varphi\rangle$ is $\hat{P}_\varphi = |\varphi\rangle\langle\varphi|$. Analogously, if we have an n-dimensional subspace of orthonormalized vectors $\{|\varphi_i\rangle\}$ ($i = 1, 2, \ldots, n$), then the respective projector for this subspace is the sum

$$\hat{P} = \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i| \,. \tag{VII.28}$$

It is easy to check two basic properties of the projection operators:

1. The projection operators are Hermitian—this follows from the definition in (VII.28) and the rules of handling the bra and ket vectors discussed previously.

2. The projection operators are *idempotent*, $\hat{P}^2 = \hat{P}$. In fact,

$$\begin{aligned} \hat{P}^2 &= \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i| \sum_{j=1}^{n} |\varphi_j\rangle\langle\varphi_j| \\ &= \sum_{i,j=1} |\varphi_i\rangle \underbrace{\langle\varphi_i|\varphi_j\rangle}_{\delta_{ij}} \langle\varphi_j| = \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i| = \hat{P} \,. \end{aligned} \tag{VII.29}$$

It is easy to see that any vector that can be expanded as a linear combination of the vectors $|\varphi_i\rangle$ spanning the subspace studied is left unchanged when the projection operator $\hat{P}$ is acting on it, whereas any vector which is orthogonal to this subspace is annihilated. In other words, any vector completely lying in the subspace of $\{|\varphi_i\rangle\}$ is an eigenvector of $\hat{P}$ with an eigenvalue 1, and any vector orthogonal to the subspace is an eigenvector with an eigenvalue 0.

The projector $\hat{P}$ remains invariant if one replaces the vectors $\{|\varphi_i\rangle\}$ by any other set of orthonormalized vectors $\{|\psi_i\rangle\}$ that completely lie in the same subspace. As both sets of vectors $\{|\varphi_i\rangle\}$ and $\{|\psi_i\rangle\}$ are orthonormalized, they should be connected by a *unitary* transformation $\mathbf{U}$:

$$|\psi_j\rangle = \sum_{k=1}^{n} U_{kj} |\varphi_k\rangle \,. \tag{VII.30}$$

Then the projector $\hat{P}'$ built up of the n functions $|\psi_j\rangle$ is

$$\begin{aligned}
\hat{P}' &= \sum_{j=1}^{n} |\psi_j\rangle\langle\psi_j| = \sum_{j=1}^{n} |\sum_{k=1}^{n} U_{kj}\varphi_k\rangle\langle\sum_{l=1}^{n} U_{lj}\varphi_l| \\
&= \sum_{j,k,l=1}^{n} U_{kj}U_{lj}^{*}|\varphi_k\rangle\langle\varphi_l| = \sum_{k,l=1}^{n} \underbrace{\left(\sum_{j=1}^{n} U_{kj}(\mathbf{U}^{\dagger})_{jl}\right)}_{\delta_{kl}} |\varphi_k\rangle\langle\varphi_l| \\
&= \sum_{k=1}^{n} |\varphi_k\rangle\langle\varphi_k| = \hat{P} \,. \qquad \text{(VII.31)}
\end{aligned}$$

Often we have to consider a subspace that is defined by some *non-orthogonal* vectors $\{|\varphi_i\rangle\}$. Then one can build up the projector on that subspace by constructing a set of orthonormalized vectors in the same subspace. Let us assume that the overlap matrix of the vectors $\{|\varphi_i\rangle\}$ spanning the subspace is $\boldsymbol{\sigma}$, i.e.,

$$\langle\varphi_i|\varphi_j\rangle = \sigma_{ij} \,. \qquad \text{(VII.32)}$$

Then the probably most straightforward way of producing an orthonormalized set of vectors $\{|\psi_k\rangle\}$, spanning the same subspace as $\{|\varphi_i\rangle\}$, is the use of Löwdin's symmetric orthogonalization (Section 2 of Chapter 3):

$$|\psi_k\rangle = \sum_{i=1}^{n} (\boldsymbol{\sigma}^{-1/2})_{ik}|\varphi_i\rangle \,. \qquad \text{(VII.33)}$$

Then the projector on the subspace is

$$\begin{aligned}
\hat{P} &= \sum_{k=1}^{n} |\psi_k\rangle\langle\psi_k| = \sum_{k=1}^{n} |\sum_{i=1}^{n} (\boldsymbol{\sigma}^{-1/2})_{ik}|\varphi_i\rangle\langle\sum_{j=1}^{n} (\boldsymbol{\sigma}^{-1/2})_{jk}\varphi_j| \\
&= \sum_{i,j,k=1}^{n} |\varphi_i\rangle(\boldsymbol{\sigma}^{-1/2})_{ik}(\boldsymbol{\sigma}^{-1/2})_{kj}\langle\varphi_j| = \sum_{i,j=1}^{n} |\varphi_i\rangle(\boldsymbol{\sigma}^{-1})_{ij}\langle\varphi_j| \qquad \text{(VII.34)}
\end{aligned}$$

—where we have taken into account account the Hermiticity of matrix $\boldsymbol{\sigma}^{-1/2}$.

6. Resolution of identity

If the vectors $\{|\varphi_i\rangle\}$ form a complete orthonormalized set of functions (basis) in the given (Hilbert) space of finite or infinite dimensions, then every function $|\psi\rangle$ can be expanded ("almost everywhere") as

$$|\psi\rangle = \sum_i c_i|\varphi_i\rangle \qquad \text{(VII.35)}$$

with the values of the coefficients (Fourier-amplitudes)

$$c_i = \langle \varphi_i | \psi \rangle \ . \tag{VII.36}$$

Substituting (VII.36) into (VII.35) we have

$$\begin{aligned} |\psi\rangle &= \sum_i c_i |\varphi_i\rangle = \sum_i \langle \varphi_i | \psi \rangle |\varphi_i\rangle = \sum_i |\varphi_i\rangle\langle\varphi_i|\psi\rangle \\ &= \left(\sum_i |\varphi_i\rangle\langle\varphi_i| \right) |\psi\rangle \ . \end{aligned} \tag{VII.37}$$

That means that the parentheses contain the identity operator $\hat{I}$:

$$\hat{I} = \sum_i |\varphi_i\rangle\langle\varphi_i| \tag{VII.38}$$

in the form called "resolution of identity" by means of the functions of a complete basis set. (As (VII.37)–(VII.38) express the completeness of the set $\{|\varphi_i\rangle\}$, the equality of type (VII.37) is also called "closure relationship.")

As $\hat{P} = |\varphi_i\rangle\langle\varphi_i|$ is the projection operator on the basis vector $|\varphi_i\rangle$, the resolution of identity (VII.38) expresses the fact that every vector can be presented as the sum of its projections on all the basis vectors. In other words, the projection of every vector on the complete space equals the vector itself—in agreement with the general property that any vector completely lying in the (sub)space considered is an eigenvector with eigenvalue 1 of the projector on that (sub)space.

7. Spectral resolution of Hermitian operators

Let us consider a Hermitian operator $\hat{A}$. As is known from mathematics, its eigenvectors $|\varphi_i\rangle$ form a complete orthonormalized set; one can, therefore, build a resolution of identity with their use:

$$\hat{I} = \sum_i |\varphi_i\rangle\langle\varphi_i| \ . \tag{VII.39}$$

Let us now consider operator $\hat{A}$ acting on an arbitrary function $|\psi\rangle$ and insert this $\hat{I}$ after $\hat{A}$:

$$\begin{aligned} \hat{A}|\psi\rangle &= \hat{A}\hat{I}|\psi\rangle = \hat{A}\left(\sum_i |\varphi_i\rangle\langle\varphi_i| \right) |\psi\rangle \\ &= \left(\sum_i a_i |\varphi_i\rangle\langle\varphi_i| \right) |\psi\rangle \ , \end{aligned} \tag{VII.40}$$

where we have taken into account that $|\varphi_i\rangle$ is assumed to be an eigenvector of $\hat{A}$, i.e., $\hat{A}|\varphi_i\rangle = a_i|\varphi_i\rangle$, a_i being the respective eigenvalue. (VII.40) means that any Hermitian operator can be presented in the form of its "spectral resolution"

$$\hat{A} = \sum_i a_i|\varphi_i\rangle\langle\varphi_i| \tag{VII.41}$$

by using its eigenvectors and eigenvalues.

8. The case of non-Hermitian operators—biorthogonal sets of functions

In most cases one can appropriately generalize the relationship in (VII.41) for non-Hermitian operators, as well, by utilizing the biorthogonal character of their left and right eigenvectors. Exceptions are the "pathological" cases in which the operator in question has fewer eigenvectors than the dimension of the linear space in which it is defined. (In such cases the matrix of the operator cannot be diagonalized by a similarity transformation. We shall not consider such singular cases, representing only mathematical, rather than physical interest. Moreover, these cases cannot even be identified in the practical numerical work because an infinitesimal change in the matrix elements eliminates this special property of the matrix.)

Let $\hat{A}$ be a non-Hermitian operator with $\{|\varphi_i\rangle\}$ and $\{\langle\psi_j|\}$ being its right and left eigenvectors, respectively:

$$\begin{aligned} \hat{A}|\varphi_i\rangle &= a_i|\varphi_i\rangle \\ \langle\psi_j|\hat{A} &= b_j\langle\psi_j| \ . \end{aligned} \tag{VII.42}$$

Now, we shall proceed similarly to the approach by which the orthogonality of eigenvectors of Hermitian operators and matrices is usually proved. We multiply the second equation in (VII.42) by a $|\varphi_i\rangle$

$$\langle\psi_j|\hat{A}|\varphi_i\rangle = b_j\langle\psi_j|\varphi_i\rangle \tag{VII.43}$$

and observe that $\hat{A}|\varphi_i\rangle = a_i|\varphi_i\rangle$ according to the first equation in (VII.42), so we get

$$a_i\langle\psi_j|\varphi_i\rangle = b_j\langle\psi_j|\varphi_i\rangle \ . \tag{VII.44}$$

Equation (VII.44) indicates that either

a) $a_i \neq b_j$ — and then $\langle\psi_j|\varphi_i\rangle = 0$, or

b) $a_i = b_j$ — and then in general, $\langle\psi_j|\varphi_i\rangle \neq 0$. We may choose the normalization of $\langle\psi_j|$ and/or $|\varphi_i\rangle$ as to ensure that $\langle\psi_j|\varphi_i\rangle = 1$.

If follows from these results that the left and right eigenvalues of $\hat{A}$ are equal by pairs and the left and right eigenvectors form mutually *biorthogonal* sets.

(In the case of degenerate $a_i = b_j$ values, they can be chosen in such manner.) We shall denote by $\sim$ over the letter φ the biorthogonal counterpart of $|\varphi_j\rangle$:

$$|\tilde{\varphi}_j\rangle = |\psi_j\rangle \ . \tag{VII.45}$$

It has the property

$$\langle\tilde{\varphi}_j|\varphi_i\rangle = \delta_{ij} \tag{VII.46}$$

and also

$$\langle\tilde{\varphi}_j|\hat{A} = a_j\langle\tilde{\varphi}_j| \ . \tag{VII.47}$$

One can construct a resolution of identity by using these biorthogonal sets of functions:

$$\hat{I} = \sum_i |\varphi_i\rangle\langle\tilde{\varphi}_i| \ . \tag{VII.48}$$

In fact, let us consider the action of the sum on the right-hand-side of (VII.48) on an arbitrary function $|\psi\rangle = \sum_j c_j|\varphi_j\rangle$:

$$\sum_i |\varphi_i\rangle\langle\tilde{\varphi}_i|\sum_j c_j|\varphi_j\rangle = \sum_{i,j} c_j|\varphi_i\rangle \underbrace{\langle\tilde{\varphi}_i|\varphi_j\rangle}_{\delta_{ij}} = \sum_j c_j|\varphi_j\rangle \tag{VII.49}$$

indicating the correctness of (VII.48). (Note that the form $\hat{I} = \sum_i |\tilde{\varphi}_i\rangle\langle\varphi_i|$ is also valid.)

Now, one may write for arbitrary $|\psi\rangle$

$$\begin{aligned} \hat{A}|\psi\rangle &= \hat{A}\hat{I}|\psi\rangle = \hat{A}\sum_i |\varphi_i\rangle\langle\tilde{\varphi}_i|\psi\rangle \\ &= \sum_i \hat{A}|\varphi_i\rangle\langle\tilde{\varphi}_i|\psi\rangle = \left(\sum_i a_i|\varphi_i\rangle\langle\tilde{\varphi}_i|\right)|\psi\rangle \end{aligned} \tag{VII.50}$$

i.e. the spectral resolution of a non-Hermitian operator in terms of its left and right eigenvectors $\langle\tilde{\varphi}_i|$ and $|\varphi_i\rangle$, respectively, is

$$\hat{A} = \sum_i a_i|\varphi_i\rangle\langle\tilde{\varphi}_i| \ . \tag{VII.51}$$

For a Hermitian operator the left and right eigenvectors are adjoints of each other (i.e., $\langle\tilde{\varphi}_i| = \langle\varphi_i|$), and (VII.51) reduces to (VII.41).

9. The trace of the projector

The trace of the projection operator $\hat{P}$ is equal to the number n of the dimensions of the subspace in question. We consider

$$\hat{P} = \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i| \ , \tag{VII.52}$$

which is defined in an m-dimensional linear space (m may be also infinite) that is referred to an orthonormalized basis $\{|\psi_j\rangle\}$, and therefore, the vectors $|\varphi_i\rangle$ can be expanded in this basis set as:

$$|\varphi_i\rangle = \sum_{j=1}^{m} c_j^i |\psi_j\rangle \ . \tag{VII.53}$$

The trace of $\hat{P}$ is then

$$\mathrm{Tr}\{\hat{P}\} = \sum_{k=1}^{m} \langle\psi_k|\hat{P}|\psi_k\rangle = \sum_{k=1}^{m} \langle\psi_k| \sum_{i=1}^{n} |\varphi_i\rangle\langle\varphi_i|\psi_k\rangle \ . \tag{VII.54}$$

By substituting the expansion in (VII.53), we get

$$\begin{aligned} \mathrm{Tr}\{\hat{P}\} &= \sum_{k=1}^{m} \langle\psi_k| \sum_{i=1}^{n} \sum_{j=1}^{m} c_j^i \psi_j\rangle\langle \sum_{l=1}^{m} c_l^i \psi_l|\psi_k\rangle \\ &= \sum_{i=1}^{n} \sum_{j,k,l=1}^{m} c_j^i c_l^{i*} \delta_{kj}\delta_{lk} = \sum_{i=1}^{n} \sum_{k=1}^{m} c_k^i c_k^{i*} = \sum_{i=1}^{n} 1 = n \ , \end{aligned} \tag{VII.55}$$

where we have utilized the orthonormalized character of the basis $\{|\psi_k\rangle\}$ and the fact that each $|\varphi_i\rangle$ is normalized to unity.

As it is easy to see, the same result holds to the trace of the matrix **P** representing operator $\hat{P}$ in the orthonormalized basis $\{|\psi_k\rangle\}$.

Appendix VIII
Collection of formulae for Rayleigh–Schrödinger perturbation theory (nondegenerate case)

$\hat{H} = \hat{H}^0 + \hat{V}$ – the perturbed Hamiltonian

$\hat{H}\Psi_i = E_i\Psi_i$ – the perturbed Schrödinger equation

$\hat{H}\Psi_l^0 = E_l^0\Psi_l^0$ – the unperturbed Schrödinger equation

Energy

$$E = E_i^0 + \sum_{n=1}^{\infty} \varepsilon_i^{(n)}$$

where (up to fourth order):

$$\begin{aligned}
\varepsilon_i^{(1)} &= V_{ii} \\
\varepsilon_i^{(2)} &= -\sum_{\substack{j \\ (j\neq i)}} \frac{|V_{ij}|^2}{E_j^0 - E_i^0} \\
\varepsilon_i^{(3)} &= \sum_{\substack{j,k \\ (j,k\neq i)}} \frac{V_{ik}V_{kj}V_{ji}}{(E_k^0 - E_i^0)(E_j^0 - E_i^0)} - \sum_{\substack{k \\ (k\neq i)}} \frac{V_{ii}|V_{ik}|^2}{(E_k^0 - E_i^0)^2}
\end{aligned}$$

$$
\begin{aligned}
\varepsilon_i^{(4)} &= -\sum_{\substack{j,k,l \\ (j,k,l\neq i)}} \frac{V_{ik}V_{kj}V_{jl}V_{li}}{(E_k^0-E_i^0)(E_j^0-E_i^0)(E_l^0-E_i^0)} \\
&\quad + \sum_{\substack{j,k \\ (j,k\neq i)}} \frac{|V_{ij}|^2|V_{ik}|^2}{(E_j^0-E_i^0)(E_k^0-E_i^0)^2} - \sum_{\substack{k \\ (k\neq i)}} \frac{V_{ii}^2|V_{ki}|^2}{(E_k^0-E_i^0)^3} \\
&\quad + \sum_{\substack{j,k \\ (j,k\neq i)}} \frac{V_{ii}V_{ik}V_{kj}V_{ji}}{(E_k^0-E_i^0)(E_j^0-E_i^0)} \left(\frac{1}{E_k^0-E_i^0}+\frac{1}{E_j^0-E_i^0}\right)
\end{aligned}
$$

(Also see the formulas in Section 1.4.2 of Chapter 4.)

Wave function

$$
\Psi_i = \Psi_i^0 + \sum_{n=1}^{\infty} \Phi_i^{(n)} \; ; \quad \Phi_i^{(n)} = \sum_{\substack{k \\ (k\neq i)}} p_{ik}^{(n)} \Psi_k^0 \, ,
$$

where (up to third order):

$$
\begin{aligned}
p_{ik}^{(1)} &= \frac{-V_{ki}}{E_k^0-E_i^0} \\
p_{ik}^{(2)} &= \sum_{\substack{j \\ (j\neq i)}} \frac{V_{kj}V_{ji}}{(E_k^0-E_i^0)(E_j^0-E_i^0)} - \frac{V_{ii}V_{ki}}{(E_k^0-E_i^0)^2} \\
p_{ik}^{(3)} &= -\sum_{\substack{j,l \\ (j,l\neq i)}} \frac{V_{kj}V_{jl}V_{li}}{(E_k^0-E_i^0)(E_j^0-E_i^0)(E_l^0-E_i^0)} \\
&\quad + \sum_{\substack{j \\ (j\neq i)}} \frac{|V_{ij}|^2V_{ki}}{(E_j^0-E_i^0)(E_k^0-E_i^0)^2} - \frac{V_{ii}^2V_{ki}}{(E_k^0-E_i^0)^3} \\
&\quad + \sum_{\substack{j \\ (j\neq i)}} \frac{V_{ii}V_{kj}V_{ji}}{(E_k^0-E_i^0)(E_j^0-E_i^0)} \left(\frac{1}{E_k^0-E_i^0}+\frac{1}{E_j^0-E_i^0}\right)
\end{aligned}
$$

Appendix IX
Direct products of matrices

Let **A** and **B** be matrices of dimensions $m_A \times n_A$ and $m_B \times n_B$, with the elements A_{ij} and B_{ij}, respectively. Then matrix **C** of dimensions $m_A m_B \times n_A n_B$ with the elements

$$C_{IJ} = C_{(ij)(kl)} = A_{ik} B_{jl} \tag{IX.1}$$

is called the direct product of matrices **A** and **B** and is denoted as

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} \ . \tag{IX.2}$$

In (IX.1) subscript I enumerates the pairs of subscripts (i, j) in the sequence (1,1), (1,2), . . ., (1,m_B), (2,1), (2,2), . . ., (m_A,m_B), and J enumerates the pairs (k, l) in the sequence (1,1), (1,2), . . ., (1, n_B), (2,1), (2,2), . . ., (n_A,n_B). In the particular case of vectors (one-column matrices) the direct product

$$\mathbf{c} = \mathbf{a} \otimes \mathbf{b} \tag{IX.3}$$

of the vectors **a** and **b** of dimensions m_A and m_B, respectively, is a vector of dimension $m_A m_B$ having the elements

$$c_I = c_{(ij)} = a_i b_j \ , \tag{IX.4}$$

where the definition of I is the same as given previously.

One can easily check the relationships

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= \mathbf{AC} \otimes \mathbf{BD} \\ (\mathbf{A} \otimes \mathbf{B})(\mathbf{c} \otimes \mathbf{d}) &= \mathbf{Ac} \otimes \mathbf{Bd} \ . \end{aligned} \tag{IX.5}$$

(These formulas are meaningful, of course, only if the number of columns of matrix **A** is equal to the number of rows of matrix **C** and vector **c**, respectively,

and similar conditions should hold for matrices **B** and **D**, or for matrix **B** and vector **d**.)

The direct product of three (or more) matrices can be defined by associativity

$$\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \tag{IX.6}$$

and so on; each matrix element is numbered by a triplet (quadruple, etc.) of subscripts corresponding to the individual matrices **A**, **B**, **C** (etc.).

Let us consider two linear variational problems defined in the linear spaces of functions $\{\varphi_i^A\}$ and $\{\varphi_j^B\}$ of dimensions m_A and m_B, respectively, and the corresponding secular equations

$$\begin{aligned} \mathbf{H}^A\mathbf{c}^A &= E^A\mathbf{c}^A \\ \mathbf{H}^B\mathbf{c}^B &= E^B\mathbf{c}^B \ . \end{aligned} \tag{IX.7}$$

One can introduce the direct-product space of dimension $m_A m_B$, referred to the basis functions $\varphi_I^{AB} = \varphi_{(ij)}^{AB} = \varphi_i^A\varphi_j^B$. If equations (IX.7) hold for the two sets $\{\varphi_i^A\}$, $\{\varphi_j^B\}$, then in the direct-product space φ_I^{AB} one also has an eigenvalue equation

$$\mathbf{H}^{AB}\mathbf{c}^{AB} = (E^A + E^B)\mathbf{c}^{AB} \ , \tag{IX.8}$$

where

$$\begin{aligned} \mathbf{H}^{AB} &= \mathbf{H}^A \otimes \mathbf{1}^B + \mathbf{1}^A \otimes \mathbf{H}^B \ ; \\ \mathbf{c}^{AB} &= \mathbf{c}^A \otimes \mathbf{c}^B \ , \end{aligned} \tag{IX.9}$$

with $\mathbf{1}^A$, $\mathbf{1}^B$ being the unit matrices of dimensions $m_A \times m_A$ and $m_B \times m_B$, respectively. In fact

$$\begin{aligned} \mathbf{H}^{AB}\mathbf{c}^{AB} &= (\mathbf{H}^A \otimes \mathbf{1}^B)(\mathbf{c}^A \otimes \mathbf{c}^B) + (\mathbf{1}^A \otimes \mathbf{H}^B)(\mathbf{c}^A \otimes \mathbf{c}^B) \\ &= \mathbf{H}^A\mathbf{c}^A \otimes \mathbf{1}^B\mathbf{c}^B + \mathbf{1}^A\mathbf{c}^A \otimes \mathbf{H}^B\mathbf{c}^B = (E^A + E^B)\mathbf{c}^A \otimes \mathbf{c}^B \ . \end{aligned} \tag{IX.10}$$

In the case of the nonorthogonal basis functions $\{\varphi_i^A\}$ and $\{\varphi_j^B\}$ one has the generalized equations

$$\begin{aligned} \mathbf{H}^A\mathbf{c}^A &= E^A\mathbf{S}^A\mathbf{c}^A \ ; \\ \mathbf{H}^B\mathbf{c}^B &= E^B\mathbf{S}^B\mathbf{c}^B \ . \end{aligned} \tag{IX.11}$$

As is easy to see, the overlap (metric) matrix in the direct-product space is

$$\mathbf{S}^{AB} = \mathbf{S}^A \otimes \mathbf{S}^B \tag{IX.12}$$

and the generalized eigenvalue equation becomes

$$\mathbf{H}^{AB}\mathbf{c}^{AB} = (E^A + E^B)\mathbf{S}^{AB}\mathbf{c}^{AB} \tag{IX.13}$$

with

$$\mathbf{H}^{AB} = \mathbf{H}^{A} \otimes \mathbf{S}^{B} + \mathbf{S}^{A} \otimes \mathbf{H}^{B} \ . \tag{IX.14}$$

Note

One may note that two quantum mechanical interactions are strictly additive if they correspond to situations described by Hamiltonian matrices (IX.9) or (IX.14). This is the case for supersystems consisting of noninteracting subsystems. In other cases one usually observes some deviation from a strict additivity. In some instances nonadditivity effects can be treated as a perturbation.

Appendix X
Permutations

A permutation $\hat{P}$ is an operation of changing the order of some objects; if they had the original sequence $1, 2, \ldots, N$ before the permutation, that will be changed to the sequence $P_1, P_2, \ldots P_n$ after it. (The set $\{P_1, P_2, \ldots, P_N\}$ is the same as $\{1, 2, \ldots, N\}$, but the elements are in a different order.) Each permutation $\hat{P}$ is defined by its own sequence $P_1, P_2, \ldots, P_N$. This means that, when applying $\hat{P}$, one has to put in the first place the element that was in the P_1-th place, in the second place the one that was in the P_2-th, and so on; in general, the kth place will be occupied by the element that originally was the P_kth. (Often a given order $P_1, P_2, \ldots, P_N$ of the elements $1, 2, \ldots, N$ is also called a permutation.)

It is well known that the number of different permutations of N objects (including the "identical permutation" $\hat{I}$, leaving the order unchanged) is $N! = 1 \cdot 2 \cdot 3 \ldots (N-1)N$. If we consider two permutations $\hat{P}$ and $\hat{Q}$—$\hat{P}$ is characterized by the sequence $\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_N$ and $\hat{Q}$ by the sequence $Q_1, Q_2, \ldots, Q_N$—then the "product" of permutations $\hat{P}$ and $\hat{Q}$ is a third permutation $\hat{R} = \hat{P}\hat{Q}$ that can be obtained by applying first the permutation $\hat{Q}$ and then the permutation $\hat{P}$. (In general $\hat{P}\hat{Q} \neq \hat{Q}P$.) Applying permutation $\hat{Q}$ we get the sequence $Q_1, Q_2, \ldots, Q_N$; when permutation $\hat{P}$ is applied to this sequence, then one has to put in the first place the P_1th element of this sequence, which is $Q_{(P_1)}$. In general, the element $Q_{(P_k)}$ goes to the kth place, i.e., the permutation $\hat{R} = \hat{P}\hat{Q}$ produces the sequence $R_1, R_2, \ldots, R_N = Q_{(P_1)}, Q_{(P_2)}, \ldots, Q_{(P_N)}$.

Further important properties of permutations that we utilize:

a) Each permutation $\hat{P}$ has a unique inverse permutation $\hat{P}^{-1}$, such as that their product is the identical permutation $\hat{I}$. Permutation $\hat{Q} = \hat{P}^{-1}$ is characterized by an order $Q_1, Q_2, \ldots, Q_N$, where Q_k is determined as the place in which element k can be found in the sequence $P_1, P_2, \ldots, P_N$. The inverses "from

left" and "from right" coincide, which can easily be seen by multiplying equality $\hat{Q}\hat{P} = \hat{I}$ by $\hat{P}$ from the left: $\hat{P}\hat{Q}\hat{P} = \hat{P}$, from which $\hat{P}\hat{Q} = \hat{I}$, as well.

b) If one applies all the $N!$ permutations to any starting sequence $Q_1, Q_2, \ldots, Q_N$, then one gets $N!$ different sequences; this means that the *set* of sequences obtained is independent of the starting sequence—only their order depends on the latter. Accordingly, the *set* of all permutations $\hat{R} = \hat{P}\hat{Q}$ for a given (fixed) $\hat{Q}$ is the same as the set of all $\hat{P}$s.

c) Each permutation is assigned a *parity*. The parity of $\hat{P}$ is +1 if one can get the sequence $P_1, P_2, \ldots, P_N$ by an even number of transpositions (interchanges of two elements) and -1 if the number of transpositions is odd. The parity of $\hat{P}$ can be denoted as $(-1)^p$, where p is the number of transpositions of $\hat{P}$. The parity of the product $\hat{R} = \hat{P}\hat{Q}$ is the product of the parities of $\hat{P}$ and $\hat{Q}$: $(-1)^r = (-1)^{p+q}$.

Appendix XI
An orthogonalization algorithm

There are cases in which one has got a vector in an $N+1$-dimensional space and wishes to build up an orthonormalized basis, requiring the given vector to be the first basis vector. Of course, the classical (Gramm–)Schmidt orthogonalization is applicable for that purpose. It is, however, costly and has the disadvantage that the different vectors are not treated in a symmetrical manner. In the given specific case one can use instead the following simple algorithm.

The task is to generate a full set of orthonormalized vectors orthogonal to some vector

$$\mathbf{q} = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ \cdots \\ q_N \end{pmatrix} \tag{XI.1}$$

We assume that vector $\mathbf{q}$ is normalized ($\sum_{i=0}^{N} |q_i|^2 = 1$), and define

$$\begin{aligned} cos\varphi &= q_0 \\ sin\varphi &= \left(\sum_{i=1}^{N} |q_i|^2 \right)^{1/2} \\ x_i &= \frac{q_i}{sin\varphi} \qquad (i = 1, 2, \ldots, N) \end{aligned} \tag{XI.2}$$

Then the coefficients of the new basis set can be combined into the unitary matrix

$$\begin{pmatrix} cos\varphi & -x_1 sin\varphi & -x_2 sin\varphi & \dots & -x_i sin\varphi & \dots \\ x_1 sin\varphi & 1+x_1^2(cos\varphi-1) & x_2x_1(cos\varphi-1) & \dots & x_ix_1(cos\varphi-1) & \dots \\ x_2 sin\varphi & x_1x_2(cos\varphi-1) & 1+x_2^2(cos\varphi-1) & \dots & x_ix_2(cos\varphi-1) & \dots \\ \dots & & \dots & \dots & \dots & \\ x_i sin\varphi & x_1x_i(cos\varphi-1) & x_2x_i(cos\varphi-1) & \dots & 1+x_i^2(cos\varphi-1) & \dots \\ \dots & & \dots & \dots & & \\ x_N sin\varphi & x_1x_N(cos\varphi-1) & x_2x_N(cos\varphi-1) & \dots & x_ix_N(cos\varphi-1) & \dots \end{pmatrix} \tag{XI.3}$$

It is easy to check that the first column of this matrix is identical to vector **q** as well as that all the columns are orthonormalized.

Index

1/Z expansion, 281
Accuracy of the variational energy, 21
Adiabatic approximation, 16
Adiabatic reactions, 16
Algebraic complement, 142–143, 150
Algorithm based on Brillouin theorem, 169
All singles CI method, 186
Analogy between differentials and variations, 303
Angular correlation, 258
Ansatz, 22, 263, 276
Antisymmetrization between distant subsystems, 135
Antisymmetrized product of strongly orthogonal geminals, 263
Antisymmetrizing operator, 129
  applied to spin-orbital indices, 133
  commutation properties of, 134
  projection character of, 130
APSG method, 263
Asymptotical behavior of wave function at large distances, 287
Atomcule, 9
Atomic units, 309
Aufbau principle, 185
Basis sets, 202
Bifurcation point of UHF method, 192
Biorthogonal functions, 320
Bond multiplicity, 235
Bond order, 235
Bond order index, 237, 248
  for open-shell system, 237
  invariance of, 237
  relations to exchange density, 243
Bond orders in three-center bonds, 247
Born–Oppenheimer approximation, 12, 15
Born–Oppenheimer separation, 13–16
Bra-ket formalism, 311
  analogy with matrix formalism, 313
Bra vectors, 311
Brillouin theorem, 166, 223
  algorithm based on, 169
Brillouin theorem and variation principle, 166
Brillouin theorem for a determinant with a stationary energy, 167
Brillouin theorem for the determinant with the absolute minimum of energy, 166
Brillouin theorem
  use of, 171, 187, 199
Brillouin–Wigner perturbation theory, 95
Canonical UHF equations, 176
CAS–SCF method, 263
CCA method, 271
Center of mass
  in classical mechanics, 295
  in quantum mechanics, 1, 6
Cholesky decomposition, 54
CI expansion of the many-electron wave function, 252
CI method, 260
CI–SD method, 262
  lack of size consistency, 262
Cofactor, 142
Column-vectors of orbital coefficients, 198
Commutation properties of antisymmetrizing operator, 134
Commutator, 33
Configuration interaction expansion, 252
Convergence of SCF calculations, 185
Correlational normalization, 72, 255
Correlation energy
  definition of, 254
Coulomb operator, 171
  Hermiticity of, 174
Coulson–Fischer wave function, 264
Coupled cluster approach, 269, 271
  size consistency of, 276
Coupling coefficients in the ROHF method, 210
  values of, 215

Coupling terms in Born–Oppenheimer approximation, 15
Covalent terms, 259
Critical point of UHF method, 191
Cusp condition
  electronic cusp, 286
  nuclear cusp, 283
Degenerate Rayleigh–Schrödinger perturbation theory, 93
ΔSCF method, 183
Diagonalization, 48
Diborane molecule, 247
Differential Hellmann–Feynman theorem, 26–27, 30, 32, 42
Different orbitals for different spins method, 188
Dirac's bra-ket formalism, 311
Direct CI method, 262
Direct product of matrices, 98, 325
Direct product space, 102
Dissociation catastrophe of RHF method, 190
DODS method, 188, 223
Doublet state, 215
Doubly excited determinants
  importance of, 255
Doubly filled orbitals, 187
Dynamical correlation, 254
Eckart's inequality, 23
EHF method, 194
Electron correlation, 251, 253, 258, 264
  and Nesbet's theorem, 253
  methods of calculation, 260
  two-electron example, 258
Electron density, 30, 228
  operator of, 228
Electronic Schrödinger equation, 13
Electrostatic force acting on nucleus, 30
Euler's theorem for homogeneous functions, 305
Exchange and chemical bonding, 243
Exchange energy and Wiberg index, 245
Exchange operator, 171
  Hermiticity of, 174
Exchange pair density, 242
Excitation energy, 186
Excited states, 24, 186
  singlet and triplet, 196
Exponential Ansatz in coupled cluster theory, 276
Expressions for perturbational energy corrections
  collection of formulae, 323
  using Wigner's $2n+1$ theorem, 88
Extended Hartree-Fock method, 194
Extended pairing theorem, 157
Factorization of the overlap determinant, 142
FCI method, 260
Finite basis Hartree–Fock theory, 198
Finite basis HF energy expression, 208
Fockian, 173
Fock matrix, 200
Fock operator, 173
  Hermiticity of, 175
  invariance of, 173
Forces acting on nuclei, 28
Free atoms, 9
Free motion of the center of mass, 3
Free valence index, 239, 248
  relations to spin density matrix, 239
Full CI method, 260
Gaussian product theorem, 287
Geminals, 263
Generalized Brillouin theorem, 168, 223
Generalized matrix eigenvalue problem, 47
  use of Löwdin's canonic orthogonalization, 66
  use of Löwdin's symmetric orthogonalization, 53
Generalized valence bond method, 263
Genuine UHF solution, 188
Gradient of the energy
  derivation by using Brillouin-type theorem, 217
  derivation by using Lagrangian multipliers, 221
Gross atomic population, 231
  invariance of, 234
Gross orbital population, 230
GVB method, 263
Hartree–Fock energy expression, 177, 184
Hartree–Fock equations, 170, 223
  derivation by Lagrangian multipliers, 177
  derivation by specified variations, 180
  derivation from Brillouin theorem, 171
Hartree–Fock method, 165
Hartree–Fock–Roothaan equations, 198, 200
  derivation of, 199
Hellman–Feynman forces, 217
Hellmann–Feynman theorem, 26, 40, 42
  conditions of validity, 30
  differential, 26–27, 30, 42
  integral, 31
  integrated, 31
Hermitization method of Kuprievich, 24–25, 206, 213, 224
HF equations, 170
HF method, 165
HFR energy expression, 208
HFR equations, 200, 223
  derivation of, 199
  formal introduction of, 201
HFR method, 198
HOMO, 185
Homogeneous functions
  Euler's theorem, 305
Homolytical dissociation, 190
Hylleraas functional, 88, 91
Hypervirial theorem, 316
Independent electron
  model of, 165
Independent Hamiltonians
  sum of, 3

Independent variables, 3
In–out correlation, 258
Integral Hellmann–Feynman theorem, 31–32
Interelectronic distances, 252, 254, 279
Intermediate normalization, 72, 255
Internal coordinates, 7
Interpair correlation, 257
Intrapair correlation, 256
Invariance of the Fock operator, 173
Ionic terms, 259
Ionization potential, 183
  vertical, 183
Jacobi coordinates, 6
Karadakov's extended pairing theorem, 157
Ket vectors, 311
Kinetic energy, 34, 307
  electronic, 13
  nuclear, 13
Koopmans theorem, 182, 223
  approximation to ionization potentials, 183
Lagrangian multipliers, 22, 25, 49, 177, 211, 221, 223
  problems connected with, 179, 223
Laplace theorem, 142
  generalized, 143
LCAO energy expression, 208
LCAO method, 198
Left–right correlation, 257
Level shift method, 185
Linear combination of atomic orbitals, 198
Linear independence of the basis, 62
  measure of, 62–63
Linear variational method, 45
Löwdin basis, 54
Löwdin orthogonalization, 52
  two-dimensional example, 59
Löwdin's canonic orthogonalization, 62, 64
  use in eigenvalue equations, 66
Löwdin's general formulae of matrix elements for nonorthogonal orbitals, 140
Löwdin's pairing theorem, 153
Löwdin's partitioning method, 115
Löwdin's spin projection operator, 194
Löwdin's symmetric orthogonalization, 52
  stationary property of, 55
  use in eigenvalue equations, 53
Löwdin's symmetry dilemma, 195
LUMO, 185
Many body perturbation theory, 264
Many-electron spin states, 124
  number of, 124, 126
Matrix eigenvalue equation, 47
  generalized, 47
  "standard", 47
Matrix eigenvalue problem, 47
  use of Löwdin's canonic orthogonalization, 66
  use of Löwdin transformation, 53
Matrix elements between determinant wave functions, 140, 145, 149
Matrix elements for nonorthogonal orbitals, 145, 149
Matrix of projection on the occupied subspace, 203
MBPT method, 264
MC–SCF methods, 263
Minors of overlap determinant, 142–143, 150
Mixing of the occupied orbitals, 138, 144
Møller–Plesset partitioning, 264
Møller–Plesset perturbation theory, 264
Mulliken's gross atomic population
  invariance of, 234
Mulliken's gross orbital population, 230
Mulliken's net atomic population, 232
  invariance of, 234
Mulliken's net orbital population, 231
Mulliken's overlap population, 232
  invariance of, 234
Mulliken's population analysis, 228, 247
  disadvantage of, 248
Nesbet's theorem, 256
Net atomic population, 232
  invariance of, 234
Net orbital population, 231
Nondegenerate Rayleigh–Schrödinger perturbation theory, 69
  algebraic expansion, 72
  compact matrix notations, 75
  use of reduced resolvent, 76
Normalization integral for Slater determinants
  orthogonal orbitals, 144
Nuclear Schrödinger equation, 14
Occupied orbital, 181–182
Occupied subspace, 183
One-electron operators
  matrix elements for nonorthogonal orbitals, 145
  matrix elements for orthogonal orbitals, 147
Open-shell singlet, 215
Operator of projection on the occupied subspace, 202
Orbital coefficients, 198
Orbital energies
  relation to total energy, 184
Orbital energy, 182
Ordered configuration, 252
Orthogonalization algorithm, 331
Overlap between Slater determinants, 140
  Löwdin's general formula for nonorthogonal orbitals, 141
  Slater rules for orthogonal orbitals, 144
Overlap density, 231
Overlap of determinant wave functions
  case of orthogonal orbitals, 144
  general case of nonorthogonal orbitals, 141
Overlap population, 231–232
  invariance of, 234

Pair correlation energy, 256
Pairing theorem, 153
  extended, 157
Pauli matrices, 122
Permutations
  properties of, 329
Perturbational theory, 264
PES, 12
P-matrix, 202
  idempotency property $\mathbf{PSP} = \mathbf{P}$, 203
Point-like discontinuity, 189
Population analysis, 228
Potential energy, 13, 34
Potential energy (hyper-)surface, 12
Projected Hamiltonian, 25, 260
Projection operators, 316
Projectors in the LCAO framework, 204
Pure singlet character of RHF wave function, 192
Radial correlation, 258
Rayleigh quotient, 19
Rayleigh–Schrödinger perturbation theory
  algebraic expansion, 72
  compact matrix notations, 75
  degenerate, 93
  nondegenerate, 69
  size consistency of, 100, 102
  use of reduced resolvent, 76
Reduced mass, 2, 11
Reduced resolvent, 76
  invariance of, 80
Reference state, 255
Relative coordinates, 7
Relative motion, 3
Resolution of identity, 318
Restricted Hartree–Fock method, 187
Restricted open-shell Hartree–Fock method, 209
RHF energy expression, 209
RHF equations, 188
RHF method, 187
Ritz method, 45
RMP2 energy formula, 268
RMP method, 268
ROHF energy expression, 210
ROHF equations, 211, 224
ROHF method, 209
Rydberg constant, 5
$S^{-1/2}$ transformation, 53
Saddle point, 188
Scaling, 35–36
  connection with virial theorem, 36
SCF, 165
SCF algorithm
  necessarily convergent, 170
SCF calculations
  convergence of, 185
SCF procedure, 176
Secular equation, 47
Self-consistent field, 165
Self-repulsion term, 173, 185–186, 197, 205, 220, 244, 268
Similarity transformation, 48
Singlet and triplet excitations from RHF ground state, 196
Singlet excited state, 196
Singlet instability, 189
Size consistency, 99
Size consistency of the Rayleigh–Schrödinger perturbation theory
  formal considerations, 100
  perturbational expansions, 102
  proof of, 102
Size–consistency problem, 98
Size extensivity, 99
Slater determinant
  definition of, 126
  expressed by the antisymmetrizing operator, 129
  invariance properties of, 138, 144
  matrix elements, 140, 145, 149
  two-electron examples, 128
Slater rules, 140
  one-electron operators, 147–148
  overlap, 144
  two-electron operators, 151, 153
Spectral resolution, 319
Spin contamination, 193
Spin density, 230
Spin density matrix, 230
Spin functions, 122
  orthogonality of, 123
Spinless density matrix, 229
Spin operators, 122
Spin orbitals, 121
Spin projected extended Hartree–Fock method, 194, 223
Spin projection operator, 194, 223
Static correlation, 254
Strong orthogonality, 263
Subspace of occupied orbitals, 140
Summation over spin variables, 123
Sum of independent Hamiltonians, 3
Symmetric orthogonalization, 52
Symmetry adapted RHF wave function, 189
Symmetry and the RHF method, 189
Symmetry dilemma, 195, 223
Symmetry-restricted HF function, 189
Symmetry-unrestricted HF function, 189
Theorem about orbitals of specific structure, 160
  existence theorem, 161
  a posteriori determination, 162
Three center—four-electron bond, 247
Three center—two electron bond, 247
Time dependence of a physical quantity, 32
Trace of projector, 321
Triplet excited state, 196

Triplet instability, 188
Triplet state, 215
Two-body problem
  in classical mechanics, 299
  in quantum mechanics, 1
Two-electron operators
  matrix elements for nonorthogonal orbitals, 149
  matrix elements for orthogonal orbitals, 151–152
UHF energy expression, 177
UHF equations, 171, 173
  canonical form of, 176
UHF equations for successive optimization of orbitals, 204
UHF method, 192, 223
UMP2 energy formula, 267
UMP method, 267
Unitary transformation of occupied orbitals, 139
Unrestricted Hartree–Fock equations, 171
Valence index, 238, 248
  relations to bond order indices, 238
Variational energy
  accuracy of, 21
Variational–perturbational method, 88
Variational principle for excited states, 42
Variation principle, 19
  as equivalent of the Schrödinger equation, 21
  for determinant wave functions, 166
  for excited states, 24
  for the ground state, 19
  useful form of, 21
Variations
  analogy with differentials, 303
Vertical ionization potential, 183
Virial
  invariance of, 308
Virial of external forces, 37
Virial ratio, 41
Virial theorem
  and chemical bonding, 41
  connection with scaling, 36
  for Coulombic systems, 34
  in classical mechanics, 307
  in quantum mechanics, 32–34
  in the Born–Oppenheimer approximation, 36, 40
Virtual orbitals, 184, 257
Wave function forces, 31, 217
Wiberg index, 236
  for open-shell system, 236
Wigner's $2n + 1$ theorem, 75, 81, 87
  accuracy of the expectation value, 81
  calculating higher order energy corrections, 87