SPRINGER REFERENCE

Wenjian Liu
Editor

Handbook of Relativistic Quantum Chemistry

Springer

Handbook of Relativistic Quantum Chemistry

Wenjian Liu
Editor

Handbook of Relativistic Quantum Chemistry

With 150 Figures and 87 Tables

Springer

Editor
Wenjian Liu
Beijing National Laboratory for Molecular Sciences
Institute of Theoretical and Computational Chemistry
State Key Laboratory of Rare Earth Materials Chemistry
and Applications
College of Chemistry and Molecular Engineering, and
Center for Computational Science and Engineering
Peking University
Beijing, People's Republic of China

ISBN: 978-3-642-40765-9 ISBN: 978-3-642-40766-6 (eBook)
ISBN: 978-3-642-40767-3 (print and electronic bundle)
DOI 10.1007/978-3-642-40766-6

Library of Congress Control Number: 2016950439

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Germany
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface

Just two years after Schrödinger proposed in 1926 the (nonrelativistic) quantum mechanical wave equation of electrons, Dirac established in 1928 the relativistic counterpart by first quantizing the special theory of relativity set up by Einstein in 1905. Any difference between the Dirac and Schrödinger equations of electrons was then called "relativistic effect." However, just one year later, that is, in 1929, Dirac himself stated that "relativistic effects are of no importance in the consideration of atomic and molecular structure and ordinary chemical reactions." The underlying assumptions were that the average speeds of valence electrons are very low compared with that of light, such that relativistic effects are very small for valence electrons, and that relativistic effects are indeed important for core electrons but are canceled out for valence properties. Such wrong assumptions were taken naively for granted for nearly half a century, until the mid-1970s, when relativistic effects were found to be indeed very important for electronic structure, sometimes even of light atoms. Since then, relativistic quantum chemistry witnessed fast progresses, including deep understandings of not only relativistic effects but also novel approximate two-component (A2C) relativistic theories such as second-order DKH (Douglass-Kroll-Hess), ZORA (zeroth-order regular approximation), and DPT (direct perturbation theory) developed between 1985 and 1995. The beginning of the new millennium turned out to be also the beginning of a new era of relativistic quantum chemistry, as symbolized by the exact two-component (X2C) relativistic theory introduced and made mature between 2005 and 2010. Undoubtedly, X2C is going to be the new workhorse of relativistic quantum chemistry. Notwithstanding this achievement, the no-pair approximation underlying both A2C and X2C theories has some fundamental defects. How to go beyond this approximation yet without diving into full quantum electrodynamics (QED) is then the final high point of relativistic quantum chemistry. After several unsuccessful tries, a proper effective QED (eQED) approach was finally obtained in a bottom-up fashion, i.e., without recourse to QED itself at all. This eQED paves a seamless bridge between relativistic quantum chemistry and full QED, which used to be two mutually exclusive subfields of relativistic molecular quantum mechanics. In the subfield of QED, both new formulations and applications have been achieved in the last decade. Especially, effective means to combine QED with many-body theory have been designed to achieve unprecedented high accuracy in spectroscopic calculations.

All the aforementioned fundamental developments in relativistic molecular quantum mechanics, the union of relativistic quantum chemistry and QED, are fully covered by this book but otherwise not covered by any other books. To facilitate understandings of such methodological developments, sufficient pedagogic introductions are also provided. Therefore, the book should be useful for both users and developers of relativistic quantum mechanical methods and tools. It is somewhat unfortunate that a collection of representative applications of such methods, although planned, was not yet accomplished but which will be included in a new edition of the book.

The editor is very grateful to the section editors for their efforts in identifying excellent chapter writers. The editorial team wants to acknowledge the chapter writers for their great contributions and also the staff of Springer, including June, Neha, and Stephen, whose patience and professional copyedit were of great help to materialize the book.

Beijing Wenjian Liu
September 2016

Contents

Part I Introduction to Relativistic Quantum Chemistry **1**
Christoph van Wüllen

1 Dirac Operator and Its Properties 3
Jacek Karwowski

2 Nuclear Charge Density and Magnetization Distributions 51
Dirk Andrae

3 One-Particle Basis Sets for Relativistic Calculations 83
Kenneth G. Dyall

4 Relativistic Self-Consistent Fields 107
Christoph van Wüllen

Part II Introduction to Quantum Electrodynamics **129**
Paul Indelicato

5 Introduction to Bound-State Quantum Electrodynamics 131
Paul Indelicato and Peter J. Mohr

6 QED Effects and Challenges 243
Anton N. Artemyev

7 Effective QED Hamiltonians 267
Anton N. Artemyev

8 Two-Time Greens Function Method 287
Anton N. Artemyev

9 Unifying Many-Body Perturbation Theory with Quantum Electrodynamics 313
Ingvar Lindgren and Paul Indelicato

Part III Relativistic Hamiltonians **343**
Wenjian Liu

10 With-Pair Relativistic Hamiltonians 345
Wenjian Liu

11 No-Pair Relativistic Hamiltonians: Q4C and X2C 375
Wenjian Liu

12 Sequential Decoupling of Negative-Energy States in Douglas–Kroll–Hess Theory 395
Markus Reiher

13 Spin Separation of Relativistic Hamiltonians 411
Zhendong Li and Wenjian Liu

14 Relativistic Effective Core Potentials 449
Michael Dolg

Part IV Relativistic Wave Functions and Density Functionals 479
Wenjian Liu

15 Basic Structures of Relativistic Wave Functions 481
Sihong Shao, Zhendong Li, and Wenjian Liu

16 Coalescence Conditions of Relativistic Wave Functions 497
Sihong Shao, Zhendong Li, and Wenjian Liu

17 Relativistic Explicit Correlation: Problems and Solutions 531
Wenjian Liu, Sihong Shao, and Zhendong Li

18 Relativistic Density Functional Theory 547
Eberhard Engel

Part V Relativistic Quantum Chemical Methods and Applications 579
Jochen Autschbach and Jun Li

19 Relativistic Many-Body Aspects of the Electron Electric Dipole Moment Searches Using Molecules 581
Bhanu P. Das, Malaya Kumar Nayak, Minori Abe, and V. S. Prasannaa

20 Relativistic Calculations of Atomic Clock 611
Bijaya Kumar Sahoo

21 Relativistic Theories of NMR Shielding 657
Yunlong Xiao, Wenjian Liu, and Jochen Autschbach

22 Relativistic Theory of Nuclear Spin-Rotation Tensor 693
Yunlong Xiao, Wenjian Liu, and Kenneth Ruud

23 Relativistic Methods for Calculating Electron Paramagnetic Resonance (EPR) Parameters 725
Hélène Bolvin and Jochen Autschbach

24 Zero-Field Splitting in Transition Metal Complexes: Ab Initio Calculations, Effective Hamiltonians, Model Hamiltonians, and Crystal-Field Models 765
Rémi Maurice, Ria Broer, Nathalie Guihéry, and Coen de Graaf

25 Relativistic Equation-of-Motion Coupled-Cluster Theory (EOM-CC) 797
Fan Wang

26 High-Accuracy Relativistic Coupled-Cluster Calculations for the Heaviest Elements 825
Ephraim Eliav, Anastasia Borschevsky, and Uzi Kaldor

27 Relativistic Quantum Chemistry for Chemical Identification of the Superheavy Elements 857
Valeria Pershina

Index 901

About the Editor

Wenjian Liu is a Cheung Kong Professor of Theoretical Chemistry at Peking University (since 2001) and an elected member of the International Academy of Quantum Molecular Science (2014). After receiving his PhD at Peking University in 1995, he spent 6 years as a postdoctoral researcher at Max Planck Institute Dresden and Ruhr University Bochum, Germany. His research interests focus on the development of relativistic quantum mechanical theories and methods for molecular electronic structure and magnetic/electric properties. Besides strong publication and citation records, he has gained several worldwide distinctions, including the Friedrich Wilhelm Bessel Research Award of the Alexander von Humboldt Foundation (2007), the Annual Medal of International Academy of Quantum Molecular Science (2006), and the Pople Medal of Asia-Pacific Association of Theoretical and Computational Chemists (2006).

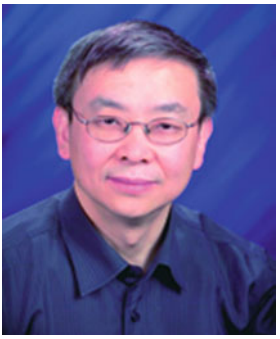
Section Editors

Christoph van Wüllen
TU Kaiserslautern, FB Chemie
Erwin-Schrödinger-Str.
D-67663 Kaiserslautern, Germany
Email: vanwullen@chemie.uni-kl.de

Paul Indelicato
Laboratoire Kastler Brossel
Université Pierre et Marie Curie and CNRS
Boîte 74, 4 Place Jussieu, F-75252
Cedex 05
Paris, France
Email: paul.indelicato@lkb.upmc.fr

Wenjian Liu
Beijing National Laboratory for Molecular Sciences
Institute of Theoretical and Computational Chemistry
State Key Laboratory of Rare Earth Materials
Chemistry and Applications
College of Chemistry and Molecular Engineering, and
Center for Computational Science and Engineering
Peking University
Yiheyuan Road 5, Haidian District
100871 Beijing
People's Republic of China
Email: liuwj@pku.edu.cn; liuwjbdf@gmail.com

Jochen Autschbach
Department of Chemistry
University at Buffalo
State University of New York
312 Natural Sciences Complex
Buffalo, NY 14260-3000, USA
Email: jochena@buffalo.edu



Jun Li
Department of Chemistry
Tsinghua University
100084 Beijing
P. R. China
Email: junli@mail.tsinghua.edu.cn

Contributors

Minori Abe Department of Chemistry, Tokyo Metropolitan University, Hachioji-city, Tokyo, Japan

Dirk Andrae Institut für Chemie und Biochemie, Freie Universität Berlin, Berlin, Germany

Anton N. Artemyev Institute of Physics, Kassel University, Kassel, Germany

Jochen Autschbach Department of Chemistry, University at Buffalo, State University of New York, Buffalo, NY, USA

Hélène Bolvin Laboratoire de Physique et de Chimie Quantiques, Université Toulouse 3, Toulouse, France

Anastasia Borschevsky Van Swinderen Institute for Particle Physics and Gravity, University of Groningen, AG Groningen, The Netherlands

Ria Broer Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands

Bhanu P. Das Theoretical Physics and Astrophysics Group, Indian Institute of Astrophysics, Bangalore, India

Michael Dolg Institute for Theoretical Chemistry, University of Cologne, Cologne, Germany

Kenneth G. Dyall Dirac Solutions, Portland, OR, USA

Ephraim Eliav School of Chemistry, Tel Aviv University, Tel Aviv, Israel

Eberhard Engel Center for Scientific Computing, J.W. Goethe-Universität Frankfurt, Frankfurt/Main, Germany

Coen de Graaf Departament de Química Física I Inorgánica, Universitat Rovira i Virgili, Tarragona, Spain

Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Nathalie Guihéry Laboratoire de Chimie et Physique Quantiques, IRSAMC/UMR5626, Université de Toulouse 3, Toulouse, France

Paul Indelicato Laboratoire Kastler Brossel, UPMC-Sorbonne Universités, Collège de France and CNRS, Paris, France

PSL Research University, Paris, France

Uzi Kaldor School of Chemistry, Tel Aviv University, Tel Aviv, Israel

Jacek Karwowski Institute of Physics, Nicolaus Copernicus University, Toruń, Poland

Zhendong Li Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China

Ingvar Lindgren Department of Physics, University of Gothenburg, Gothenburg, Sweden

Wenjian Liu Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China

Rémi Maurice SUBATECH, UMR CNRS 6457, IN2P3/EMN Nantes/Université de Nantes, Nantes, France

Peter J. Mohr National Institute of Standards and Technology, Gaithersburg, MD, USA

Malaya Kumar Nayak Bhabha Atomic Research Centre, Trombay, Mumbai, India

Valeria Pershina GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany

V. S. Prasannaa JST, CREST, Kawaguchi, Saitama, Japan

Markus Reiher Laboratorium für Physikalische Chemie, ETH Zürich, Zürich, Switzerland

Kenneth Ruud Department of Chemistry, Centre for Theoretical and Computational Chemistry, University of Tromsø—The Arctic University of Norway, Tromsø, Norway

Bijaya Kumar Sahoo Theoretical Physics Division, Physical Research Laboratory (PRL), Ahmedabad, Gujarat, India

Sihong Shao LMAM and School of Mathematical Sciences, Peking University, Beijing, China

Christoph van Wüllen Fachbereich Chemie, Technische Universität Kaiserslautern, Kaiserslautern, Germany

Fan Wang Institute of Atomic and Molecular Physics, Sichuan University, Chengdu, China

Yunlong Xiao Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China

Part I

Introduction to Relativistic Quantum Chemistry

Christoph van Wüllen

Dirac Operator and Its Properties

1

Jacek Karwowski

Contents

Introduction . . . 4
Equations of Motion . . . 6
Schrödinger Equation . . . 8
Dirac Equation . . . 9
Representations in the Spinor Space . . . 13
Relativistic Covariance of the Dirac Equation . . . 13
Gamma Matrices . . . 14
Clifford Algebra . . . 14
Covariant Form of the Dirac Equation . . . 15
Transformation Properties of the Dirac Spinor . . . 16
Eigenvalue Problem of the Free Dirac Operator . . . 20
Constants of Motion . . . 21
Free-Particle Spinors . . . 22
Position and Velocity Operators, Zitterbewegung . . . 22
External Fields . . . 25
Dirac Equation for a Particle in an External Field . . . 25
Gauge Invariance . . . 27
The Non-relativistic Limit and the Direct Approach . . . 28
Elimination of the Small Component . . . 30
Gordon Decomposition . . . 32
Charge Conjugation, Positrons . . . 32
Exactly Solvable Models . . . 34
Harmonic Oscillators . . . 34
Central Force . . . 37

J. Karwowski (✉)
Institute of Physics, Nicolaus Copernicus University, Toruń, Poland
e-mail: jka@fizyka.umk.pl

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_22

Two Dirac Particles 40
Non-interacting Particles in an Electrostatic Field 42
Two Interacting Particles 46
Summary 47
References 48

Abstract

The formulation of a theory which unifies quantum mechanics and special theory of relativity, performed by Dirac nearly a century ago, required introduction of new mathematical and physical concepts which led to models which, on one hand, are very successful in terms of interpretation of the physical reality but, on the other, still create some challenge, both conceptual and computational. A central notion of relativistic quantum mechanics is a construct known as the *Dirac operator*. It may be defined as the result of factorization of a second-order differential operator in the Minkowski space. The result of this factorization, in the case of the relativistic model of a single electron, gives the relativistic Dirac Hamiltonian. The spectrum of this Hamiltonian describes the energy of the electron, but it is unbounded from below. This weird property leads to many computational problems, artifacts and misunderstandings in the interpretation of the results derived from the Dirac model. This chapter contains a brief and self-contained description of the basic properties of the Dirac operator.

Keywords

Brown-Ravenhall disease • Clifford algebra • Complex coordinate rotation • Dirac eigenvalue problem • Dirac equation • Essential spectrum • Exactly-solvable models • Lorentz covariance • Minimax principle • Non-relativistic limit • Pauli Hamiltonian • Spinor

Introduction

The second half of the third decade of the twentieth century brought one of the most important discoveries in the history of science. At the end of 1925 Werner Heisenberg constructed the matrix form of quantum mechanics. Several months later Erwin Schrödinger proposed his wave-equation-based formulation. Finally, in 1928 Paul Adrien Maurice Dirac performed a successful unification of quantum mechanics and special theory of relativity. The formulation of the equation which opens the way to this unification, referred to as the Dirac equation, required the introduction of a new mathematical construct known as the Dirac operator. In the simplest way the Dirac operator may be defined as the result of factorization of a second-order differential operator in the Minkowski space.

In this chapter the properties of the Dirac operator relevant to quantum chemical applications are briefly described. The subject is covered from a variety of perspectives in many textbooks and monographs, and many new directions of development are still under construction. The selection of topics addressed in this chapter has been motivated by their importance for the understanding of basic ideas related

to the properties of the Dirac operator and by the foreseen perspectives of further development of the theory. The author did his best to be objective in his choice, but he is aware that the selection is biased by his personal preferences. The author of this section tried to cite mainly textbooks, monographs, and review papers. Only exceptionally, when the subject has not been adequately covered in a review, the original works have been cited.

For the basic understanding of some elementary problems related to the subject of this chapter, the classical textbooks by Davydov [1], Messiah [2], Itzykson and Zuber [3], and, in particular, Sakurai [4] are recommended. A broad pedagogical presentation with illustrative solutions of many interesting problems has been given by Greiner [5]. Modern, and more advanced, presentations may be found, in textbooks by Landau [6], Pilkuhn [7], Das [8], Schwabl [9], and Scadron [10], listed in an increasingly demanding sequence, as far as the formal background of the reader is concerned. A particular position among books concerned with the Dirac operator is occupied by the monograph by Thaller [11]. It has been written for theoretical physicists and mathematicians, and, consequently, it does not offer an easy reading for a theoretical chemist. However, to the author's knowledge, it is the most complete and rigorous presentation of the issues related to the one-particle Dirac problem available in the literature.

In principle, the concept of the Dirac operator is restricted to the relativistic quantum model of a single fermion. However, the subject is of chemical interest because of its many-particle, in fact many-electron, extensions. Four recent monographs cover different aspects of this direction of the development. The one by Reiher and Wolf [12] offers probably the most complete summary of different aspects of relativistic quantum chemistry. In particular, a list of 975 carefully selected references is most helpful for a reader willing to study this field from different perspectives. Another monograph, by Dyall and Fægri [13], covers various aspects of modern methods of relativistic quantum chemistry. Computational methodology is presented in detail in a monograph by Grant [14]. Finally, the one by Lindgren [15] shows the present status of the developments aimed at combining quantum electrodynamics and relativistic quantum mechanics into one computationally manageable theory. Very recently new concepts directed towards the construction of a bridge between theories based on relativistic theory of many-electron systems and quantum electrodynamics have been developed by Lindgren [16], Liu and Lindgren [17], and Liu [18]. Section *Relativistic Hamiltonians* of this handbook gives a broad description of the most important constructions based on the concept of the Dirac operator, relevant in the theory of many-electron systems.

In the next section two derivations of the Dirac equation are presented, and relations between spin-dependent Schrödinger equation and the Dirac equation are discussed. The following section is concerned with some formal features of the Dirac formalism, in particular with the transformation properties of the Dirac spinors and with relations between the Dirac matrices and the Clifford algebra. The eigenvalue problem of the free Dirac operator and some unusual features of the Dirac formalism related to the definition of the position and velocity of a Dirac particle are discussed in the next section. Properties of the Dirac operator describing an electron in external fields, with particular emphasis put on the bound

properties of the energy spectrum, are the subject of section "External Fields." In the same section links between equations describing electrons and positrons are briefly discussed. Two important exactly solvable models, namely, a harmonic oscillator and an electron in a Coulomb field, are the subject of section "Exactly Solvable Models." The next section is devoted to the problem of two Dirac particles. Finally, in the last section brief remarks about not fully resolved questions are given.

The following notation conventions are used: sans serif symbols denote operators (e.g., H, p_k) and quantum numbers (e.g., n, j, l) while the standard ones – the corresponding classical quantities (H, p_k) and integer indices (e.g., n, j, l). Unit matrix $n \times n$ is denoted $\mathbf{1}_n$. In most cases its dimension can easily be deduced from the context, and then it is not given explicitly. Symbol $\mathbf{1}$ also stands for the unit operator. Vectors or tensors in the Galilean three-dimensional (3D) space are denoted by boldface symbols and their components by the standard symbols with Latin indices. Thus, for example, the position and the momentum vectors in the 3D space are, respectively, $\boldsymbol{r} = \{x_1, x_2, x_3\}$ and $\boldsymbol{p} = \{p_1, p_2, p_3\}$. In the four-dimensional (4D) Minkowski space-time, the Cartesian metric is used, i.e., the metric tensor is defined as 4×4 unit matrix and, consequently, there is no distinction between covariant and contravariant coordinates of a vector. The Greek indices identify components of vectors and tensors (e.g., x_μ, p_μ, $\mu = 1, 2, 3, 4$). The coordinate and the momentum vectors are defined, respectively, as

$$x_\mu\big|_{\mu=1}^{4} = \{x_1, x_2, x_3, x_4\} = \{\boldsymbol{r}, ict\},$$
$$p_\mu\big|_{\mu=1}^{4} = \{p_1, p_2, p_3, p_4\} = \{\boldsymbol{p}, iE/c\}, \tag{1}$$

where c, t, and E denote, respectively, velocity of light, time, and energy. Implicit summation over repeated Greek indices is assumed. Note that some authors use a pseudo-Cartesian metric with the metric tensor $\mathbf{g} = \mathrm{diag}\{1, -1, -1, -1\}$. In this case one has to distinguish between covariant and contravariant coordinates. In particular, $x_\mu = g_{\mu\nu}\, x^\nu = \{x_0, \boldsymbol{r}\}$ with $x_0 = ct$.

Hartree atomic units are used, i.e., the reduced Planck constant $\hbar = h/2\pi = 1$, the elementary charge $e = 1$, and the rest mass of electron $m_\mathrm{e} = 1$. Thus, the charge of electron is equal to $q_\mathrm{electron} = -e = -1$ and the charge of proton is $q_\mathrm{proton} = e = +1$. Velocity of light $c \approx 137.035\,9895$, fine structure constant $\alpha = e^2/\hbar c$, and energy unit $E_\mathrm{h} = \alpha^2 m_\mathrm{e} c^2 = m_\mathrm{e}\, e^4/\hbar^2 = 1$ hartree. In some cases (in particular in the next section), for convenience of less experienced readers, some constants are given explicitly. Also, whenever it is convenient, the mass m and the charge q are used.

Equations of Motion

The time evolution equation of a closed quantum system in the coordinate representation and in the Schrödinger picture reads [2]

$$\left[i\hbar \frac{\partial}{\partial t} - \mathsf{H}(\boldsymbol{r}) \right] \psi(\boldsymbol{r}, t) = 0, \tag{2}$$

where H is the Hamilton operator (Hamiltonian) of the system and the remaining symbols have their usual meaning. Depending on the physical context and on the form of the Hamiltonian, this equation may be referred to as the Schrödinger equation, Dirac equation, Weyl equation, causal equation of the matter field PSI, etc. Quantum equations may be constructed using the correspondence principle. This can be accomplished by the replacement of the classical quantities by the appropriate operators. In particular for energy E and momentum $\boldsymbol{p}$,

$$E \;\mapsto\; \mathsf{E} = i\hbar\frac{\partial}{\partial t}, \quad p_k \;\mapsto\; \mathsf{p}_k = -i\hbar\frac{\partial}{\partial x_k}, \quad k = 1, 2, 3. \tag{3}$$

Since the Hamiltonian corresponds to the energy of the system, solutions of its eigenvalue problem give stationary state wave functions $\phi_E(\boldsymbol{r})$ and the corresponding energies E:

$$\mathsf{H}\,\phi_E(\boldsymbol{r}) = E\,\phi_E(\boldsymbol{r}). \tag{4}$$

From here one can derive solutions of Eq. (2) as

$$\psi(\boldsymbol{r}, t) = \sum\!\!\!\!\!\!\int_E \mathrm{e}^{-iEt/\hbar}\, c(E)\, \phi_E(\boldsymbol{r})\, dE. \tag{5}$$

Let us assume that the wave function $\psi(\boldsymbol{r}, t)$ fulfills an equation $\Omega\,\psi = 0$ and $\mathbb{S}$ is an operator transforming variables of this equation. Then the transformed equation reads $\Omega'\psi' = 0$, where $\Omega' = \mathbb{S}\Omega\mathbb{S}^{-1}$ and $\psi' = \mathbb{S}\psi$. The results derived from a model described by this equation are independent of the transformation if $\mathbb{S}\Omega\mathbb{S}^{-1} = \Omega$, i.e., if $[\mathbb{S}, \Omega] = 0$. In such a case $\mathbb{S}$ is called *a symmetry transformation*. The most universal symmetry is related to transformations between two inertial (i.e., moving with a constant velocity relative to each other) reference frames. This symmetry implies that equations of the motion must retain the same form in all inertial reference frames. This property is described as *covariance* of the equations. The group of transformations which provide equivalent description of physical phenomena in all inertial frames is referred to as the *covariance group*. The non-relativistic theories are covariant with respect to the Galilei group. All inertial frames share the same universal time. Positions of the particles are determined by three coordinates forming three-component position vectors $\boldsymbol{r} = \{x_1, x_2, x_3\}$ which transform between two inertial frames according to the Galilei transformation.

However, the physical reality is believed to be invariant with respect to the Lorentz transformation. The theories which are Lorentz covariant are referred to as *relativistic*. In the relativistic theories three coordinates and time, momentum and energy, etc. form four-vectors (1). The components of a four-vector transform between two inertial frames (the "unprimed" one and the "primed" one) according to the Lorentz transformation:

$$x'_\mu = a_{\mu\nu}\, x_\nu, \tag{6}$$

where $a_{\mu\rho}a_{\nu\rho} = \delta_{\mu\nu}$, $\left(a^{-1}\right)_{\mu\nu} = a_{\nu\mu}$, i.e., $\boldsymbol{a}$ is an orthogonal 4×4 matrix. The general Lorentz transformation can be built as a composition of an ordinary rotation $\boldsymbol{\omega} = \omega\,\mathbf{n}$ by a real angle ω around the axis defined by the unit vector $\mathbf{n}$ in the 3D subspace of the 4D Minkowski space, and a special Lorentz transformation (also referred to as the Lorentz boost) which may be expressed as a rotation by an imaginary angle $\overline{\omega} = i\,\chi = i\,\mathrm{arctanh}(v/c)$ around the axis defined by $\mathbf{v}/v$ where $\mathbf{v}$ is the velocity of the primed reference frame relative to the unprimed one.

A requirement of the covariance with respect to either Galilei or Lorentz transformations imposes strong restrictions on the structure of a theory and on the form of the basic equations. In a non-relativistic theory, three coordinates of a point $\boldsymbol{r}$, momentum $\boldsymbol{p}$, and vector potential $\boldsymbol{A}$ are three-component vectors, while time t, energy E, and electrostatic potential φ are invariants (scalars). Therefore, the Schrödinger equation is invariant with respect to the Galilei transformation if the Hamiltonian is Galilei invariant. In relativistic theories not only $\boldsymbol{r}$ and $x_4 = ict$ form a four-component vector. Also momentum $\boldsymbol{p}$ and energy $p_4 = iE/c$, current $\boldsymbol{j}$ and density $j_4 = ic\rho$, vector potential $\mathbf{A}$, and electrostatic potential $A_4 = i\,\varphi/c$ form four-vectors. The covariance conditions are here more complicated. In particular, the transformation properties of the wave function depend on the spin of the particles. These topics are discussed in section "Relativistic Covariance of the Dirac Equation" of this chapter.

Conclusion: *In a Lorentz-covariant model, the space coordinates of a particle and the time corresponding to this particle have to appear on an equal footing.* In particular, if an equation is of the first order in time, it must be of the first order in all coordinates. In the general equation of motion (2), time plays a special role. One can say that this equation is written in a non-covariant form. It may be either Galilei or Lorentz covariant. Its possible Lorentz covariance is hidden and may be discovered only after analyzing simultaneously the form of the Hamiltonian and the structure of the equation.

Schrödinger Equation

The free-particle Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi(\boldsymbol{r},t) = \mathsf{H}_0^{\mathrm{S}}\psi(\boldsymbol{r},t), \tag{7}$$

with

$$\mathsf{H}_0^{\mathrm{S}} = \frac{\mathbf{p}^2}{2m} = -\frac{\hbar^2}{2m}\triangle \tag{8}$$

describes quantum dynamics of non-relativistic particles and opens a way to the formulation of the non-relativistic quantum mechanics. The equation is scalar; it is defined in

$$\mathscr{H}_1^{\mathrm{S}} = L^2\left(\mathbb{R}^3\right) \tag{9}$$

Hilbert space and is invariant with respect to the Galilei transformation. Since time and coordinate derivatives are of different orders, it cannot be Lorentz covariant.

In order to describe spin-1/2 particles within Galilei-covariant formalism, one should use, instead of Eq. (7), its analog defined in a larger Hilbert space,

$$\mathscr{H}_1^{\mathrm{P}} = \mathscr{H}_1^{\mathrm{S}} \oplus \mathscr{H}_1^{\mathrm{S}} = L^2\left(\mathbb{R}^3\right) \otimes \mathbb{C}^2 \tag{10}$$

with

$$\mathsf{H}_0^{\mathrm{P}} = \frac{(\boldsymbol{\sigma} \cdot \mathbf{p})^2}{2m}, \tag{11}$$

where the superscript P stands for *Pauli* and $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \sigma_3\}$ are Pauli spin 2×2 matrices. They are defined by the following relations:

$$\sigma_k^2 = \mathbf{1}, \quad \sigma_k\sigma_l + \sigma_l\sigma_k = 0 \ \text{ if } \ k \neq l, \quad \sigma_1\sigma_2 - \sigma_2\sigma_1 = 2i\sigma_3 \ \text{ with } \ \text{cyclic}\{1,2,3\}. \tag{12}$$

These relations are invariant with respect to unitary transformations of the matrices. Therefore, if they are fulfilled by one set of matrices, then they are also fulfilled by all sets related to this one by a unitary transformation. In particular, in the Pauli representation

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{13}$$

Pauli matrices fulfill a very useful identity known as the *Dirac relation* [12]:

$$(\boldsymbol{\sigma} \cdot \mathbf{A})\,(\boldsymbol{\sigma} \cdot \mathbf{B}) = (\mathbf{A} \cdot \mathbf{B}) + i\boldsymbol{\sigma} \cdot [\mathbf{A} \times \mathbf{B}], \tag{14}$$

where **A** and **B** are operators. Consequently,

$$\mathsf{H}_0^{\mathrm{P}} = \frac{\mathbf{p}^2}{2m}\,\mathbf{1}_2. \tag{15}$$

In this case ψ is not a scalar but a two-component Pauli spinor, i.e.,

$$\psi = \begin{bmatrix} \psi_\alpha \\ \psi_\beta \end{bmatrix}, \tag{16}$$

where ψ_α and ψ_β correspond to two projections of spin.

Dirac Equation

Relativistic generalizations of the Schrödinger equation appeared to be by far nontrivial. The relativistic Hamilton function of a free particle

$$H_0^{\mathrm{rel}} = E = \sqrt{\boldsymbol{p}^2c^2 + m^2c^4} \tag{17}$$

prevents a straightforward use of the rules of the correspondence principle (3) because of the square root of the momentum-dependent quadratic form. In an early attempt (in 1925) Erwin Schrödinger applied the correspondence principle after taking square of Eq. (17). The resulting equation is the following:

$$\left(\frac{\partial^2}{\partial x_\mu^2} - \varkappa^2\right)\psi = 0, \tag{18}$$

where $\varkappa = mc/\hbar$, has been abandoned by Schrödinger since it is second order in time and, thus, the resulting density is not positive definite. Several months later Oskar Klein, Walter Gordon, and, independently, Vladimir Fock derived the same equation, now known as the Klein-Gordon equation. This equation

- Is relativistically invariant,
- May be generalized to account for an external field,
- Does not describe electrons,
- Describes spinless particles.

Equation (18) yields the continuity equation

$$\frac{\partial j_\mu}{\partial x_\mu} = 0 \quad \text{with} \quad j_\mu \sim \psi^* \frac{\partial \psi}{\partial x_\mu} - \frac{\partial \psi^*}{\partial x_\mu}\psi \tag{19}$$

with nonpositive definite density $\rho \sim j_4$. The properties of the Klein-Gordon equation and its applications in quantum field theory are broadly discussed in the literature, e.g., [1–6].

A relativistic equation describing correctly a spin-1/2 particle, derived in 1928 by P. A. M. Dirac and known as the *Dirac equation*, has not been immediately accepted because of its unusual properties. At about the same time also Hendrik Anthony Kramers derived another form of the correct first-order equation, but he did not publish his result until 1933 [7, 19], a year after the presentation of a similar analysis by Bartel Leendert van der Waerden in his book on group theory [20].

Derivation by Kramers

Following the discussion of the Schrödinger equation which led to defining a spin-dependent non-relativistic Hamiltonian (11), the relativistic relation between energy and momentum (17) may be also expressed as

$$E^2\,\mathbf{1} = c^2\,(\boldsymbol{\sigma}\cdot\boldsymbol{p})^2 + m^2c^4\mathbf{1} \tag{20}$$

The correspondence principle yields

$$[\mathsf{E}\,\mathbf{1} - c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p})][\mathsf{E}\,\mathbf{1} + c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p})]\Psi = m^2c^4\,\Psi, \tag{21}$$

where Ψ is a two-component function. Defining $\Psi^{\mathrm{r}} \equiv [\mathsf{E}\,\mathbf{1} + c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p})]\,\Psi/mc^2$, and $\Psi^{\mathrm{l}} \equiv \Psi$, one obtains

$$\begin{aligned}[\mathsf{E}\,\mathbf{1}_2 - c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p})]\,\Psi^{\mathrm{r}} &= mc^2\Psi^{\mathrm{l}},\\ [\mathsf{E}\,\mathbf{1}_2 + c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p})]\,\Psi^{\mathrm{l}} &= mc^2\Psi^{\mathrm{r}},\end{aligned} \tag{22}$$

i.e.,

$$i\hbar\frac{\partial\boldsymbol{\Psi}}{\partial t} = \mathsf{H}_0^{\mathrm{W}}\boldsymbol{\Psi}, \tag{23}$$

where

$$\boldsymbol{\Psi} = \begin{bmatrix}\Psi^{\mathrm{r}}\\ \Psi^{\mathrm{l}}\end{bmatrix} \tag{24}$$

is a four-component wave function and

$$\mathsf{H}_0^{\mathrm{W}} = \begin{bmatrix} c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p}) & mc^2\,\mathbf{1}\\ mc^2\,\mathbf{1} & -c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p})\end{bmatrix}. \tag{25}$$

Equation (23) is known as the *Dirac equation in the representation of Weyl*.

Derivation by Dirac

A relativistic equation of the first order in time has to be also of the first order in coordinates. Therefore, the Hamilton operator deduced from the correspondence principle applied to Eq. (17) should read

$$\mathsf{H}_0^{\mathrm{D}} = c(\alpha_1\mathsf{p}_1 + \alpha_2\mathsf{p}_2 + \alpha_3\mathsf{p}_3) + \beta\,mc^2, \tag{26}$$

where $\alpha_j,\ j = 1,2,3$, and β do not depend on either coordinates or time and

$$\mathsf{H}_0^{\mathrm{D}}\cdot\mathsf{H}_0^{\mathrm{D}} = \left(c^2\,\mathbf{p}^2 + m^2c^4\right)\,\mathbf{1}_4. \tag{27}$$

By combining Eqs. (26) and (27), one readily gets

$$\begin{aligned}\alpha_k\alpha_j + \alpha_j\alpha_k &\equiv \{\alpha_k,\alpha_j\} = 0,\quad \text{if}\ \ k \neq j,\\ \alpha_k\beta + \beta\alpha_k &\equiv \{\alpha_k,\beta\} = 0,\\ \alpha_k^2 = \beta^2 &= \mathbf{1},\quad j,k = 1,2,3.\end{aligned} \tag{28}$$

Thus, $\boldsymbol{\alpha} = \{\alpha_1,\alpha_2,\alpha_3\}$ and β form a set of anticommuting matrices. They have to be Hermitian because the Hamiltonian and the momentum operators are Hermitian. Then, the eigenvalues of these matrices can only be ± 1. The cycle invariance of

trace yields $\mathrm{Tr}\,(\alpha_k) = \mathrm{Tr}\,(\beta^2\alpha_k) = -\mathrm{Tr}\,(\beta\alpha_k\beta) = -\mathrm{Tr}\,(\alpha_k)$. A similar calculation for β gives $\mathrm{Tr}\,(\beta) = -\mathrm{Tr}\,(\beta)$. Therefore,

$$\mathrm{Tr}\,(\alpha_k) = \mathrm{Tr}\,(\beta) = 0. \tag{29}$$

From here we deduce that the number of positive and negative eigenvalues has to be the same, i.e., the order of the matrices is even. The Pauli matrices form a complete set of 2×2 matrices, but only three of them anticommute (the fourth one is the unit matrix which commutes with all matrices). Therefore the smallest possible dimension of $\boldsymbol{\alpha}$ and β matrices is 4. Such a set can easily be constructed, as, for example,

$$\boldsymbol{\alpha} = \begin{bmatrix} \mathbf{0} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & \mathbf{0} \end{bmatrix}, \quad \beta = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{bmatrix}, \tag{30}$$

where all entries represent 2×2 matrices and $\boldsymbol{\sigma}$ are Pauli matrices.

Hamiltonian (26) with matrices (30) is a Hermitian square root of the second-order differential operator given by Eq. (27) and is referred to as the *free Dirac operator*. It is defined in

$$\mathscr{H}_1^{\mathrm{D}} = \mathscr{H}_1^{\mathrm{P}} \oplus \mathscr{H}_1^{\mathrm{P}} = L^2\left(\mathbb{R}^3\right) \otimes \mathbb{C}^4 \tag{31}$$

Hilbert space. By the substitution of the Dirac operator to Eq. (2), the relativistic time evolution equation may be written as

$$i\hbar \frac{\partial \boldsymbol{\Psi}}{\partial t} = \mathsf{H}_0^{\mathrm{D}}\, \boldsymbol{\Psi}. \tag{32}$$

The free Dirac operator may be conveniently expressed as

$$\mathsf{H}_0^{\mathrm{D}} = c\,(\boldsymbol{\alpha} \cdot \mathbf{p}) + \beta\, mc^2 = \begin{bmatrix} mc^2\,\mathbf{1} & c\,(\boldsymbol{\sigma} \cdot \mathbf{p}) \\ c\,(\boldsymbol{\sigma} \cdot \mathbf{p}) & -mc^2\,\mathbf{1} \end{bmatrix}. \tag{33}$$

The wave function has four components. Its transformation properties are specific for objects classified as *Dirac spinors* (cf. section "Transformation Properties of the Dirac Spinor"). One should stress that $\boldsymbol{\Psi}$ is *not* a four-vector and its four-component structure has no relation to the 4 dimensions of the Minkowski space. This kind of coincidence appears only in the case of the Dirac wave functions. The relativistic wave functions describing particles with spin 0 (Klein-Gordon) or with spin 1 (Proca) have, respectively, one and ten components [5].

The Dirac wave function can be expressed as

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{bmatrix} = \begin{bmatrix} \Psi_{\mathrm{L}} \\ \Psi_{\mathrm{S}} \end{bmatrix}, \quad \boldsymbol{\Psi}^\dagger = \left[\psi_1^*, \psi_2^*, \psi_3^*, \psi_4^*\right] = \left[\Psi_{\mathrm{L}}^\dagger, \Psi_{\mathrm{S}}^\dagger\right] \tag{34}$$

where $\Psi_{\mathrm{L}}/\Psi_{\mathrm{S}}$ are two-component functions referred to as *large*/*small* components of $\boldsymbol{\Psi}$. As it is shown in section "The Non-relativistic Limit and the Direct Approach," in the non-relativistic limit, $\Psi_{\mathrm{S}} \to 0$. Equations (27) and (32) imply that each component of the Dirac spinor is a solution of the Klein-Gordon equation (but the inverse is not true).

Representations in the Spinor Space

If $\mathscr{T}$ is a non-singular 4×4 matrix, then the equations

$$i\hbar \frac{\partial \boldsymbol{\Psi}}{\partial t} = \mathsf{H}\boldsymbol{\Psi} \quad \text{and} \quad i\hbar \frac{\partial \boldsymbol{\Psi}'}{\partial t} = \mathsf{H}'\boldsymbol{\Psi}', \tag{35}$$

where

$$\mathsf{H}' = \mathscr{T}\,\mathsf{H}\,\mathscr{T}^{-1} \quad \text{and} \quad \boldsymbol{\Psi}' = \mathscr{T}\boldsymbol{\Psi} \tag{36}$$

are equivalent. In particular, the Dirac equation in the standard (Dirac-Pauli) representation (32) and the Dirac equation in the representation of Weyl (23) are related by

$$\mathscr{T}_{\mathrm{W}} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & -\mathbf{1} \end{bmatrix} = \mathscr{T}_{\mathrm{W}}^{-1}. \tag{37}$$

Two other representations of some practical importance are known as the Majorana representation and the supersymmetric representation. In the Majorana representation, the Dirac equation is purely real. Its properties are discussed in detail in [4, 5, 9, 11]. The supersymmetric representation is related to the standard one by a unitary transformation

$$\mathscr{T}_{\mathrm{SUSY}} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{1} & i\,\mathbf{1} \\ i\,\mathbf{1} & \mathbf{1} \end{bmatrix}. \tag{38}$$

In this representation the Dirac operator

$$\mathsf{H}_0^{\mathrm{SUSY}} = \begin{bmatrix} \mathbf{0} & c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p}) - i\,mc^2\,\mathbf{1} \\ c\,(\boldsymbol{\sigma}\cdot\,\mathbf{p}) + i\,mc^2\,\mathbf{1} & \mathbf{0} \end{bmatrix},$$

after including external fields (see section "External Fields"), is particularly useful in studies of supersymmetric systems [11].

Relativistic Covariance of the Dirac Equation

Originally the Dirac equation was expressed in terms of matrices $\boldsymbol{\alpha}$ and β. This mode of presentation is convenient when the Hamiltonian form of the equation is discussed and when one is interested in the non-relativistic limit. However, for the

discussion of general properties of the Dirac equation, in particular for studies on its transformations between different inertial frames, the form in which time and coordinates are treated in the same way is more appropriate. In order to express the equation in the form which is usually referred to as *covariant*, one has to introduce another set of anticommuting 4×4 matrices, called *gamma matrices*. Each representation of the Dirac equation is associated with its own set of gamma matrices, and, for a given representation, there are several different ways of defining these matrices. In the present discussion the matrices appropriate for the standard (Dirac-Pauli) representation and defined in the way advocated by Sakurai [4] are used.

Gamma Matrices

The four gamma matrices are defined as

$$\gamma_\mu\big|_{\mu=1}^{4} = \{-i\,\beta\boldsymbol{\alpha},\,\beta\} = \{\boldsymbol{\gamma},\,\beta\}\,, \tag{39}$$

or more explicitly

$$\boldsymbol{\gamma} = \begin{bmatrix} \mathbf{0} & -i\boldsymbol{\sigma} \\ i\boldsymbol{\sigma} & \mathbf{0} \end{bmatrix}, \qquad \gamma_4 = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{bmatrix}. \tag{40}$$

Similarly to $\boldsymbol{\alpha}$ and β, gamma matrices are Hermitian, are squares of the unit matrix and anticommute:

$$\gamma_\mu = \gamma_\mu^\dagger, \quad \gamma_1^2 = \gamma_2^2 = \gamma_3^2 = \gamma_4^2 = \mathbf{1}, \quad \text{and, if } \mu \neq \nu, \;\; \{\gamma_\mu, \gamma_\nu\} = 0. \tag{41}$$

It results from (41) that $[\gamma_\mu, \gamma_\nu] = 2\gamma_\mu\gamma_\nu$. A fifth matrix γ_5 defined as

$$\gamma_5 \equiv \gamma_1\gamma_2\gamma_3\gamma_4 = \begin{bmatrix} \mathbf{0} & -\mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{bmatrix} \tag{42}$$

also fulfills conditions (41) characteristic for the gamma matrices:

$$\gamma_5 = \gamma_5^\dagger, \quad \gamma_5^2 = \mathbf{1}, \quad \text{and,} \quad \{\gamma_\mu, \gamma_5\} = 0, \quad [\gamma_\mu, \gamma_5] = 2\gamma_\mu\gamma_5. \tag{43}$$

Clifford Algebra

Matrices γ_p, $p = 1, 2, 3, 4, 5$, and the unit matrix $\mathbf{1}$ make a set of six linearly independent matrices. A complete basis in the space of complex 4×4 matrices forms an algebra. This algebra is isomorphic to the algebra of 4×4 complex matrices, one of the Clifford algebras, named after William Kingdon Clifford who defined it at the

end of the nineteenth century. It is composed of 16 linearly independent matrices Γ_n, $n = 1, 2, \ldots, 16$. The remaining 10 independent matrices are obtained from products of two different gamma matrices:

$$\sigma_{pq} \equiv \frac{1}{2i}\left[\gamma_p, \gamma_q\right] = -i\,\gamma_p\gamma_q, \quad p, q = 1, 2, 3, 4, 5, \quad p < q. \tag{44}$$

As one can see, $\sigma_{qp} = -\sigma_{pq}$. Here is the explicit form of these matrices:

$$\sigma_{12} = \begin{bmatrix} \sigma_3 & 0, \\ 0 & \sigma_3 \end{bmatrix} \equiv \Sigma_3, \quad \text{cyclic}\{1, 2, 3\} \tag{45}$$

$$\sigma_{k4} = \begin{bmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{bmatrix} \equiv \alpha_k, \quad k = 1, 2, 3 \tag{46}$$

Matrices Γ_n, $n = 1, 2, \ldots, 16$, fulfill several basic relations:

1. From $\Gamma_n^2 = \mathbf{1}$ results that $\Gamma_n = \Gamma_n^{-1}$ and $\mathrm{Tr}\left(\Gamma_n^2\right) = 4$
2. If $\Gamma_n, \Gamma_m \neq \mathbf{1}$ then $\mathrm{Tr}\,(\Gamma_n\,\Gamma_m) = 0$ since $\{\Gamma_n, \Gamma_m\} = 0$, and $\mathrm{Tr}\,(\Gamma_n) = 0$ because $\mathrm{Tr}\,(\Gamma_n) = \mathrm{Tr}\left(\Gamma_m^2\Gamma_n\right) = -\mathrm{Tr}\,(\Gamma_m\Gamma_n\Gamma_m) = -\mathrm{Tr}\left(\Gamma_m^{-1}\Gamma_n\Gamma_m\right) = -\mathrm{Tr}\,(\Gamma_n)$.

Theorem 1. *An arbitrary* 4×4 *matrix* G *may be expressed in terms of* Γ_n *matrices as*

$$G = \frac{1}{4}\sum_{n=1}^{16} \mathrm{Tr}\,(G\,\Gamma_n)\;\Gamma_n. \tag{47}$$

Covariant Form of the Dirac Equation

Using definitions (1) and (39), one can combine Eqs. (32) and (33) to the so-called *covariant form of the Dirac equation*:

$$\left[\gamma_\mu \frac{\partial}{\partial\, x_\mu} + \varkappa\right] \boldsymbol{\Psi}(\boldsymbol{x}) = 0, \qquad \varkappa = \frac{mc}{\hbar}, \tag{48}$$

where $\boldsymbol{x} = x_\mu\big|_{\mu=1}^{4}$. In order to write the equation conjugate to (48), one has to introduce a new form of conjugation of the Dirac spinor:

$$\overline{\boldsymbol{\Psi}} = \boldsymbol{\Psi}^\dagger\beta \equiv \boldsymbol{\Psi}^\dagger\gamma_4 = \left[\psi_1^*, \psi_2^*, -\psi_3^*, -\psi_4^*\right]. \tag{49}$$

The Hermitian conjugate of Eq. (48) reads

$$\frac{\partial\,\overline{\boldsymbol{\Psi}}(\boldsymbol{x})}{\partial\, x_\mu}\gamma_\mu - \varkappa\,\overline{\boldsymbol{\Psi}}(\boldsymbol{x}) = 0. \tag{50}$$

By multiplying Eqs. (48) and (50), respectively, by $\overline{\boldsymbol{\Psi}}$ from the left and by $\boldsymbol{\Psi}$ from the right, subsequent subtraction and simple rearrangement of the result, one gets the continuity equation

$$\frac{\partial\, j_\mu}{\partial\, x_\mu} = 0, \quad j_\mu = i\, c\, \overline{\boldsymbol{\Psi}}\, \gamma_\mu\, \boldsymbol{\Psi} = (\boldsymbol{j}\,, i c \rho)\,. \tag{51}$$

In the 3D notation, using (49) and (39), one gets

$$\mathbf{j} = c\ \boldsymbol{\Psi}^\dagger \boldsymbol{\alpha}\, \boldsymbol{\Psi}, \quad \rho = \boldsymbol{\Psi}^\dagger \boldsymbol{\Psi}. \tag{52}$$

All Dirac equations expressed in terms of matrices which fulfill conditions (41) are related by a unitary transformation. Thus, all these equations are equivalent in terms of the physics they describe. This remarkable property of the Dirac equation results from the *fundamental Pauli theorem*:

Theorem 2. *If two sets of* 4×4 *matrices satisfy* $\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$ *and* $\{\gamma'_\mu, \gamma'_\nu\} = 2\delta_{\mu\nu}$, $\mu, \nu = 1, 2, 3, 4$, *then there exists a non-singular and unique, up to a multiplicative constant, matrix* $\mathscr{T}$ *such that* $\gamma'_\mu = \mathscr{T} \gamma_\mu \mathscr{T}^{-1}$.

A proof of this theorem may be found, e.g., in the monograph by Sakurai [4]. One should remember that the structure of the wave function depends on the set of gamma matrices used: If $\boldsymbol{\Psi}$ corresponds to $\boldsymbol{\gamma}$, then $\mathscr{T}\, \boldsymbol{\Psi}$ corresponds to the $\boldsymbol{\gamma}'$.

Transformation Properties of the Dirac Spinor

The Dirac equation is covariant if after a Lorentz transformation it has the same form as before the transformation, i.e., if in the primed coordinate system it reads

$$\left[\gamma'_\mu \frac{\partial}{\partial\, x'_\mu} + \varkappa\right] \boldsymbol{\Psi}'(\boldsymbol{x}') = 0, \tag{53}$$

and if an explicit prescription which relates $\boldsymbol{\Psi}(\boldsymbol{x})$ and $\boldsymbol{\Psi}'(\boldsymbol{x}')$ exists. Matrices γ'_μ in Eq. (53) have to fulfill relations (41). Hence, according to the fundamental Pauli theorem, there exists a non-singular matrix $\mathscr{T}$ such that $\gamma'_\mu = \mathscr{T} \gamma_\mu \mathscr{T}^{-1}$ and the Dirac equations corresponding to all $\mathscr{T}$ are equivalent. Therefore, without any loss of generality, one can set $\mathscr{T} = \mathbf{1}_4$, i.e., $\boldsymbol{\gamma}' = \boldsymbol{\gamma}$.

Matrix $\boldsymbol{a}$ of the Lorentz transformation (6) is coordinate independent. Therefore, one can expect that

$$\boldsymbol{\Psi}'(\boldsymbol{x}') = \mathscr{S} \boldsymbol{\Psi}(\boldsymbol{x}), \tag{54}$$

where $\mathscr{S}$ is a non-singular 4×4 matrix independent of $\boldsymbol{x}$. Since $\partial/\partial x'_\mu = a_{\mu\nu}\partial/\partial x_\nu$, Eq. (53) may be rewritten as

$$\left[\mathscr{S}^{-1}\gamma_\mu \mathscr{S} a_{\mu\nu}\frac{\partial}{\partial x_\nu} + \varkappa\right]\boldsymbol{\Psi}(\boldsymbol{x}) = 0. \tag{55}$$

As one can see, the Dirac equation is covariant, i.e., Eqs. (48) and (55) are the same, if

$$\mathscr{S}^{-1}\gamma_\mu \mathscr{S} a_{\mu\nu} = \gamma_\nu \quad \Rightarrow \quad \mathscr{S}\gamma_\nu \mathscr{S}^{-1} = \gamma_\mu a_{\mu\nu}. \tag{56}$$

From Eq. (56) one can derive $\mathscr{S}$ corresponding to a selected $\boldsymbol{a}$. Detailed derivations may be found, e.g., in [1, 4, 5, 10, 11]. In the next subsections only the final results are presented.

According to (56), the covariance of the Dirac equation implies that $\boldsymbol{\gamma}$ transforms like a four-vector. A detailed discussion of the transformation properties of the gamma matrices may be found in Section 5.B of [10] and in Chapter 2 of [11]. Note that the transformation rule of the gamma matrices is analogous to the one of the Pauli matrices $\boldsymbol{\sigma}$ [10, 12].

Proper Lorentz Transformation

Under a rotation by ω in 1-2 plane, the transformation matrix for the Pauli spinors is equal to [1, 4]

$$\mathscr{S}^{(12)}_{\text{Pauli}} = \mathbf{1}_2 \cos\tfrac{\omega}{2} + i\,\sigma_3 \sin\tfrac{\omega}{2}. \tag{57}$$

Under the same rotation in 1-2 plane of the Minkowski space, the Dirac spinors transform according to

$$\mathscr{S}^{(12)}_{\text{rot}} = \mathbf{1}_4 \cos\tfrac{\omega}{2} + i\,\Sigma_3 \sin\tfrac{\omega}{2} = \mathbf{1}_4 \cos\tfrac{\omega}{2} + i\,\sigma_{12} \sin\tfrac{\omega}{2} \tag{58}$$

and under a special case of the Lorentz boost – a rotation in 1-4 plane by an imaginary angle $\overline{\omega} = i\,\chi$, where $\chi = \operatorname{arctanh}(v/c)$, according to

$$\mathscr{S}^{(14)}_{\text{boost}} = \mathbf{1}_4 \cosh\tfrac{\chi}{2} - \sigma_{14} \sinh\tfrac{\chi}{2}. \tag{59}$$

As one can see, $\mathscr{S}^{(jk)\dagger}_{\text{rot}} = \mathscr{S}^{(jk)-1}_{\text{rot}}$, but $\mathscr{S}^{(k4)\dagger}_{\text{boost}} = \mathscr{S}^{(k4)}_{\text{boost}}$. This means that the transformation $\mathscr{S}^{(k4)}_{\text{boost}}$ is not unitary. It is important to note that

$$\mathscr{S}^\dagger = \gamma_4 \mathscr{S}^{-1} \gamma_4 \tag{60}$$

and $[\mathscr{S}, \gamma_5] = 0$.

Space Inversion and Parity

The form of the Dirac equation is conserved under the space inversion, i.e., under the replacement $\boldsymbol{r} \rightarrow -\boldsymbol{r}$, if the transformation operator $\mathscr{S}_{\text{inv}}$ is equal to γ_4, i.e., if

$$\mathscr{S}_{\text{inv}}\,\boldsymbol{\Psi}(\boldsymbol{x}) = \boldsymbol{\Psi}'(\boldsymbol{x}') = \gamma_4\,\boldsymbol{\Psi}(\boldsymbol{x}) \tag{61}$$

[1, 2, 4]. But $\boldsymbol{\Psi}'(\boldsymbol{x}') = \mathsf{P}_0\boldsymbol{\Psi}'(\boldsymbol{x})$ where P_0 performs the transformation of coordinates without changing the form of the function, i.e., it is the inversion operator of the non-relativistic quantum mechanics. Therefore,

$$\boldsymbol{\Psi}'(\boldsymbol{x}) = \mathsf{P}_0\,\boldsymbol{\Psi}'(\boldsymbol{x}') = \mathsf{P}_0\,\gamma_4\,\boldsymbol{\Psi}(\boldsymbol{x}) \equiv \mathsf{P}\,\boldsymbol{\Psi}(\boldsymbol{x}), \tag{62}$$

where

$$\mathsf{P} = \mathsf{P}_0\,\gamma_4 \tag{63}$$

is the *parity operator*.

Infinitesimal Rotation

The angular momentum is a generator of rotations in 3D space, similarly to how energy generates translations in time and linear momentum generates translations in 3D space [1]. In the Minkowski space the form of the Dirac spinor under an infinitesimal rotation in 1–2 plane changes according to [5]

$$\boldsymbol{\Psi}'(\boldsymbol{x}) = (\mathbf{1} + i\,\mathsf{J}_3\,\delta\omega)\,\boldsymbol{\Psi}(\boldsymbol{x}), \tag{64}$$

where

$$\mathsf{J}_3 = \mathsf{L}_3 + \mathsf{S}_3, \quad \mathsf{S}_3 = \frac{1}{2}\Sigma_3 \tag{65}$$

Thus, the generator of rotations in the Minkowski space is composed of two parts: the orbital part described by L_3 and the spin part described by S_3. The first one acts in $\mathbb{R}^3$ space and the second one in the spinor space, $\mathbb{C}^4$. Therefore, Eq. (65) should be written as $\mathsf{J}_3 = \mathsf{L}_3 \otimes \mathbf{1}(\mathbb{C}^4) + \mathbf{1}(\mathbb{R}^3) \otimes \mathsf{S}_3$, where $\mathbf{1}(\mathbb{C}^4)$ and $\mathbf{1}(\mathbb{R}^3)$ are the appropriate unit operators. However, usually the simplified notation, as in Eq. (65), is used. Obviously, neither the orbital angular momentum nor the spin operators are covariant. A covariant description of spin is given by the *Pauli-Lubanski* vector. For a detailed discussion, a reader is referred to, e.g., [5].

Bilinear Covariants $\overline{\boldsymbol{\Psi}}\,\boldsymbol{\Gamma}\,\boldsymbol{\Psi}$

Let us consider transformation properties of the bilinear forms $\overline{\boldsymbol{\Psi}}\,\Gamma\,\boldsymbol{\Psi}$, where Γ is an element of the Clifford algebra. If $\boldsymbol{\Psi}'(\boldsymbol{x}') = \mathscr{S}\boldsymbol{\Psi}(\boldsymbol{x})$, then, according to Eqs. (49) and (60), $\boldsymbol{\Psi}'(\boldsymbol{x}')^\dagger = \boldsymbol{\Psi}(\boldsymbol{x})^\dagger\,\mathscr{S}^\dagger = \overline{\boldsymbol{\Psi}}(\boldsymbol{x})\mathscr{S}^{-1}\gamma_4$. From here $\boldsymbol{\Psi}'(\boldsymbol{x}')^\dagger\gamma_4 =$

Table 1 Complete set of the bilinear covariants

Scalar	Pseudoscalar	Vector	Pseudovector	Antisymmetric tensor
$\overline{\boldsymbol{\Psi}}\boldsymbol{\Psi}$	$\overline{\boldsymbol{\Psi}}\gamma_5\boldsymbol{\Psi}$	$\overline{\boldsymbol{\Psi}}\gamma_\mu\boldsymbol{\Psi}$	$i\,\overline{\boldsymbol{\Psi}}\gamma_5\gamma_\mu\boldsymbol{\Psi}$	$\overline{\boldsymbol{\Psi}}\sigma_{\mu\nu}\boldsymbol{\Psi}$

$\overline{\boldsymbol{\Psi}}(\boldsymbol{x})\mathscr{S}^{-1}$, and $\overline{\boldsymbol{\Psi}}'(\boldsymbol{x}') = \overline{\boldsymbol{\Psi}}(\boldsymbol{x})\mathscr{S}^{-1}$. Consequently, $\overline{\boldsymbol{\Psi}}'(\boldsymbol{x}')\boldsymbol{\Psi}'(\boldsymbol{x}') = \overline{\boldsymbol{\Psi}}(\boldsymbol{x})\boldsymbol{\Psi}(\boldsymbol{x})$. This means that $\overline{\boldsymbol{\Psi}}\boldsymbol{\Psi}$ is invariant under the Lorentz transformation, i.e., it is a scalar. In a similar way, using Eq. (56), one can show that $\left(\overline{\boldsymbol{\Psi}}\gamma_\mu\boldsymbol{\Psi}\right)' = a_{\mu\nu}\left(\overline{\boldsymbol{\Psi}}\gamma_\nu\boldsymbol{\Psi}\right)$, i.e., $\overline{\boldsymbol{\Psi}}\gamma_\nu\boldsymbol{\Psi}$ is a four-vector. In particular, j_μ defined in Eq. (52) is a four-vector and $\rho = \boldsymbol{\Psi}^\dagger\boldsymbol{\Psi}$ is not a scalar but the fourth component of this vector. A complete set of bilinear Lorentz covariants is given in Table 1.

Equation (51) may be rewritten as

$$\nabla\mathbf{j} + \frac{\partial\rho}{\partial t} = 0. \tag{66}$$

From here, using the Gauss theorem, one can see that the integral of ρ taken over the domain of $\boldsymbol{\Psi}$ is time-independent. This result justifies the standard normalization condition for the Dirac wave function:

$$\int_V \boldsymbol{\Psi}^\dagger\boldsymbol{\Psi}\, d^3x = 1. \tag{67}$$

The conservation of the norm defined in this way was not obvious a priori, since in the relativistic theory ρ is the fourth component of the four-current vector rather than a scalar. In the quantum theory of many-electron systems, the relativistic wave functions are usually normalized according to Eq. (67). However, in many other areas of application of quantum physics, the bilinear scalar, $\overline{\boldsymbol{\Psi}}\boldsymbol{\Psi}$, is normalized instead.

In the Pauli representation, eight Γ matrices, $\Gamma_{\mathrm{L}} = \left\{\mathbf{1}, \gamma_4, i\gamma_5\gamma_k, \sigma_{jk}\right\}$, are block-diagonal, and the remaining eight, $\Gamma_{\mathrm{S}} = \{\gamma_k, i\gamma_5\gamma_4, \gamma_5, \sigma_{k4}\}$, are block-antidiagonal. The corresponding bilinear covariants read

$$\begin{aligned}
\overline{\boldsymbol{\Psi}}\Gamma_{\mathrm{L}}\boldsymbol{\Psi} &= \left[\Psi_{\mathrm{L}}^\dagger, -\Psi_{\mathrm{S}}^\dagger\right]\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}\end{bmatrix}\begin{bmatrix}\Psi_{\mathrm{L}}\\ \Psi_{\mathrm{S}}\end{bmatrix} = \Psi_{\mathrm{L}}^\dagger\mathbf{A}\Psi_{\mathrm{L}} - \Psi_{\mathrm{S}}^\dagger\mathbf{B}\Psi_{\mathrm{S}},\\
\overline{\boldsymbol{\Psi}}\Gamma_{\mathrm{S}}\boldsymbol{\Psi} &= \left[\Psi_{\mathrm{L}}^\dagger, -\Psi_{\mathrm{S}}^\dagger\right]\begin{bmatrix}\mathbf{0} & \mathbf{A}\\ \mathbf{A}^\dagger & \mathbf{0}\end{bmatrix}\begin{bmatrix}\Psi_{\mathrm{L}}\\ \Psi_{\mathrm{S}}\end{bmatrix} = \Psi_{\mathrm{L}}^\dagger\mathbf{A}\Psi_{\mathrm{S}} - \Psi_{\mathrm{S}}^\dagger\mathbf{A}^\dagger\Psi_{\mathrm{L}}
\end{aligned} \tag{68}$$

In the non-relativistic limit, i.e., for $c \to \infty$, the small component of the Dirac spinor vanishes. Thus, $\lim_{c\to\infty}\overline{\boldsymbol{\Psi}}\Gamma_{\mathrm{L}}\boldsymbol{\Psi} = \Psi_{\mathrm{L}}^\dagger\mathbf{A}\Psi_{\mathrm{L}}$ and $\lim_{c\to\infty}\overline{\boldsymbol{\Psi}}\Gamma_{\mathrm{S}}\boldsymbol{\Psi} = 0$. Therefore, the bilinear covariants associated with Γ_{L} / Γ_{S} are referred to as *large*/*small* covariants. Since the set of 16Γ matrices is complete, one cannot generate more linearly independent bilinear covariants of the form $\overline{\boldsymbol{\Psi}}\Gamma\boldsymbol{\Psi}$.

Eigenvalue Problem of the Free Dirac Operator

The free Dirac operator (33) commutes with $\mathbf{p}$. Therefore, solutions of Eq. (32) corresponding to a given energy E may be written as a product of a scalar part depending on time and coordinates and a spinor part:

$$\boldsymbol{\Psi}(\boldsymbol{x}) = \frac{\mathrm{e}^{i(\boldsymbol{p}\cdot\boldsymbol{r}-Et)}}{(2\pi)^2}\boldsymbol{\Phi}, \tag{69}$$

where the spinor part of the wave function

$$\boldsymbol{\Phi} = \begin{bmatrix}\phi_1\\ \phi_2\\ \phi_3\\ \phi_4\end{bmatrix} = \begin{bmatrix}\Phi_{\mathrm{L}}\\ \Phi_{\mathrm{S}}\end{bmatrix}, \quad \boldsymbol{\Phi}^\dagger = \left[\phi_1^*, \phi_2^*, \phi_3^*, \phi_4^*\right] = \left[\Phi_{\mathrm{L}}^\dagger, \Phi_{\mathrm{S}}^\dagger\right] \tag{70}$$

does not depend on time and, in the case of a free particle, on coordinates. Note that except for some special cases, as free electron or electron in a homogeneous magnetic field (section "Electron in a Magnetic Field"), the separation of coordinate-independent spinors from the Dirac wave function cannot be performed and $\boldsymbol{\Phi}$ depends on the electron coordinates. Since this separation is possible in the free-electron case, the eigenvalue equation of $\mathsf{H}_0^{\mathrm{D}}$ may be expressed as a set of homogeneous linear equations

$$\begin{bmatrix}mc^2 - E & c\,(\boldsymbol{\sigma}\cdot\boldsymbol{p})\\ c\,(\boldsymbol{\sigma}\cdot\boldsymbol{p}) & -mc^2 - E\end{bmatrix}\begin{bmatrix}\Phi_{\mathrm{L}}\\ \Phi_{\mathrm{S}}\end{bmatrix} = 0. \tag{71}$$

Equation (71) has nontrivial solutions if

$$\det\begin{vmatrix}mc^2 - E, & c\,(\boldsymbol{\sigma}\cdot\boldsymbol{p})\\ c\,(\boldsymbol{\sigma}\cdot\boldsymbol{p}), & -mc^2 - E\end{vmatrix} = 0 \tag{72}$$

i.e., if $E = \lambda\, E_p$, where $\lambda = \pm 1$ and

$$E_p = mc^2\sqrt{1 + \left(\frac{\boldsymbol{p}}{mc}\right)^2}. \tag{73}$$

Then, the spectrum of the free-particle Dirac operator is given by

$$E \in \left(-\infty, -mc^2\right] \cup \left[mc^2, \infty\right). \tag{74}$$

The spectrum consists of two continua. The positive one spreads from the rest energy of the particle, mc^2, to $+\infty$. It corresponds to the non-relativistic energy of a free particle. The negative continuum, spreading from $-mc^2$ to $-\infty$, has no physical meaning. Mathematically this implies that the plain Dirac equation cannot describe a physically stable system since the ground state does not exist.

The Hilbert space in which the Dirac operator is defined contains vectors corresponding to the positive and to the negative energies. Since a quantum mechanical state is defined as a vector in the Hilbert space of the Hamiltonian, one can define the *positive-energy states* and the *negative-energy states*, though the latter ones have no physical meaning. It is important to note that in the charge-conjugate equation, the negative- energy states correspond to positrons with positive energies (cf. section "Charge Conjugation, Positrons").

Constants of Motion

As it was already noticed, $\left[\mathsf{H}_0^{\mathrm{D}},\ \mathbf{p}\right] = 0$. Then, similarly as in the non-relativistic theory, for a free Dirac particle, the linear momentum is conserved. However, as one can easily check, neither orbital angular momentum $\mathbf{L}$ nor spin

$$\mathbf{S} = \frac{1}{2}\,\boldsymbol{\Sigma}, \qquad \boldsymbol{\Sigma} = \{\Sigma_1, \Sigma_2, \Sigma_3\},$$

commutes with $\mathsf{H}_0^{\mathrm{D}}$, but $\mathbf{J} = \mathbf{L} + \mathbf{S}$ does commute. This last observation and the way the Dirac spinors transform under infinitesimal rotations (64) are important for the understanding of the role of spin: from the formal point of view, spin is optional in the Schrödinger theory, while spin is a necessary attribute of the Dirac particles. Since the eigenvalues of S_3 are equal to $\pm\,1/2$, the Dirac equation describes particles with spin $1/2$. Two other operators which commute with $\mathsf{H}_0^{\mathrm{D}}$ are helicity:

$$\mathsf{h} = \frac{(\boldsymbol{\Sigma}\cdot\ \mathbf{p})}{p}, \tag{75}$$

and parity P, defined in Eq. (63). Helicity is a 4×4 matrix with eigenvalues $h = \pm1$. The states with helicity $+1/-1$ (spin parallel/opposite to the direction of the motion) are referred to as right/left handed. However, neither $\mathbf{p}$ and $\mathbf{J}$ nor P and h commute with each other. Since P changes sign of $\mathbf{p}$, it does not commute with $\mathbf{p}$. Therefore, one can select the following sets of five commuting operators which include $\mathsf{H}_0^{\mathrm{D}}$ and, thus, can be used to construct and label its eigenfunctions:

- Plane waves: $\mathsf{H}_0^{\mathrm{D}}$, p_1, p_2, p_3, and h;
- Spherical waves: $\mathsf{H}_0^{\mathrm{D}}$, J^2, J_3, p^2, and either h or P.

Free-Particle Spinors

The free-particle spinors corresponding to the plane waves may be chosen as simultaneous eigenfunctions of $\mathsf{H}_0^{\mathrm{D}}$, $\boldsymbol{p}$, and h. In the case of a free particle, the operator

$$\Lambda = \frac{\mathsf{H}_0^{\mathrm{D}}}{E_p} \tag{76}$$

commutes both with $\mathsf{H}_0^{\mathrm{D}}$ and with all operators commuting with the Hamiltonian. Its eigenvalues $\lambda = +1 / -1$ correspond to the positive-/negative-energy solutions [cf. Eqs. (72) and (73)]. If axis 3 of the coordinate system is directed along the momentum of the particle, i.e., if $\boldsymbol{p} = \{0, 0, p_3\}$, then $\mathsf{h} = \Sigma_3\, p_3/|p_3|$. Thus, in this case, also Σ_3 is a constant of the motion. As a consequence, the free-particle spinors may be chosen as simultaneous eigenfunctions of $\mathsf{H}_0^{\mathrm{D}}$, $\boldsymbol{p}$, h, Λ, and Σ_3. They are labeled as $\boldsymbol{\Phi}^{\lambda}_{p,s}$, where $p = p_3$ and $s = \pm 1$ are eigenvalues of Σ_3.

Some simple algebra applied to Eq. (71) yields

$$\boldsymbol{\Phi}^{+}_{p,1} = \begin{bmatrix} N^+ \\ 0 \\ N^- \\ 0 \end{bmatrix}, \quad \boldsymbol{\Phi}^{+}_{p,-1} = \begin{bmatrix} 0 \\ -N^+ \\ 0 \\ N^- \end{bmatrix}, \quad \boldsymbol{\Phi}^{-}_{p,1} = \begin{bmatrix} -N^- \\ 0 \\ N^+ \\ 0 \end{bmatrix}, \quad \boldsymbol{\Phi}^{-}_{p,-1} = \begin{bmatrix} 0 \\ N^- \\ 0 \\ N^+ \end{bmatrix}, \tag{77}$$

where

$$N^+ = \sqrt{\frac{E_p + mc^2}{2E_p}}, \quad N^- = \frac{p_3}{|p_3|}\sqrt{\frac{E_p - mc^2}{2E_p}}. \tag{78}$$

For $\boldsymbol{p} = 0$, i.e., for the particle at rest, $N^+ = 1$ and $N^- = 0$. The transformation of $\boldsymbol{p} = 0$ spinors to the form corresponding to an arbitrary $\boldsymbol{p}$ may also be obtained using the Lorentz transformation. A detailed analysis may be found, e.g., in [1,5,6].

Position and Velocity Operators, Zitterbewegung

One could expect that an electron in a positive-energy state should fall into a negative-energy state and emit a photon with an energy which may be infinite, since the spectrum is unlimited. In order to solve this difficulty, Dirac proposed that all the negative-energy states are filled under normal conditions and the Pauli exclusion principle prevents transitions to these states. An excitation of one negative-energy electron results in creation of a hole in the "Dirac vacuum" and of one positive-energy electron. This process corresponds to the creation of an

electron-positron pair. Thus, in this model *the number of particles is not conserved*. The conserved quantity is the *total charge*.

However, it is hard to classify this interpretation as satisfying. It may only be applied to fermions, while the Klein-Gordon equation describing 0-spin bosons has also unbounded from below energy spectrum. The Dirac model describes a single electron, and, at the same time, it is unable to explain the behavior of a free electron without assuming that it is surrounded by an infinite number of non-interacting electrons occupying the negative-energy states. This contradiction, inseparable from the Dirac model, results in numerous artifacts and stimulated work on the formulation of a more general Lorentz-covariant theory in which the energy spectrum of a system of electrons is bounded from below. It has been realized that the Dirac equation should be interpreted as a classical equation for the matter field, as the Maxwell equation is the classical equation for the electromagnetic field. A theory which correctly describes the physical reality, quantum electrodynamics (QED), has been formulated during 1927–1937 by Dirac himself and by Wolfgang Pauli, Eugene Wigner, Werner Heisenberg, and Enrico Fermi. QED unifies theories describing matter (as the Dirac or the Klein-Gordon models) and radiation (the Maxwell equation). The resulting equations have been obtained by the *second quantization* procedure. In effect all contradictions of the original Dirac model have been solved. The unboundedness from below of the energy spectrum has been removed, the possibility of creation and annihilation of particles has been introduced, and emission and absorption of photons have been correctly described. Electrons and positrons have been described on an equal footing (in the original Dirac model, positrons are interpreted as "holes in the vacuum" while the only "real particles" are electrons) [2–4,11]. Nevertheless, the Dirac model of vacuum survived as the very first intuitive quantum model of vacuum. The vacuum has been defined as a polarizable medium from which new particles can be created. Observable effects, as, e.g., *vacuum polarization* and related to the vacuum fluctuations *self energy*, can be explained using this idea. Many concepts of the Dirac vacuum have been retained in quantum field theories, and a substantial part of terminology used to describe the phenomena related to the physical vacuum may be traced back to the original ideas of Dirac [21].

The Dirac theory is considerably simpler than QED and leads to a much more precise description of the physical reality than the approaches derived from the non-relativistic Schrödinger equation. Therefore, it is the basis for nearly all relativistic methods of description of many-electron systems, though the unboundedness from below of the spectrum is the origin of serious conceptual and computational problems.

One of the consequences of the unboundedness from below of the Dirac spectrum is unusual properties of the velocity operator. In the Schrödinger theory,

$$\mathbf{v}_{\mathrm{S}} = \left(\frac{d\,\mathbf{r}}{dt}\right)_{\mathrm{S}} = i\,[\mathsf{H}_{\mathrm{S}},\ \mathbf{r}] = \frac{\mathbf{p}}{m}. \tag{79}$$

In the Dirac case,

$$\mathbf{v}_{\mathrm{D}} = \left(\frac{d\,\mathbf{r}}{dt}\right)_{\mathrm{D}} = i\,[\mathsf{H}_{\mathrm{D}},\ \mathbf{r}] = c\,\boldsymbol{\alpha}. \tag{80}$$

This result is very strange in many aspects. First, the eigenvalues of α_k are ± 1. Therefore, the eigenvalues of $\mathbf{v}_{\mathrm{D}}$ are equal to $\pm c$. Besides, different components of $\mathbf{v}_{\mathrm{D}}$ do not commute. Then, a measurement of one component of the Dirac velocity is incompatible with a measurement of another one. This is also strange since different components of the momentum operator do commute. Moreover, $\mathbf{v}_{\mathrm{D}}$ does not commute with $\mathsf{H}_0^{\mathrm{D}}$. Therefore, the velocity is not a constant of the motion despite the fact that the particle is free.

An analysis of the time dependence of $\boldsymbol{\alpha}$ and of the coordinate operator $\mathbf{r}$ leads to the conclusion that both of them execute very rapid oscillations with an angular frequency $(2mc^2)/\hbar \approx 1.5 \cdot 10^{21}\mathrm{s}^{-1}$. This motion, named by Schrödinger *Zitterbewegung*, is due to an interference between the positive- and negative-energy components of the wave packet describing the electron. Intuitively it may be interpreted as a consequence of a permanent creation and annihilation of the so-called *virtual electron-positron pairs*. In order to understand the background of this process, it is convenient to introduce *even* and *odd* operators. An even operator acting on a function which belongs to the positive (negative) energy subspace of a free electron transforms it into a function which belongs to the same subspace. An odd operator transforms a positive (negative)-energy-subspace function into a function which belongs to the complementary subspace. An operator, say Ω, may be decomposed into its even part $[\Omega]$ and its odd part $\{\Omega\}$:

$$\Omega = [\Omega] + \{\Omega\}, \tag{81}$$

where

$$[\Omega] = \frac{1}{2}(\Omega + \Lambda\Omega\Lambda), \quad \{\Omega\} = \frac{1}{2}(\Omega - \Lambda\Omega\Lambda). \tag{82}$$

After some algebra one can see that

$$[\,\mathbf{v}_{\mathrm{D}}] = c\,[\boldsymbol{\alpha}] = \frac{c^2\,\mathbf{p}}{E}. \tag{83}$$

This result is in agreement with the standard definition of velocity. A detailed discussion of this subject may be found, e.g., in [1,4,5,8,9,11]. Operators

$$\Pi_{\pm} = \frac{1}{2}(\mathbf{1} \pm \Lambda), \tag{84}$$

are projection operators to the positive (Π_+) and to the negative (Π_-) energy subspace. For example, $\Pi_+\Omega\Pi_+$ acts in the free-electron positive-energy space only. One may think about modifying the theory by restricting the Hilbert space to

its positive-energy part. Technically this can be done using projection operators (84) and, in fact, is done in some approximate models. Since $\Pi_\pm$ commutes with the free Dirac operator, an initial state of a free particle belonging to the positive-energy space will remain in this space forever. However, position-dependent operators do not commute with $\Pi_\pm$. They are composed of an even and an odd part and couple the positive- and the negative-energy space. For example, the even part of the position operator is given by

$$[\,\mathbf{x}] = \mathbf{x} + \frac{i}{2E}\,(\,\mathbf{v}_\mathrm{D} - [\,\mathbf{v}_\mathrm{D}])\,. \tag{85}$$

Due to the presence of the term proportional to $\mathbf{v}_\mathrm{D}$, the position of a Dirac particle is diffuse in the area comparable to the Compton wavelength $\hbar/mc$.

External Fields

The notion of a particle in an external field is an idealized concept. Neither the quantum structure of the field nor the influence of the particle on the field is taken into account. The external field is used to replace in an effective way the real interaction. For example, a simple, exactly solvable problem of an electron in the Coulomb potential may be used to effectively describe a hydrogen-like atom – a highly complicated mathematically and unsolvable analytically system of two interacting particles. The external potentials may have different transformation properties (scalar, pseudoscalar, vector, pseudovector, tensor) [5, 11]. In this chapter the discussion is limited to scalar and electromagnetic vector potentials. It is also assumed that the potentials are time independent. This means that only closed systems are considered.

Dirac Equation for a Particle in an External Field

The Dirac operator in a field described by a scalar potential $\mathsf{V}_\mathrm{sc}(\boldsymbol{r})$ may be obtained from the free Dirac operator (33) by the replacement

$$mc^2 \;\mapsto\; mc^2 + \mathsf{V}_\mathrm{sc}(\boldsymbol{r}). \tag{86}$$

Scalar interactions for electrons and for positrons are the same – they do not depend on the electric charge of the particle. Therefore, they can be attractive for electrons and positrons at the same time. In quantum chromodynamics, scalar potentials are used to model the confinement of quarks, but in theory of many-electron systems, they appear very seldom, since interactions of non-electromagnetic origin (as, e.g., the strong ones or the gravitational ones) are of rather marginal importance. However, already in 1973, a possibility of establishing a limit on the scalar coupling constant by discussing the Coulomb potential with a scalar contribution

was discussed [22]. Until now there is no experimental evidence for the existence of long-range scalar interactions. Such interactions are mediated by a scalar massless boson which is experimentally unknown and does not exist in the *Standard Model*. But the atomic spectroscopy is one of the most sensitive tools for studying interactions. Therefore, studies on possibilities of some admixture of the scalar coupling to the electromagnetic one may be useful in the search of phenomena which are not described by the *Standard Model*.

The coupling of a Dirac particle to an electromagnetic field may be obtained by applying the *principle of minimal coupling* [5, 9–11], originally developed in classical mechanics. According to this principle, in order to account for an external field described by an electromagnetic four-potential

$$\mathsf{A}_\mu = \{\mathbf{A}, i\,\varphi/c\}, \tag{87}$$

one has to modify the free-particle equations by the replacement

$$\mathsf{p}_\mu \;\mapsto\; \pi_\mu \equiv \mathsf{p}_\mu - q\,\mathsf{A}_\mu, \tag{88}$$

where q is the charge of the particle. Note that in many textbooks in Eq. (88) instead of $q\,\mathsf{A}_\mu$, $q\,\mathsf{A}_\mu/c$ appears. The factor $1/c$ depends on the choice of units and has nothing to do with relativistic effect. In particular, it is present in the *Gaussian cgs* system and absent in the *SI*. Thus, the free-particle Dirac equation (48) generalized to account for the external fields reads

$$\left[\gamma_\mu \mathsf{D}_\mu + mc^2 + \mathsf{V}_{\rm sc}\right] \boldsymbol{\Psi}(\boldsymbol{x}) = 0, \tag{89}$$

where

$$\mathsf{D}_\mu = \frac{\partial}{\partial x_\mu} - i\,q\,\mathsf{A}_\mu(\boldsymbol{r}), \tag{90}$$

is the *gauge covariant derivative*. Combining Eqs. (33), (86), and (88), one obtains the Dirac operator for a particle in an external field:

$$\mathsf{H}^{\rm D} = c\,(\boldsymbol{\alpha}\cdot\boldsymbol{\pi}) + \mathsf{V}_{\rm el}\,\mathbf{1} + \left(mc^2 + \mathsf{V}_{\rm sc}\right)\,\beta, \tag{91}$$

where $\mathsf{V}_{\rm el} = q\,\varphi$, and its eigenvalue equation

$$\begin{bmatrix} \left(\mathsf{V}^+ + mc^2 - E\right)\mathbf{1} & c\,(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ c\,(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & \left(\mathsf{V}^- - mc^2 - E\right)\mathbf{1} \end{bmatrix} \begin{bmatrix} \Phi_{\rm L}(\boldsymbol{r}) \\ \Phi_{\rm S}(\boldsymbol{r}) \end{bmatrix} = 0, \tag{92}$$

where $\mathsf{V}^\pm = \mathsf{V}_{\rm el} \pm \mathsf{V}_{\rm sc}$. In the following it is assumed that $q = -1$, i.e., the charge of the Dirac particle is equal to the charge of electron.

The external potentials modify spectrum of the Dirac operator. In particular, for a broad class of potentials, discrete eigenvalues corresponding to physically bound

states appear in the energy gap $(-mc^2, +mc^2)$. The eigenvalues of H^{D} may be defined as the critical values of the Rayleigh quotient

$$Q[\boldsymbol{\Phi}] = \frac{\langle \boldsymbol{\Phi} | \mathsf{H}^{\mathrm{D}} | \boldsymbol{\Phi} \rangle}{\langle \boldsymbol{\Phi} | \boldsymbol{\Phi} \rangle}, \tag{93}$$

where $\boldsymbol{\Phi}$ is in the domain of H^{D}. Since the Dirac operator is unbounded from below, the eigenvalues cannot be determined by a straightforward application, a minimization procedure based on the Hylleraas-Undheim-MacDonald theorem [23–26], commonly used in the non-relativistic quantum mechanics. However, over the last half a century, numerous variational methods of finding the eigenvalues have been elaborated [12–14, 23–25]. In particular, the lowest discrete eigenvalue in the energy gap, i.e. the physical ground state energy, may be derived from the *minimax principle*:

$$E = \min_{\{\Phi_{\mathrm{L}}\}} \left[\max_{\{\Phi_{\mathrm{S}}\}} Q[\boldsymbol{\Phi}] \right] \tag{94}$$

originally formulated as a recipe for reaching the stationary points of the energy hypersurface in the space of variational parameters [24, 26] and rigorously proved after several years [27]. The consecutive eigenvalues may be obtained in the usual way imposing appropriate orthogonality constraints on the wave function.

For some specific potentials, resonances embedded in the continuum, described by the wave functions containing both localized and continuum-type contributions and corresponding to auto-ionizing states, are also present [11, 28]. However, for most of physically significant potentials, the essential spectrum of the Dirac operator (74) remains unchanged (the essential spectrum of a Hermitian operator is a subset of its complete spectrum; its complement is the discrete spectrum, i.e., the set of isolated eigenvalues of finite multiplicity [11, 25]). In particular, the essential spectrum of a particle in an external field is the same as of the free particle if the potential vanishes in infinity. For a discussion of an external potential which modifies the essential spectrum, see section "Harmonic Oscillators."

Gauge Invariance

The gauge transformation of the electromagnetic four-vector potential is defined as

$$\mathsf{A}_\mu(\boldsymbol{r}) \;\mapsto\; \mathsf{A}_\mu(\boldsymbol{r}) + \frac{\partial f(\boldsymbol{r})}{\partial x_\mu}, \tag{95}$$

where $f(\boldsymbol{r})$ is a scalar function. Under this transformation, the electromagnetic fields remain unchanged. The results of measurements should be gauge independent, i.e., the physical contents of any theory cannot depend on the gauge. In particular,

also the Dirac equation should be gauge invariant. One can easily check that if the gauge transformation of the wave function is described by

$$\boldsymbol{\Psi}(\boldsymbol{x}) \;\mapsto\; \exp\left[i\ f(\boldsymbol{r})\right]\boldsymbol{\Psi}(\boldsymbol{x}), \tag{96}$$

then

$$\mathsf{D}_\mu\boldsymbol{\Psi}(\boldsymbol{x}) \;\mapsto\; \exp\left[i\ f(\boldsymbol{r})\right]\mathsf{D}_\mu\boldsymbol{\Psi}(\boldsymbol{x}). \tag{97}$$

Since scalar potential is gauge independent, Eqs. (96) and (97) imply gauge invariance of Eq. (89). However, note that the gauge transformations of $\boldsymbol{\Psi}$ and $\partial\boldsymbol{\Psi}/\partial x_\mu$ are inconsistent, and therefore the non-covariant derivative does not preserve the gauge symmetry of the Dirac equation.

The Non-relativistic Limit and the Direct Approach

After shifting the energy scale by mc^2, i.e., defining

$$\mathscr{E} = E - mc^2, \tag{98}$$

and performing some simple algebra, the eigenvalue equation (92) becomes

$$\begin{bmatrix} \left(\mathsf{V}^+ - \mathscr{E}\right)\mathbf{1} & (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & -2m\left[1 + (\mathscr{E} - \mathsf{V}^-)/(2mc^2)\right]\mathbf{1} \end{bmatrix} \begin{bmatrix} \Phi_{\mathrm{L}}(\boldsymbol{r}) \\ c\,\Phi_{\mathrm{S}(r)} \end{bmatrix} = 0. \tag{99}$$

In the non-relativistic limit, $(\mathscr{E} - \mathsf{V}^-)/(2mc^2) = 0$ and Eq. (99) transforms to

$$\begin{bmatrix} \left(\mathsf{V}^+ - \mathscr{E}\right)\mathbf{1} & (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & -2m\mathbf{1} \end{bmatrix} \begin{bmatrix} \Phi_{\mathrm{L}}(\boldsymbol{r}) \\ c\,\Phi_{\mathrm{S}}(\boldsymbol{r}) \end{bmatrix} = 0. \tag{100}$$

This equation, a non-relativistic limit of the Dirac equation defined in the same Hilbert space as the original Dirac equation, is known as the Lévy-Leblond equation. The corresponding equation of motion is Galilei covariant – it transforms according to a spinor representation of the Galilei group. The normalization conditions of the wave functions may be derived as follows: Given a Hamiltonian eigenvalue problem, $(\mathsf{H} - \Delta E)\Phi = 0$, where H and Δ are matrix operators. Then the normalization condition for Φ reads $\langle\Phi|\Delta|\Phi\rangle = 1$. Therefore, for the Dirac equations (92) and (99) $\langle\Phi_{\mathrm{L}}|\Phi_{\mathrm{L}}\rangle + \langle\Phi_{\mathrm{S}}|\Phi_{\mathrm{S}}\rangle = 1$, while for the Lévy-Leblond one (100), $\langle\Phi_{\mathrm{L}}|\Phi_{\mathrm{L}}\rangle = 1$.

Equation (100) yields

$$\Phi_{\mathrm{S}}(\boldsymbol{r}) = \frac{(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})}{2mc}\,\Phi_{\mathrm{L}}(\boldsymbol{r}), \tag{101}$$

i.e., if $|\boldsymbol{\sigma}\cdot\boldsymbol{\pi}| \ll mc$, then $|\Phi_{\mathrm{S}}| \ll |\Phi_{\mathrm{L}}|$. This justifies the names of these components of the Dirac wave function in the Pauli representation. The elimination of $c\Phi_{\mathrm{S}}$ from Eq. (100) gives

$$\left[\frac{(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}{2m} + \mathsf{V}^+ - \mathscr{E}\right]\Phi(\boldsymbol{r}) = 0, \tag{102}$$

i.e., the spin-dependent Schrödinger equation for $\Phi = \Phi_{\mathrm{L}}$. In the non-relativistic equation (102), the electrostatic and the scalar potentials are non-distinguishable since they contribute on an equal footing to V^+. Equation (102), with the help of the Dirac relation (14), may be transformed to the *Pauli equation*:

$$\left[\frac{1}{2m}\left(\boldsymbol{\pi}^2 + \boldsymbol{\sigma}\cdot\mathbf{B}\right) + \mathsf{V}^+ - \mathscr{E}\right]\Phi(\boldsymbol{r}) = 0, \tag{103}$$

where $\mathbf{B} = \nabla\times\mathbf{A}$ is the external magnetic field. It should be stressed that Eqs. (100) and (102) are equivalent *only* if relation (101) is fulfilled. This relation, known as the *kinetic balance condition* [12, 13], plays very important role in designing approximate methods of solving Dirac eigenvalue problems.

A Dirac equation similar to the Lévy-Leblond equation can be obtained by setting $\mathsf{V}^- = 0$. This may correspond to exact models with two kinds of fields: (1) a particle in an external magnetic field with $\mathsf{V}_{\mathrm{el}} = \mathsf{V}_{\mathrm{sc}} = 0$ and (2) a rather unusual special case of $\mathsf{V}_{\mathrm{el}} = \mathsf{V}_{\mathrm{sc}} \neq 0$. Another option is an approximation $\mathsf{V}^-/2mc^2 \mapsto 0$, originally proposed by Moore under the name *direct Dirac equation* [29]. The resulting equation reads

$$\begin{bmatrix} \left(\mathsf{V}^+ - \mathscr{E}\right)\mathbf{1} & (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) \\ (\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) & -2\,\mathfrak{M}\,\mathbf{1} \end{bmatrix}\begin{bmatrix} \Phi_{\mathrm{L}}(\boldsymbol{r}) \\ c\Phi_{\mathrm{S}}(\boldsymbol{r}) \end{bmatrix} = 0, \tag{104}$$

where

$$\mathfrak{M} = m\left(1 + \frac{\mathscr{E}}{2m}\alpha^2\right) \tag{105}$$

and $\alpha = 1/c$ is the fine structure constant. An intuitive rationale for the last equation can be derived from Eq. (17) which may be rewritten as $\mathscr{E} = p^2/(2\,\mathfrak{M})$. The elimination of Φ_{S} gives the direct Dirac equation in a Schrödinger-like form:

$$\left[\frac{(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})^2}{2\mathfrak{M}} + \mathsf{V}^+ - \mathscr{E}\right]\Phi_{\mathrm{L}}(\boldsymbol{r}) = 0. \tag{106}$$

Note that the difference between Φ_{L} of Eq. (106) an Φ of Eq. (102) is of a fundamental importance. In particular, Φ_{L} corresponding to different values of

$\mathscr{E}$ are not orthonormal since the Hamiltonian depends on $\mathscr{E}$. In this case the orthonormality condition is

$$\int_V \left[\Phi_{\mathrm{L}}^{\dagger}(\boldsymbol{r};\mathscr{E})\Phi_{\mathrm{L}}(\boldsymbol{r};\mathscr{E}') + \Phi_{\mathrm{S}}^{\dagger}(\boldsymbol{r};\mathscr{E})\Phi_{\mathrm{S}}(\boldsymbol{r};\mathscr{E}')\right] d^3\mathbf{r} = \delta(\mathscr{E}-\mathscr{E}'), \tag{107}$$

while in the non-relativistic one

$$\int_V \Phi^{\dagger}(\boldsymbol{r};\mathscr{E})\Phi(\boldsymbol{r};\mathscr{E}')\, d^3\mathbf{r} = \delta(\mathscr{E}-\mathscr{E}'), \tag{108}$$

where δ stands for either Dirac delta (if the spectrum is continuous) or the Kronecker one (if the spectrum is discrete).

Elimination of the Small Component

In the general case the elimination of Φ_{S} from the eigenvalue equation of H^{D} yields

$$\left[(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\frac{\mathsf{R}}{2\mathfrak{M}}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi}) + \mathsf{V}^{+} - \mathscr{E}\right]\Phi_{\mathrm{L}}(\boldsymbol{r}) = 0, \tag{109}$$

where

$$\mathsf{R} = \frac{1}{1-\mathsf{w}}, \tag{110}$$

$$\mathsf{w} = \frac{\mathsf{V}^{-}}{2\mathfrak{M}}\alpha^2, \tag{111}$$

and

$$\Phi_{\mathrm{S}}(\boldsymbol{r}) = \frac{\mathsf{R}\,\alpha}{2\mathfrak{M}}(\boldsymbol{\sigma}\cdot\boldsymbol{\pi})\Phi_{\mathrm{L}}(\boldsymbol{r}). \tag{112}$$

Since $\mathfrak{M}$ depends on $\mathscr{E}$, Eq. (109) is not an eigenvalue problem – the dependence on $\mathscr{E}$ is nonlinear. However, solutions $\mathscr{E}$ of this equation are equal to the eigenvalues of the corresponding Dirac operator, and the set of these solutions *is bounded from below* by the lowest eigenvalue of H_{D} located in the gap of its essential spectrum [25]. This property of Eq. (109) has important practical consequences. In particular, $\mathscr{E}$ may be obtained by minimization procedures which are technically difficult in the case of *basis-set* calculations but rather straightforward when one is using *grid* methods. It also explains the well-known (but not always correctly understood) property of the numerical Dirac-Fock algorithms: When solving the corresponding

equations, no effects of the unboundedness from below of the Dirac operator have ever been observed [14].

By applying the Dirac relation (14), one can separate the spin-dependent part [13]:

$$\left[\frac{\boldsymbol{\pi}\mathsf{R}\boldsymbol{\pi}}{2\mathfrak{M}}+\frac{i\,\boldsymbol{\sigma}}{2\mathfrak{M}}\left[\boldsymbol{\pi}\times\mathsf{R}\boldsymbol{\pi}\right]+\mathsf{V}^{+}-\mathscr{E}\right]\Phi_{\mathrm{L}}(\boldsymbol{r})=0. \tag{113}$$

The R-containing terms may be expressed as

$$\boldsymbol{\pi}\mathsf{R}\boldsymbol{\pi}=\mathsf{R}\boldsymbol{\pi}^{2}+\mathsf{R}^{2}(\mathbf{p}\,\mathsf{w})\boldsymbol{\pi} \tag{114}$$

$$i\boldsymbol{\sigma}[\boldsymbol{\pi}\times\mathsf{R}\boldsymbol{\pi}]=\mathsf{R}(\boldsymbol{\sigma}\cdot\boldsymbol{B})+\mathsf{R}^{2}\boldsymbol{\sigma}\left[(\nabla\mathsf{w})\times\boldsymbol{\pi}\right]. \tag{115}$$

It should be stressed that Eq. (109) [as well as (113)] is exact and, jointly with Eq. (112), is equivalent to the original Dirac equation.

Pauli Corrections

Equation (113) may be expressed in a form correct up to terms proportional to α^2. By the substitution

$$\mathsf{R}=1+\alpha^{2}\frac{\mathsf{V}^{-}}{2m}+\mathscr{O}\left(\alpha^{4}\right),\qquad \mathsf{w}=\alpha^{2}\frac{\mathsf{V}^{-}}{2m}+\mathscr{O}\left(\alpha^{4}\right),\qquad \frac{1}{\mathfrak{M}}=\frac{1}{m}-\alpha^{2}\frac{\mathscr{E}}{2m^{2}}+\mathscr{O}\left(\alpha^{4}\right)$$

Eq. (113) becomes

$$\left[\frac{1}{2m}\left(\boldsymbol{\pi}^{2}+\boldsymbol{\sigma}\cdot\,\mathbf{B}\right)+\mathsf{V}^{+}+\mathsf{H}'+\mathscr{O}\left(\alpha^{4}\right)-\mathscr{E}\right]\Phi_{\mathrm{L}}(\boldsymbol{r})=0, \tag{116}$$

where

$$\mathsf{H}'=\frac{\alpha^{2}}{4m^{2}}\left[(\mathsf{V}^{-}-\mathscr{E})\left(\boldsymbol{\pi}^{2}+\boldsymbol{\sigma}\cdot\,\mathbf{B}\right)+(\,\mathbf{p}\,\mathsf{V}^{-})\,\boldsymbol{\pi}+\boldsymbol{\sigma}\left[(\nabla\,\mathsf{V}^{-})\times\boldsymbol{\pi}\right]\right] \tag{117}$$

is the well-known Pauli relativistic correction with the consecutive terms corresponding to the mass-velocity, Darwin, and spin-orbit terms [1, 5] in which the interaction with an external magnetic field is included. In principle, H' may be treated as the first-order perturbation to the Schrödinger energy only. However, in some approaches it is transformed in a way which allows to include them to the Hamiltonian eigenvalue equation. In particular, some of the non-integrable singularities which originate from the Coulomb potential may be eliminated by introducing a finite-size nucleus [12, 13]. The transformation of H' to the Hermitian form may be found, e.g., in [1].

Gordon Decomposition

A similar separation of the spin-dependent term may be performed for the four-current (51). After some algebra described in detail in [4, 5], one obtains the separation of the four-current to a *convection current* j_μ^{C} and a *spin current* j_μ^{S}:

$$j_\mu = j_\mu^{\mathrm{C}} + j_\mu^{\mathsf{S}}, \tag{118}$$

where

$$j_\mu^{\mathrm{C}} = \frac{1}{2m}\left[\overline{\boldsymbol{\Psi}}\left(\mathsf{p}_\mu + \mathsf{A}_\mu\right)\boldsymbol{\Psi} - \left(\left(\mathsf{p}_\mu - \mathsf{A}_\mu\right)\overline{\boldsymbol{\Psi}}\right)\boldsymbol{\Psi}\right] \tag{119}$$

and

$$j_\mu^{\mathsf{S}} = -\frac{i}{2m}\mathsf{p}_\nu(\overline{\boldsymbol{\Psi}}\sigma_{\nu\mu}\boldsymbol{\Psi}). \tag{120}$$

In three-vector notation, Eq. (119) may be rewritten as

$$\mathbf{j}^{\mathrm{C}} = \frac{1}{2m}\left[\overline{\boldsymbol{\Psi}}\boldsymbol{\pi}\,\boldsymbol{\Psi} + (\boldsymbol{\pi}^*\overline{\boldsymbol{\Psi}})\,\boldsymbol{\Psi}\right], \tag{121}$$

and if $\boldsymbol{\Psi}$ corresponds to a given energy E,

$$\rho^{\mathrm{C}} = \frac{E - \mathsf{V}_{\mathrm{el}}}{mc^2}\,\overline{\boldsymbol{\Psi}}\,\boldsymbol{\Psi}. \tag{122}$$

Similarly,

$$\mathbf{j}^{\mathrm{S}} = \frac{1}{2m}\left[\nabla\times(\overline{\boldsymbol{\Psi}}\boldsymbol{\Sigma}\boldsymbol{\Psi}) + \frac{E}{c}\left(\overline{\boldsymbol{\Psi}}\boldsymbol{\alpha}\boldsymbol{\Psi}\right)\right] \tag{123}$$

and

$$\rho^{\mathrm{S}} = -\frac{1}{2mc}\,\mathbf{p}\left(\overline{\boldsymbol{\Psi}}\boldsymbol{\alpha}\boldsymbol{\Psi}\right). \tag{124}$$

In the non-relativistic limit, i.e., if $E - \mathsf{V}_{\mathrm{el}} \to mc^2$ and $\Psi_{\mathrm{S}} \to 0$, then $\rho^{\mathrm{C}} \to \Psi_{\mathrm{L}}^\dagger\Psi_{\mathrm{L}}$ and $\rho^{\mathrm{S}} \to 0$ (cf. section "Bilinear Covariants $\overline{\boldsymbol{\Psi}}\,\Gamma\,\boldsymbol{\Psi}$").

Charge Conjugation, Positrons

The Dirac equation is valid for all particles with spin $1/2$, in particular for electrons and for positrons. If Eq. (89) describes electrons, then the equation for positrons

may be obtained by the replacement $-e \mapsto +e$ in the expression for the covariant derivative. This transformation is related to another symmetry of the Dirac equation referred to as *charge conjugation* [3–5]. Some simple algebra shows that the charge conjugation operator is equal to

$$\mathsf{C} = \gamma_2 \mathsf{K}_0, \tag{125}$$

where K_0 is the operator of complex conjugation. The form of the charge conjugation operator depends on the representation of the Dirac equation (contrary to the parity (63) operator, which is representation independent). For example, in the Majorana representation, $\mathsf{C} = \mathsf{K}_0$.

If Eq. (89) is symbolically expressed as

$$\Omega_{-e} \boldsymbol{\Psi}_{\text{electron}} = 0,$$

then $\left(\mathsf{C}\,\Omega_{-e}\mathsf{C}^{-1}\right)\mathsf{C}\boldsymbol{\Psi}_{\text{electron}} = 0$ is equivalent to

$$\Omega_{+e} \boldsymbol{\Psi}_{\text{positron}} = 0,$$

where $\Omega_{+e} = \mathsf{C}\,\Omega_{-e}\mathsf{C}^{-1}$ and $\boldsymbol{\Psi}_{\text{positron}} = \mathsf{C}\,\boldsymbol{\Psi}_{\text{electron}}$. Thus,

$$\boldsymbol{\Psi}_{\text{positron}} = \mathsf{C}\,\boldsymbol{\Psi}_{\text{electron}} \quad \text{and} \quad \boldsymbol{\Psi}_{\text{electron}} = \mathsf{C}\,\boldsymbol{\Psi}_{\text{positron}}, \tag{126}$$

where the last equation is a consequence of $C^2 = \mathbf{1}$. The application of the charge conjugation to the free-particle spinors (77) gives

$$\begin{aligned} \mathsf{C}\,\boldsymbol{\Phi}^{+}_{p,1} &= -\boldsymbol{\Phi}^{-}_{-p,-1}, \quad & \mathsf{C}\,\boldsymbol{\Phi}^{+}_{p,-1} &= \boldsymbol{\Phi}^{-}_{-p,1}, \\ \mathsf{C}\,\boldsymbol{\Phi}^{-}_{p,1} &= \boldsymbol{\Phi}^{+}_{-p,-1}, & \mathsf{C}\,\boldsymbol{\Phi}^{-}_{p,-1} &= -\boldsymbol{\Phi}^{+}_{-p,1}. \end{aligned} \tag{127}$$

Combining with Eq. (69), one can see that the transformation electron $\Leftrightarrow$ positron is equivalent to

$$\{E, \boldsymbol{p}, \boldsymbol{s}, h\} \quad \Leftrightarrow \quad \{-E, -\boldsymbol{p}, -\boldsymbol{s}, h\}. \tag{128}$$

Relation (128) implies that a state with negative energy corresponds to the antiparticle with positive energy. The energy, charge, momentum, and spin of the antiparticle are opposite to that of the particle, while the current, helicity, and velocity are the same [cf. Eqs. (52), (75), (83)].

However, the interpretation of the negative-energy states of the electronic Dirac equation as describing positrons meets some difficulties [11]. First, there exist physically valid quantum states of electrons which are superpositions of the positive- and negative-energy eigenfunctions. Second, in the quantized Dirac field theory, the field operators anticommute and this results in the change of sign of the current under charge conjugation.

Exactly Solvable Models

The exact solutions of the Dirac and Klein-Gordon equations for a free-particle, hydrogen-like atom, a particle in an external magnetic field, were obtained as a part of the process of formulation of relativistic quantum mechanics, and the results were crucial to the acceptance of the new theory. Detailed analysis of these solutions may be found in all textbooks of relativistic quantum mechanics, in particular in [1–5]. In this section solutions of several models are briefly discussed as illustrations of some selected approaches and, in general, do not replicate the textbook presentations. Three models, representing two different classes of exactly solvable problems, are discussed: (1) two modes of relativistic generalizations of the harmonic oscillator (HO) (one known as *a particle in a homogeneous magnetic field* corresponding to an axial HO and the second one referred to as *the Dirac oscillator* corresponding to a spherical HO) and (2) a charged Dirac particle in the Coulomb field, in the case of an electron referred to as "the hydrogen-like atom," represents problems with spherical symmetry.

Harmonic Oscillators

Relativistic generalizations of harmonic oscillator are neither unique nor trivial. The simplest approach: setting $\mathbf{A} = 0$, $\mathsf{V}_{\rm sc} = 0$, and $\mathsf{V}_{\rm el} \sim r^2$ leads to a quartic oscillator with no bound solutions since the effective potential approaches $-\infty$ when $r \to \infty$. A harmonic electrostatic potential describes a charged particle within an infinite uniformly charged sphere. This is rather unphysical and one may expect that a scalar rather than electrostatic potential is more appropriate to model the harmonic-type interactions. Indeed, a quartic oscillator with bound-state solutions results from $\mathsf{V}_{\rm el} = 0, \mathsf{V}_{\rm sc} \sim r^2$. None of these models can be solved analytically. However, two other models are exactly solvable. In both the electrostatic and scalar potentials are set equal to zero. In the first model, corresponding to an axial oscillator, the vector potential describes a homogeneous magnetic field. In the second one, a non-electrostatic kind of interaction leads to a spherical oscillator. Very recently an experimental realization of the second model has been discovered [30].

With $\mathsf{V}^+ = \mathsf{V}^- = 0$, Eq. (99) yields

$$\begin{bmatrix} -\mathscr{E}\,\mathbf{1} & c\,\boldsymbol{\sigma}\cdot(\,\mathbf{p}+\,\mathbf{A}^\dagger) \\ c\,\boldsymbol{\sigma}\cdot(\,\mathbf{p}+\,\mathbf{A}) & -2\mathfrak{M}c^2\,\mathbf{1} \end{bmatrix} \begin{bmatrix} \Phi_{\rm L}(\boldsymbol{r}) \\ \Phi_{\rm S}(\boldsymbol{r}) \end{bmatrix} = 0. \tag{129}$$

From here one gets

$$\Phi_{\rm S}(\boldsymbol{r}) = \frac{\boldsymbol{\sigma}\cdot(\,\mathbf{p}+\,\mathbf{A})}{2\mathfrak{M}c}\,\Phi_{\rm L}(\boldsymbol{r}) \tag{130}$$

and, using the Dirac relation (14),

$$\left[\frac{1}{2m}(\mathbf{p}+\mathbf{A}^{\dagger})(\mathbf{p}+\mathbf{A})+\frac{(\boldsymbol{\sigma}\cdot\mathbf{M})}{2m}-\mathfrak{E}\right]\Phi_{\mathrm{L}}(\boldsymbol{r})=0, \tag{131}$$

where

$$\mathfrak{E}=\mathscr{E}\frac{\mathfrak{M}}{m}=\mathscr{E}\left(1+\frac{\mathscr{E}}{2mc^2}\right) \tag{132}$$

and

$$\mathbf{M}=\nabla\times\mathbf{A}+i(\mathbf{A}^{\dagger}\times\mathbf{A})+\left(\mathbf{A}^{\dagger}-\mathbf{A}\right)\times\nabla. \tag{133}$$

There is a subtle interplay between Galilei and Lorentz covariance of these equations. Equation (131) transforms according to a spinor representation of the Galilei group and, thus, is Galilei covariant with the normalization condition for the two-component wave function: $\langle\Phi_{\mathrm{L}}|\Phi_{\mathrm{L}}\rangle=1$. On the other hand, the pair of Eqs. (131) and (130) is equivalent to Eq. (129), an eigenvalue equation derived from the Lorentz-covariant Dirac equation (the Lorentz-covariant equation is obtained when $\mathscr{E}\mapsto mc^2+i\,\partial/\partial t$) with the normalization condition for the four-component wave function: $\langle\Phi_{\mathrm{L}}|\Phi_{\mathrm{L}}\rangle+\langle\Phi_{\mathrm{S}}|\Phi_{\mathrm{S}}\rangle=1$. Note that in this case solutions of the Schrödinger equation may be used to generate the corresponding solutions of the Dirac equation. In particular, the analytical solubility of the Schrödinger equation implies that also the corresponding Dirac equation is analytically solvable. The dependence of the eigenvalues $\mathfrak{E}$ of the Hamiltonian in Eq. (131) on E is shown in Fig. 1. The spectrum is bounded from below; the states with the same $|E|$ are degenerate and in the non-relativistic limit $\mathfrak{E}\mapsto\mathscr{E}$.

Electron in a Magnetic Field

If $\mathbf{A}$ describes a magnetic field $\mathbf{B}$, then $\nabla\times\mathbf{A}=\mathbf{B}$ and $\mathbf{A}=\mathbf{A}^{\dagger}$. For a homogeneous magnetic field, $\mathbf{A}=\mathbf{B}\times\mathbf{r}/2$ and Eq. (131) becomes

$$\left[\frac{\mathbf{p}^2}{2m}+\frac{\mathsf{B}^2}{8m}\left[\mathsf{r}^2-(\mathbf{r}\cdot\hat{\mathsf{B}})^2\right]-\boldsymbol{\mu}\cdot\mathbf{B}-\mathfrak{E}\right]\Phi_{\mathrm{L}}(\boldsymbol{r})=0, \tag{134}$$

where $\hat{\mathsf{B}}=\mathbf{B}/\mathsf{B}$ is the unit vector of $\mathbf{B}$ direction,

$$\boldsymbol{\mu}=\boldsymbol{\mu}_{\mathrm{orbit}}+\boldsymbol{\mu}_{\mathrm{spin}}=-\frac{e}{2m}(\mathbf{L}+g\,\mathbf{S}), \tag{135}$$

is the magnetic moment of the particle,

$$\mathbf{L}=\mathbf{r}\times\mathbf{p},\qquad\mathbf{S}=\frac{1}{2}\boldsymbol{\sigma}, \tag{136}$$

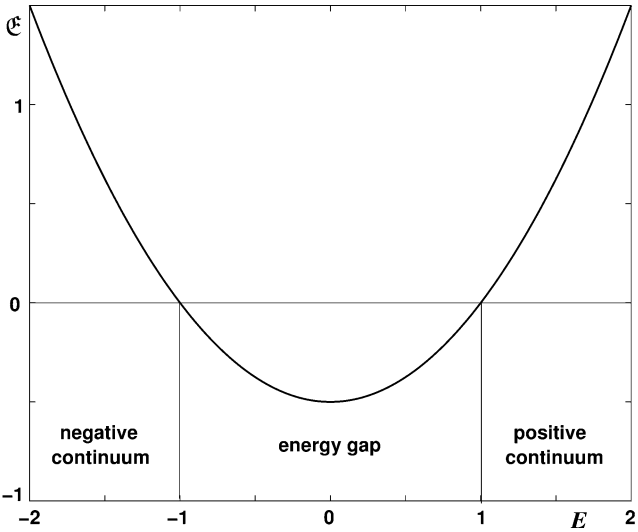


Fig. 1 Eigenvalues $\mathfrak{E}$ of Eq. (131) versus energy $E = \mathscr{E} + mc^2$ in units of mc^2. The eigenvalues are bounded from below by $-mc^2/2$ and the minimum corresponds to $E = 0$. The positive- and the negative-energy continua are mapped to $\mathfrak{E} > 0$. The states Φ_{L} differing by the sign of energy λ, i.e., corresponding to the same values of $|E|$, are degenerate

and g is the *spin g-factor*. The non-relativistic and relativistic quantum mechanics predicts $g = 2$ for all spin-1/2 particles. The experimental values (up to seven significant figures) for electron and for muon are, respectively, $g_{\mathrm{e}} = 2.002319$ and $g_\mu = 2.002332$. The deviation from 2 is due to QED effects and is very precisely predicted by the theory. For composite particles, as proton or neutron, the values of g are substantially different from 2.

If the coordinate system is selected so that $\mathbf{B} = \{0, 0, \mathsf{B}\}$, then the Hamiltonian commutes with S_3 and Φ_{L} may be expressed as a normalized two-component spinor

$$\Phi_{\mathrm{L}}(\mathbf{r}, \sigma) \equiv \Phi_{\mathsf{n}\mathsf{m}_s}(\mathbf{r}, \sigma) = \phi_{\mathsf{n}}(\mathbf{r})\chi_{\mathsf{m}_s}(\sigma), \quad \chi_{1/2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi_{-1/2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{137}$$

with $\langle \Phi_{\mathsf{n}\mathsf{m}_s} | \Phi_{\mathsf{n}'\mathsf{m}'_s} \rangle = \delta_{\mathsf{n}\mathsf{n}'}\delta_{\mathsf{m}_s\mathsf{m}'_s}$, where $\mathsf{n} = 0, 1, 2, \ldots$ is the principal quantum number and $\mathsf{m}_s = \pm 1/2$ corresponds to two projections of spin. In this case Eq. (134) reduces to the cylindrical harmonic oscillator equation. The eigenvalues (i.e., the non-relativistic energies) are given by

$$\mathfrak{E}_{\mathsf{n}\mathsf{m}_s} = \frac{\mathsf{B}}{m}\left(\mathsf{n} + \mathsf{m}_s + \frac{1}{2}\right) + \frac{p_z^2}{2m}. \tag{138}$$

For a detailed analysis, see, e.g., [1–3]. The relativistic energies may be obtained from Eq. (132)

$$E = \mathscr{E} + mc^2 = \pm mc^2 \sqrt{1 + \frac{2\mathfrak{E}}{mc^2}}. \tag{139}$$

In this case the essential spectrum is different than the one of the free Dirac operator. An infinite set of discrete energy levels numbered by $\mathsf{n}+\mathsf{m}_s$ spreads from $+mc^2[1+(p_z/mc)^2]^{1/2}$ to $+\infty$ and from $-mc^2[1+(p_z/mc)^2]^{1/2}$ to $-\infty$. The corresponding normalized four-component wave functions can be derived using Eq. (134). After some algebra, one obtains

$$\boldsymbol{\Phi}^{\lambda}_{\mathsf{n}\mathsf{m}_s} = \frac{1}{\sqrt{2\tau}} \begin{bmatrix} \sqrt{\tau+\lambda}\,\chi_{\mathsf{m}_s} \\ \frac{\lambda}{\sqrt{\tau+\lambda}} \left(\frac{\boldsymbol{\sigma}\cdot\boldsymbol{\pi}}{mc}\right) \chi_{\mathsf{m}_s} \end{bmatrix} \phi_{\mathsf{n}}(\mathbf{r}), \quad \tau = \sqrt{1 + \frac{2\mathfrak{E}}{mc^2}}. \tag{140}$$

Dirac Oscillator

One of many possible relativistic analogs of the non-relativistic harmonic oscillators, originally proposed by Cook [31] and then rediscovered by Moshinsky and Szczepaniak [32], is based on the construction of an exactly solvable Dirac equation which in the non-relativistic limit gives the Schrödinger harmonic oscillator. This kind of oscillator may also be defined as a system which is invariant with respect to a canonical transformation interchanging coordinates and momenta, i.e., its eigenvalue problem looks the same in the coordinate and in the momentum representations [33].

Equation (129) is invariant under the transformation $\mathbf{p} \leftrightarrow \mathbf{r}$ if $\mathbf{A} \sim i\,\mathbf{r}$. In particular, if $\mathbf{A} = -i\,m\omega\,\mathbf{r}$, then

$$\mathbf{M} = -2m\omega\,[\mathbf{r}\times\mathbf{p}] = -2m\omega\,\mathbf{L} \tag{141}$$

and Eq. (131) yields

$$\left[\frac{\mathbf{p}^2}{2m} + \frac{m\omega^2}{2}\,\mathbf{r}^2 + \omega\left(\mathbf{L}^2 - \mathbf{J}^2\right) - \frac{3}{4}\omega - \mathfrak{E}\right] \Phi_{\mathrm{L}} = 0, \tag{142}$$

where $\mathbf{J} = \mathbf{L} + \mathbf{S}$ and the identities $\boldsymbol{\sigma}\cdot\mathbf{L} = \mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2$, $\mathbf{S}^2 = (3/4)\,\mathbf{1}$ have been used. Further analysis is similar to the one of Eq. (134). Details may be found in [33].

Central Force

The description of the behavior of a Dirac particle in a spherically symmetric potential is a prerequisite for the construction of relativistic theories of atomic structure. The potential is centered at the origin of the coordinate system, and

the corresponding Dirac operator is obtained by setting $\mathbf{A} = 0$, $\mathsf{V}_{\rm sc} = 0$, and $\mathsf{V}_{\rm el} = \mathsf{V}(r)$. The set of commuting operators comprises H, $\mathbf{J}^2$, J_3, P, and K, where

$$\mathsf{K} = \beta\,(\boldsymbol{\Sigma}\cdot\mathbf{L}+\mathbf{1}) = \begin{bmatrix} \boldsymbol{\sigma}\cdot\mathbf{L}+\mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\sigma}\cdot\mathbf{L}-\mathbf{1} \end{bmatrix}. \tag{143}$$

The corresponding eigenvalues are denoted E, $\mathsf{j}(\mathsf{j}+1)$, m_j, $\mathsf{p} = \pm 1$, and k, where

$$\mathsf{k} = \epsilon\left(\mathsf{j}+\frac{1}{2}\right) \quad \text{and} \quad \epsilon = 2(\mathsf{j}-\mathsf{l}) = \frac{\mathsf{k}}{|\mathsf{k}|}. \tag{144}$$

Note that some authors, e.g., [4, 12, 13], use $\kappa = -\mathsf{k}$ instead.

Neither the square nor the projection of $\mathbf{L}$ commutes with the Dirac Hamiltonian. Therefore, the Dirac wave function cannot be an eigenfunction of either L^2 or L_3. However, in the Dirac-Pauli representation, large and small components of the wave function are eigenfunctions of L^2, to the eigenvalues $\mathsf{l}^{\rm L}(\mathsf{l}^{\rm L}+1)$ and $\mathsf{l}^{\rm S}(\mathsf{l}^{\rm S}+1)$, respectively, with $\mathsf{l} = \epsilon\mathsf{k} - (\epsilon+1)/2$ and $\mathsf{l}^{\rm S} = \mathsf{l}^{\rm L} + \epsilon$. Due to this feature, the Dirac-Pauli representation is particularly useful in describing the angular dependence in the spherically symmetric problems, and $\mathsf{l} \equiv \mathsf{l}^{\rm L}$ is used as an auxiliary label of the eigenfunctions. The common eigenfunctions of $\mathbf{J}^2$, J_3, and K may be expressed as [1–5]

$$\boldsymbol{\Phi}_{\mathsf{nkm}_j}(r,\theta,\phi) = \begin{bmatrix} \Phi_{\rm L}(r,\theta,\phi) \\ \Phi_{\rm S}(r,\theta,\phi) \end{bmatrix} = \frac{1}{r}\begin{bmatrix} \mathscr{Y}^{\mathsf{l}}_{\mathsf{km}_j}(\theta,\phi) & G_{\mathsf{nk}}(r) \\ i\,\mathscr{Y}^{\mathsf{l}+\epsilon}_{-\mathsf{km}_j}(\theta,\phi) & F_{\mathsf{nk}}(r) \end{bmatrix} \tag{145}$$

where $\mathscr{Y}^{\mathsf{l}}_{\mathsf{km}_j}$ may be constructed by coupling $\mathsf{m}_s = \pm 1/2$ spin eigenfunctions with spherical harmonics $Y_{\mathsf{l},\mathsf{m}_j\pm 1/2}$. Finally, the eigenvalue equation of the Dirac operator may be reduced to the following system of two radial equations:

$$\begin{bmatrix} \mathsf{V}(r)+mc^2-E, & c\left(-\frac{d}{dr}+\frac{\mathsf{k}}{r}\right) \\ c\left(\frac{d}{dr}+\frac{\mathsf{k}}{r}\right), & \mathsf{V}(r)-mc^2-E \end{bmatrix}\begin{bmatrix} G_{\mathsf{nk}}(r) \\ F_{\mathsf{nk}}(r) \end{bmatrix} = 0. \tag{146}$$

Coulomb Potential

Usually the problem of an electron moving in the field of a Coulomb potential is referred to as "hydrogen-like atom." In fact, a quantum-mechanical treatment of a hydrogen-like atom is equivalent to solving a two-body problem with a Coulomb interaction. In the non- relativistic case, it may be solved exactly (assuming the interacting particles are point-like). By the separation of the center of mass, the six-dimensional Schrödinger equation is split into two three-dimensional equations. One of them describes a free motion of the center of mass and the other one the relative motion of the electron and the nucleus. In the relativistic case, each particle has its own time and the interaction is not instant but spreads with the velocity of

light. Therefore, a relativistic problem of two interacting particles cannot be solved exactly, and the hydrogen-like atom in the Dirac theory is modeled by an electron moving in an external Coulomb field with $\mathsf{V} = -Z/r$, where Z is the "nuclear charge." The wave functions are usually expressed in the Dirac-Pauli representation and are labeled by n, j, k (or l), and m_j.

Both large and small radial components may be expressed as products of $\mathrm{e}^{-\zeta r}$ and a polynomial, where $\zeta = \sqrt{m^2c^2 - E^2/c^2}$. At the origin $G(r)$ and $F(r)$ behave as r^s, where $s = \sqrt{k^2 - Z^2\alpha^2}$. Then, for $|\mathsf{k}| > 1$ the wave function vanishes for $r = 0$. However, for $\mathsf{k} = \pm 1$, $s < 1$ and the radial function is singular. This singularity is weak and for $Z\alpha < 1$ does not obstruct the normalizability of the wave functions. The relativistic radial electron density is contracted relative to the non-relativistic one (the Dirac hydrogen atom is "smaller" than the Schrödinger one). Besides, it is nodeless since the nodes of the large and small radial components (except the $r = 0$ node) never coincide.

The Dirac operator of Eq. (146) for $\mathsf{V} = -Z/r$ has a hidden supersymmetry. Therefore, the system of radial equations, transformed to the second order, can be separated [10, 11] giving in effect two Schrödinger-like equations:

$$\left[-\frac{d^2}{d\rho^2} + \frac{s(s \pm 1)}{\rho^2} - \frac{2Z}{\rho} - 2\mathfrak{e}\right] R^{\pm}_{\mathsf{nk}} = 0 \tag{147}$$

where $\rho = r\,E/mc^2$,

$$\mathfrak{e} = mc^2\left(1 - \frac{m^2c^4}{E^2}\right) \tag{148}$$

$R^{\pm}_{\mathsf{nk}} = q_{\pm}\,G_{\mathsf{nk}} + q_{\mp}\,F_{\mathsf{nk}}$, $q_+ = \sqrt{(|k| + s)/2s}$, $q_- = (k/|k|)\sqrt{(|k| - s)/2s}$. The spectrum is bounded from below since the negative continuum is shifted to the positive-energy space and overlaps with the positive continuum. The discrete eigenvalues are equal to

$$\mathfrak{e}_{\mathsf{n}|\mathsf{k}|} = -\frac{Z^2}{2\tilde{\mathsf{n}}^2} \tag{149}$$

where $\tilde{\mathsf{n}} = \mathsf{n} + s - |k|$. By combining Eqs. (148) and (149), one can get a simple expression for the discrete energy levels

$$\mathscr{E}_{\mathsf{n}|\mathsf{k}|} = -\frac{Z^2}{\mathsf{N}(\mathsf{N} + \tilde{\mathsf{n}})}, \tag{150}$$

where $\mathsf{N} = \sqrt{\alpha^2 Z^2 + \tilde{\mathsf{n}}^2}$. The last equation is equivalent to the well-known Sommerfeld formula. For $\alpha \to 0$, i.e., in the non-relativistic limit, $s \to |\mathsf{k}|$, $\tilde{\mathsf{n}} \to \mathsf{n}$, $N \to \mathsf{n}$. The expansion of the Dirac energy into a power series of $(\alpha Z)^2$ yields

$$E_{n|k|} = mc^2 \left[1 - \frac{(\alpha Z)^2}{2n^2} - \frac{(\alpha Z)^4}{2n^3} \left(\frac{1}{|k|} - \frac{3}{4n} \right) + \mathcal{O}\left(\alpha^6 Z^6\right) \right], \tag{151}$$

where the first term is the rest energy of the electron, the second one is the non-relativistic (Schrödinger) energy, and the third one is the relativistic (Pauli) correction.

In the Dirac energy spectrum, apart from shifting the energy levels relative to the Schrödinger ones, some of the degeneracies are removed. The Dirac energy levels depend on $|k|$ but not on its sign. In other words, they depend on the total angular momentum quantum number j but not on l. Then, the energies of the states $2p_{1/2}$ and $2p_{3/2}$ are different, while the energies of $2p_{1/2}$ are the same as that of $2s_{1/2}$. The splitting of the energy levels due to j is called the *fine structure splitting*. It is relatively small for small Z, but it grows very fast with increasing nuclear charge (it is proportional to Z^4). Experimental measurements show that the energies of a real hydrogen-like atom depend on l. In particular the energies of $2s_{1/2}$ and $2p_{1/2}$ are different. For the first time this splitting was measured for the hydrogen atom in 1947 by Willis Lamb and Robert Retherford. The effect is called the *Lamb shift*. It is explained on the ground of quantum electrodynamics. Its value grows rapidly with increasing Z. For the hydrogen atom, it is equal to 1058 MHz $\approx 4 \cdot 10^{-6}$ eV; for the hydrogen-like uranium, it is about 468 eV, i.e., by eight orders of magnitude larger than in hydrogen!

The effects resulting from the finite size of the nucleus, particularly important for heavy ions, may be treated numerically [12–14]. Their influence on the structure of the electronic spectrum supplies information about the distribution of the nuclear charge and about the nuclear shape. The coupling between the nuclear and the electronic angular momenta is responsible for additional splitting of the energy spectrum – the hyperfine structure.

Two Dirac Particles

A many-electron time-evolution equation (2) cannot be Lorentz covariant – in a covariant theory each particle has to have its own time coordinate. Besides, if the particles interact, then the interaction is mediated by the virtual photons and spreads with the velocity of light. The corresponding operators have to describe the coupling with the electromagnetic field and cannot be expressed in a closed form. Also some introductory steps, like the separation of the motion of the center of mass, cannot be performed in a Lorentz-covariant model since the center of mass is not well defined in the relativistic mechanics – one has to use the *center of momentum* instead. Nevertheless, *relativistic quantum chemistry* is derived from many-electron generalizations of the Dirac equation [12]. The most common approaches stem from what is called the *Dirac-Coulomb* (DC) equation – the eigenvalue equation of the n-electron DC Hamiltonian

$$\left[\mathsf{H}_n^{\mathrm{DC}} - E_{\mathsf{q}}\right] \boldsymbol{\Phi}_{\mathsf{q}}^{(n)} = 0, \tag{152}$$

where q labels the eigenstates and $\boldsymbol{\Phi}_{\mathsf{q}}^{(n)}$ is the n-electron eigenfunction (in order to simplify the notation, the energy and the wave functions are labeled by the same index, i.e., the degeneracy is not shown explicitly). The DC Hamiltonian, in a simplified notation, may be expressed as

$$\mathsf{H}_n^{\mathrm{DC}}(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_n) = \sum_{j=1}^{n} \mathsf{H}_1^{\mathrm{D}}(\boldsymbol{r}_j) + \sum_{i<j}^{n} \frac{1}{r_{ij}}. \tag{153}$$

It is a rather awkward composition of the Dirac one-body operators $\mathsf{H}_1^{\mathrm{D}}$ with electrostatic potential generated by fixed nuclei and a non-relativistic Coulomb interaction term. A theory derived from this Hamiltonian is evidently non-covariant. Also after the interaction operator is supplemented by the retardation and magnetic corrections, the covariance is limited to the terms proportional to α^2.

The eigenvalue problem of the Dirac-Coulomb Hamiltonian, additionally to being unbounded from below, suffers from another mathematical inconvenience: the discrete and continuous spectra of its one-electron part overlap. The wave functions describing the discrete states are coupled by the interaction term to the ones describing the continua. In effect, the spectrum of the DC Hamiltonian does not contain any bound states. All its solutions either belong to a continuum or are auto-ionizing. This property of the DC eigenvalue problem, discovered by G. E. Brown and D. G. Ravenhall [34], is referred to as the Brown-Ravenhall (BR) disease, and the continuum which spreads over the entire energy range (from $-\infty$ to $+\infty$) is known as the *Brown-Ravenhall continuum*. Most common escape is to project the Hamiltonian to the positive-energy subspace of the complete Hilbert space. In practical terms, this means that the many-electron basis, in which the DC Hamiltonian is represented, is constructed as the Kronecker products of the one-electron Dirac spinors corresponding to the positive-energy-state solutions of a Dirac equation. In such a space the algebraic representation of the DC Hamiltonian corresponds to an operator bounded from below and is free from the BR disease. The projection method offers an efficient and conceptually simple solution, but it carries several drawbacks. First, the projected variational space can never approach completeness. Second, the results of the projection depend upon the choice of the one-electron Dirac Hamiltonian. Third, an application of this approach in the case of explicitly correlated wave functions is hardly possible. In practical terms, all these difficulties have been solved [12–14]. However, the approach cannot be considered as completely satisfactory from either mathematical or practical perspective.

New directions aimed at linking relativistic quantum mechanics and quantum electrodynamics have recently been formulated by Lindgren and Liu in a very recent series of papers [16–18]. A complete description of this subject is given in "Relativistic Hamiltonians" section of this handbook.

Non-interacting Particles in an Electrostatic Field

The n-particle Dirac operator describing non-interacting particles in an external field is referred to as the n-particle Dirac Hamiltonian. It is defined in the n-particle Hilbert space

$$\mathscr{H}_n^{\mathrm{D}} = \mathscr{H}_1^{\mathrm{D}\otimes n} = L^2\left(\mathbb{R}^{3n}\right) \otimes \mathbb{C}^{4^n}. \tag{154}$$

According to a definition more precise than Eq. (153), it is a direct sum of n one-particle Hamiltonians. In the coordinate representation, it is a first-order differential operator represented in the spinor space by a $4^n \times 4^n$ matrix. Its main features may be seen already for $n = 2$, and therefore, for simplicity, the further discussion is limited to two particles. A reader interested in general properties of n-particle models is referred to monographs on relativistic many-body problems ([15] and references therein). Thus, in the case of two identical particles, the Dirac Hamiltonian may be expressed as

$$\mathsf{H}_2 = \mathsf{H}_1 \oplus \mathsf{H}_1 = \mathsf{H}_1 \otimes \mathbf{1} + \mathbf{1} \otimes \mathsf{H}_1, \tag{155}$$

where H_1 and H_2 denote, respectively, one- and two-particle operators and $\mathbf{1}$ is the one-particle unit operator. One should remember that in a direct sum or in a direct product, the particle number is defined by the position of the corresponding operator. Thus, $\mathsf{H}_1 \otimes \mathbf{1}$ means that H_1 acts in the space of particle 1 and $\mathbf{1}$ – in the space of particle 2. If the particles move in the field of an external electrostatic potential

$$\mathsf{V}_2(\boldsymbol{r}_1, \boldsymbol{r}_2) = \mathsf{V}_1(\boldsymbol{r}_1) + \mathsf{V}_1(\boldsymbol{r}_2) \tag{156}$$

then [cf. Eq. (92)]

$$\mathsf{H}_1 = \begin{bmatrix} (\mathsf{V}_1 + mc^2)\,\mathbf{1} & c\,(\boldsymbol{\sigma}\cdot\mathbf{p}) \\ c\,(\boldsymbol{\sigma}\cdot\mathbf{p}) & (\mathsf{V}_1 - mc^2)\,\mathbf{1} \end{bmatrix}. \tag{157}$$

The two-electron Dirac Hamiltonian is a 16×16 matrix operator which may be expressed as

$$\mathsf{H}_2(\boldsymbol{r}_1, \boldsymbol{r}_2) = \begin{bmatrix} \mathsf{V}_2\,\mathbf{1} & c\,\mathbf{1}\otimes(\boldsymbol{\sigma}\cdot\mathbf{p}) & c\,(\boldsymbol{\sigma}\cdot\mathbf{p})\otimes\mathbf{1} & \mathbf{0} \\ c\,\mathbf{1}\otimes(\boldsymbol{\sigma}\cdot\mathbf{p}) & (\mathsf{V}_2 - 2mc^2)\,\mathbf{1} & \mathbf{0} & c\,(\boldsymbol{\sigma}\cdot\mathbf{p})\otimes\mathbf{1} \\ c\,(\boldsymbol{\sigma}\cdot\mathbf{p})\otimes\mathbf{1} & \mathbf{0} & (\mathsf{V}_2 - 2mc^2)\,\mathbf{1} & c\,\mathbf{1}\otimes(\boldsymbol{\sigma}\cdot\mathbf{p}) \\ \mathbf{0} & c\,(\boldsymbol{\sigma}\cdot\mathbf{p})\otimes\mathbf{1} & c\,\mathbf{1}\otimes(\boldsymbol{\sigma}\cdot\mathbf{p}) & (\mathsf{V}_2 - 4mc^2)\,\mathbf{1} \end{bmatrix} \tag{158}$$

where

$$(\boldsymbol{\sigma}\cdot\ \mathbf{p})\otimes\mathbf{1}=\begin{bmatrix}\mathsf{p}_3 & 0 & \mathsf{p}_1-i\mathsf{p}_2 & 0\\ 0 & \mathsf{p}_3 & 0 & \mathsf{p}_1-i\mathsf{p}_2\\ \mathsf{p}_1+i\mathsf{p}_2 & 0 & -\mathsf{p}_3 & 0\\ 0 & \mathsf{p}_1+i\mathsf{p}_2 & 0 & -\mathsf{p}_3\end{bmatrix} \tag{159}$$

$$\mathbf{1}\otimes(\boldsymbol{\sigma}\cdot\ \mathbf{p})=\begin{bmatrix}\mathsf{p}_3 & \mathsf{p}_1-i\mathsf{p}_3 & 0 & 0\\ \mathsf{p}_1+i\mathsf{p}_2 & -\mathsf{p}_3 & 0 & 0\\ 0 & 0 & \mathsf{p}_3 & \mathsf{p}_1-i\mathsf{p}_2\\ 0 & 0 & \mathsf{p}_1+i\mathsf{p}_2 & -\mathsf{p}_3\end{bmatrix}. \tag{160}$$

If the particles are identical, then, according to the Pauli principle, the Hilbert space $\mathscr{H}_n^{\mathrm{D}}$ in which the eigenvalue problem of the n-particle Dirac Hamiltonian is defined has to be reduced to its antisymmetric subspace ${}^{\mathsf{A}}\mathscr{H}_n^{\mathrm{D}}\equiv\mathscr{H}_1^{\mathrm{D}\wedge n}$. Thus, if $\boldsymbol{\Phi}_{\mathrm{a}}^{(1)}$ and $\boldsymbol{\Phi}_{\mathrm{b}}^{(1)}$ are eigenfunctions of H_1 to the eigenvalues E_{a} and E_{b}, respectively, then

$$\mathsf{H}_2\,\boldsymbol{\Phi}_{\mathrm{ab}}^{(2)}=E_{\mathrm{ab}}^{(2)}\,\boldsymbol{\Phi}_{\mathrm{ab}}^{(2)}, \tag{161}$$

where

$$\boldsymbol{\Phi}_{\mathrm{ab}}^{(2)}=\frac{1}{\sqrt{2}}\left[\boldsymbol{\Phi}_{\mathrm{a}}^{(1)}\otimes\boldsymbol{\Phi}_{\mathrm{b}}^{(1)}-\boldsymbol{\Phi}_{\mathrm{b}}^{(1)}\otimes\boldsymbol{\Phi}_{\mathrm{a}}^{(1)}\right]. \tag{162}$$

is an antisymmetric eigenfunction of H_2 to the eigenvalue $E_{\mathrm{ab}}^{(2)}=E_{\mathrm{a}}^{(1)}+E_{\mathrm{b}}^{(1)}$. The 16-component two-particle wave function expressed in terms of large and small components of the one-electron spinors reads

$$\boldsymbol{\Phi}_{\mathrm{ab}}^{(2)}=\frac{1}{\sqrt{2}}\begin{bmatrix}\Phi_{\mathrm{L\,a}}^{(1)}\otimes\Phi_{\mathrm{L\,b}}^{(1)}-\Phi_{\mathrm{L\,b}}^{(1)}\otimes\Phi_{\mathrm{L\,a}}^{(1)}\\ \Phi_{\mathrm{L\,a}}^{(1)}\otimes\Phi_{\mathrm{S\,b}}^{(1)}-\Phi_{\mathrm{L\,b}}^{(1)}\otimes\Phi_{\mathrm{S\,a}}^{(1)}\\ \Phi_{\mathrm{S\,a}}^{(1)}\otimes\Phi_{\mathrm{L\,b}}^{(1)}-\Phi_{\mathrm{S\,b}}^{(1)}\otimes\Phi_{\mathrm{L\,a}}^{(1)}\\ \Phi_{\mathrm{S\,a}}^{(1)}\otimes\Phi_{\mathrm{S\,b}}^{(1)}-\Phi_{\mathrm{S\,b}}^{(1)}\otimes\Phi_{\mathrm{S\,a}}^{(1)}\end{bmatrix}\equiv\begin{bmatrix}\Phi_{\mathrm{LL\,ab}}^{(2)}\\ \Phi_{\mathrm{LS\,ab}}^{(2)}\\ \Phi_{\mathrm{SL\,ab}}^{(2)}\\ \Phi_{\mathrm{SS\,ab}}^{(2)}\end{bmatrix}. \tag{163}$$

From here one can easily get the symmetry properties of the two-particle components:

$$\begin{aligned}(12)\Phi_{\mathrm{LL}}^{(2)}&=-\Phi_{\mathrm{LL}}^{(2)},\\ (12)\Phi_{\mathrm{LS}}^{(2)}&=-\Phi_{\mathrm{SL}}^{(2)},\\ (12)\Phi_{\mathrm{SL}}^{(2)}&=-\Phi_{\mathrm{LS}}^{(2)},\\ (12)\Phi_{\mathrm{SS}}^{(2)}&=-\Phi_{\mathrm{SS}}^{(2)},\end{aligned} \tag{164}$$

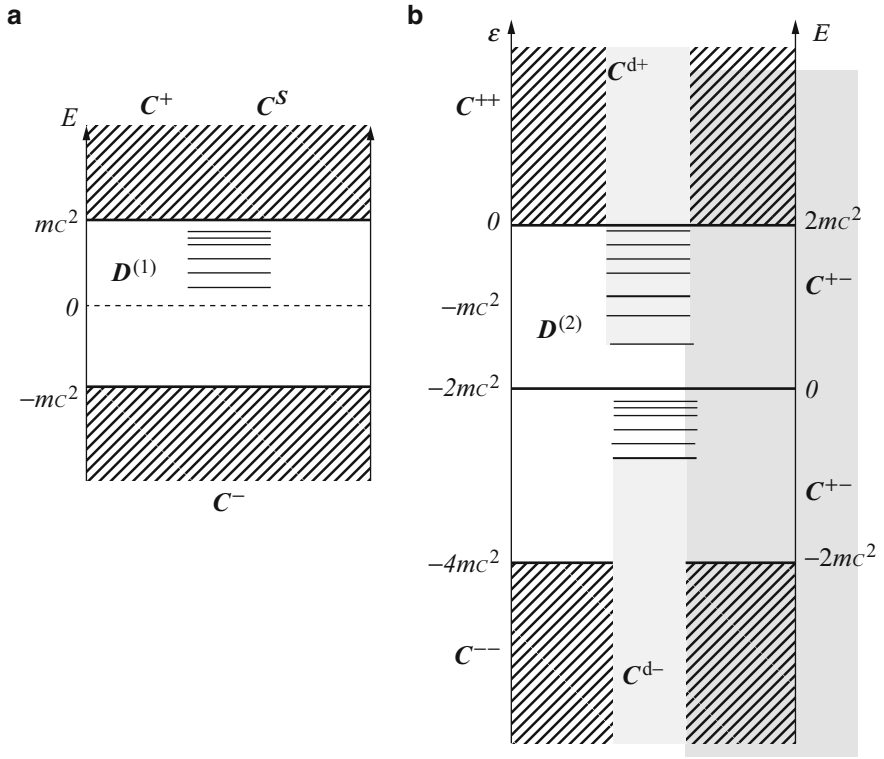


Fig. 2 Spectrum of the Dirac Hamiltonian with an external potential supporting bound states for one electron (**a**) and for two non-interacting electrons (**b**). In the one-electron case, there are two continua: $\boldsymbol{C}^{+}$ and $\boldsymbol{C}^{-}$ corresponding, respectively, to the positive- and to the negative-energy states. The discrete energies, $\boldsymbol{D}^{(1)}$, are located in the energy gap between the two continua. The continuum $\boldsymbol{C}^{S}$ of the non-relativistic Schrödinger model overlaps with $\boldsymbol{C}^{+}$. In the two-electron case, the states with both electrons occupying the same one-electron continuum generate either upper or lower continuum, $\boldsymbol{C}^{++}$ and $\boldsymbol{C}^{--}$, respectively. The Brown-Ravenhall continuum, $\boldsymbol{C}^{+-}$, spreads from $-\infty$ to $+\infty$ and comprises two-electron states with one electron in $\boldsymbol{C}^{+}$ and the other one in $\boldsymbol{C}^{-}$. If one electron occupies a discrete state and the other one $\boldsymbol{C}^{+}$ / $\boldsymbol{C}^{-}$, we get the continuum $\boldsymbol{C}^{\mathrm{d}+}$ / $\boldsymbol{C}^{\mathrm{d}-}$ associated with the one-electron ionization. The discrete two-electron energies, $\boldsymbol{D}^{(2)}$, correspond to both electrons in $\boldsymbol{D}^{(1)}$ and are located between $\boldsymbol{C}^{--}$ and $\boldsymbol{C}^{++}$

where (12) is the transposition operator of the particles and the indices ab have been omitted.

The structure of the two-electron spectrum resulting from different combinations of the one-electron states is shown in Fig. 2 and in Table 2. As one can see, apart from the positive and negative energy continua (similar to the one-electron case) and the continua associated with the ionization of one of the electrons, the Brown-Ravenhall continuum spreads over the entire energy range. All discrete states of the two-electron Hamiltonian are degenerate with the states of this continuum.

Table 2 Structure of spectrum of two-electron Dirac Hamiltonian

One-electron states		Two-electron states		
$E_a^{(1)}$	$E_b^{(1)}$	$E_{ab}^{(2)} = E_a^{(1)} + E_b^{(1)}$	Range[a]	Description
$\boldsymbol{D}^{(1)}$	$\boldsymbol{D}^{(1)}$	$\boldsymbol{D}^{(2)}$	$\left(-2mc^2, +2mc^2\right)$	Discrete
$\boldsymbol{C}^{+}$	$\boldsymbol{C}^{+}$	$\boldsymbol{C}^{++}$	$\left(2mc^2, +\infty\right)$	Positive continuum
$\boldsymbol{C}^{-}$	$\boldsymbol{C}^{-}$	$\boldsymbol{C}^{--}$	$\left(-\infty, -2mc^2\right)$	Negative continuum
$\boldsymbol{C}^{+}$	$\boldsymbol{C}^{-}$	$\boldsymbol{C}^{+-}$	$(-\infty, +\infty)$	Brown-Ravenhall continuum
$\boldsymbol{D}^{(1)}$	$\boldsymbol{C}^{+}$	$\boldsymbol{C}^{\mathrm{d}+}$	$\left(2mc^2 - E_d^{(1)}, +\infty\right)$	Positive-energy ionization
$\boldsymbol{C}^{-}$	$\boldsymbol{D}^{(1)}$	$\boldsymbol{C}^{\mathrm{d}-}$	$\left(-\infty, -E_d^{(1)}\right)$	Negative-energy ionization

[a] $E_d^{(1)}$ – discrete one-electron energy level

However, the discrete states are well defined and do not contain any contribution of the BR continuum as long as the electron interaction terms are neglected. Thus, the discrete solutions of the two-electron Dirac equation can be exactly determined. In particular, one can eliminate the influence of the BR continuum by the separation of the two-electron eigenvalue problem to two independent one-electron problems.

The structure of spectra composed of mutually overlapping discrete and continuous sections and also containing auto-ionizing states may be efficiently studied using the *complex coordinate rotation* (CCR), also known as the *complex scaling*, method [28]. The basic theorem of the method says that the transformation $\boldsymbol{r} \rightarrow \boldsymbol{r}\,\mathrm{e}^{i\theta}$ performed on the Hamiltonian does not change the bound-state energies while the continua move to the complex plane. After the transformation the Hamiltonian is non-Hermitian and its eigenvalues z are complex. The energies are equal to $E = Re(z)$, and the imaginary parts, in the case of auto-ionizing states, are related to their widths $\Gamma = -2\,Im(z)$. The spectra of one- and two-particle Dirac Hamiltonian, after the rotation by θ, are shown in Fig. 3. The discrete levels are separated from the continua, and each continuum occupies a specific area in the complex plane.

The ideas based on CCR, recently applied in numerical studies of spectral properties of two-electron Dirac operator [35], present a useful tool for a description of the structure of the *computed* spectrum. In a discrete representation, the eigenvalues of the Hamiltonian matrix corresponding to the discrete and to the continuum states are mixed and in many cases are difficult to distinguish. In the CCR calculations, using the same basis set as in the standard ones, one can see which eigenvalues correspond to the well-described discrete energies, which ones to the well-described continuum states, and which ones to poorly described mixtures. This kind of diagnostics may be done in both one- and two-electron calculations. Also the pedagogical features of CCR are worth noticing. The structure of the two-electron DC spectrum is complicated and difficult to describe using the traditional formalism. With CCR, one

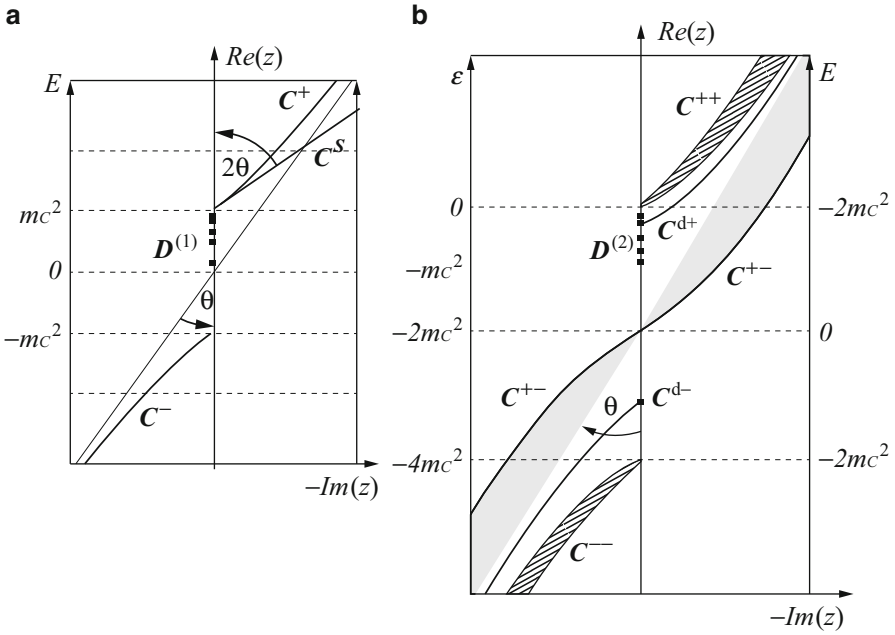


Fig. 3 The same as in Fig. 2 but after complex coordinate rotation by the angle θ. The continua are represented by *lines* and *strips* extending towards the imaginary axis. The *dots* in the real axis represent the bound-state energies. Now the Schrödinger and the positive Dirac continua, $\boldsymbol{C}^S$ and $\boldsymbol{C}^+$, respectively, are separated

can easily distinguish different kinds of continua and explain what is the meaning of the statement that *the non-projected DC Hamiltonian of interacting particles does not have bound states*. The essential difference between two Coulomb-confined non-interacting and interacting electrons becomes much easier to discuss with the picture of the CCR spectrum in hand.

In the non-relativistic limit, Eq. (161) transforms to the two-electron LL equation:

$$\begin{aligned}(\boldsymbol{\sigma}\cdot\boldsymbol{p})_2\,\Phi_{\mathrm{LS}} + (\boldsymbol{\sigma}\cdot\boldsymbol{p})_1\,\Phi_{\mathrm{SL}} &= \mathscr{E}\,\Phi^{\mathrm{LL}},\\ (\boldsymbol{\sigma}\cdot\boldsymbol{p})_2\,\Phi^{\mathrm{LL}} &= 2m\,\Phi^{\mathrm{LS}},\\ (\boldsymbol{\sigma}\cdot\boldsymbol{p})_1\,\Phi^{\mathrm{LL}} &= 2m\,\Phi^{\mathrm{SL}}.\end{aligned} \tag{165}$$

The elimination of Φ^{LS} and Φ^{SL} gives a two-electron Schrödinger equation.

Two Interacting Particles

The interaction operator couples the one-electron states with the states of the Brown-Ravenhall continuum. As a consequence, the DC Hamiltonian, as defined in Eq. (153), has no bound-state solutions. It can be used for a description of physical

properties of many-electron systems if it is projected to the positive-energy subspace of the Hilbert space. In the case of two electrons, this implies using a projected Hamiltonian

$$\mathsf{H}_2^{\mathrm{DC}+}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \Pi_+(1)\Pi_+(2)\mathsf{H}_2^{\mathrm{DC}}(\boldsymbol{r}_1, \boldsymbol{r}_2)\Pi_+(2)\Pi_+(1), \tag{166}$$

where the projection operators, defined in Eq. (84), have to be related to a specific one-electron Hamiltonian – they project on the space of the positive-energy states of this very Hamiltonian. The resulting model is known as the *no-pair approximation* since it cuts off a possibility of the creation of virtual electron-positron pair. It suffers from many drawbacks [12, 13, 15], but it is a basis for computationally stable and most commonly used approaches [12, 13]. A discussion of beyond-no-pair approaches may be found in a recent work by Liu and Lindgren [17].

A simple and transparent way of solving relativistic many-electron problems is based on using a representation of the Dirac-Coulomb Hamiltonian in the Fock space. A discussion of this subject may be found, e.g., in [12, 13].

Summary

Contrary to the logically closed non-relativistic quantum mechanics, the relativistic one suffers from many inconsistencies and artifacts. In particular, exact many-particle generalizations are not possible, and the one-particle formulation requires introduction of infinitely many particles which are not described by the mathematical formalism. All contradictions and conceptual difficulties have been removed by quantum electrodynamics, a complete theory of electromagnetic interactions. But, in chemical applications, relativistic quantum mechanics cannot be replaced by quantum electrodynamics due to complexity of the latter. However, over the last several decades, the sources of artifacts generated by the relativistic quantum mechanics have been identified and efficient algorithms for solving many-electron relativistic equations have been developed. As it was already mentioned, the construction of a bridge between relativistic quantum mechanics and QED based on solid mathematical foundations started several years ago and has been presented in recent works [15–18] and in section *Relativistic Hamiltonians*.

Apart from the works aimed at developing approaches which are directly linked to the QED limit of relativistic quantum mechanics, studies on mathematical properties of different forms of the Dirac-Coulomb Hamiltonian attract attention of some researchers. In particular, an interesting and important monograph by Moiseyev [28], concerned with non-Hermitian quantum mechanics, shows possibilities of some nonstandard ways of approaching problems which in traditional terms would be classified as “ill defined” [35].

Acknowledgements The author is most grateful to W. H. Eugen Schwarz, Wenjian Liu, and Christoph van Wüllen for their comments. A discussion with Trond Saue is gratefully acknowledged.

References

1. Davydov AS (1976) Quantum mechanics, 2nd edn. Pergamon, Oxford
2. Messiah A (1967) Quantum mechanics. North Holland, Amsterdam
3. Itzykson C, Zuber JB (1980) Quantum field theory. McGraw-Hill, New York
4. Sakurai JJ (1984) Advanced quantum mechanics. Benjamin/Cummings, Menlo Park
5. Greiner W (2000) Relativistic quantum mechanics, 3rd edn. Spinger, Berlin/Heidelberg/New York
6. Landau RH (1996) Quantum mechanics II. A second course in quantum theory. John Wiley, New York
7. Pilkuhn HM (2005) Relativistic quantum mechanics, 2nd edn. Springer, Berlin/Heidelberg/New York
8. Das A (2008) Lectures on quantum field theory. World Scientific, Singapore
9. Schwabl F (2004) Advanced quantum mechanics. Springer, Berlin/Heidelberg/New York
10. Scadron MD (1991) Advanced quantum theory. Springer, Berlin/Heidelberg/New York
11. Thaller B (1992) The Dirac equation. Springer, Berlin/Heidelberg/New York
12. Reiher M, Wolf A (2009) Relativistic quantum chemistry: the fundamental theory of molecular science. Wiley-VCH, Weinheim
13. Dyall KG, Fægri K (2007) Introduction to relativistic quantum chemistry. Oxford University Press, Oxford
14. Grant IP (2007) Relativistic quantum theory of atoms and molecules: theory and computation. Springer, Berlin/Heidelberg/New York
15. Lindgren I (2013) Relativistic many-body theory. A new field-theoretical approach. Springer, Berlin/Heidelberg/New York
16. Lindgren I (2014) Development of many-body perturbation theory: how to combine with quantum electrodynamics. Int J Quantum Chem 114:1176–1182
17. Liu W, Lindgren I (2013) Going beyond 'no-pair quantum chemistry'. J Chem Phys 139:014108
18. Liu W (2014) Advances in relativistic quantum mechanics. Phys Rep 537:59–89
19. Kramers HA (1933) Die Grundlagen der Quantentheorie: Quantentheorie des Electrons und der Strahlung. Akad. Verlagsges, Leipzig
20. van der Waerden BL (1932) Die gruppentheoretische Methode in der Quantenmechanik. J Springer, Berlin
21. Boi L (2011) The quantum vacuum: a scientific and philosophical concept, from electrodynamics to string theory and geometry of the microscopic world. The Johns Hopkins University Press, Baltimore
22. Soff G, Müller B, Rafelski J, Greiner W (1973) Solution of the Dirac equation for scalar potentials and its implications in atomic physics. Z Naturforsch A 28:1389–1396
23. Kutzelnigg W (1984) Basis set expansion of the Dirac operator without variational collapse. Int J Quantum Chem 25:107–129
24. Kutzelnigg W (1997) Relativistic Hamiltonians 'for electrons only' and the variational treatment of the Dirac equation. Chem Phys 225:203–222
25. Esteban MJ, Lewin M, Séré E (2008) Variational methods in relativistic quantum mechanics. Bull Am Math Soc 45:535–593
26. Talman JD (1986) Minimax principle for the Dirac equation. Phys Rev Lett 57:1091–1094
27. Dolbeault J, Esteban MJ, Séré E (2000) On the eigenvalues of operators with gaps. Application to Dirac operators. J Funct Anal 174:208–226
28. Moiseyev N (2011) Non-Hermitian quantum mechanics. Cambridge University Press, New York
29. Moore RA (1975) An alternative method of obtaining approximate solutions to the Dirac equation. Can J Phys 53:1240–1246
30. Franco-Villafa ne JA, Sadurni E, Barkhofen S, Kuhl U, Mortessagne F, Seligman TH (2013) First experimental realization of the Dirac oscillator. Phys Rev Lett 111:170405

31. Cook PA (1971) Relativistic harmonic oscillators with intrinsic spin structure. Lett Nuovo Cimento 1:419–426
32. Moshinsky M, Szczepaniak A (1989) The Dirac oscillator. J Phys A Math Gen 22:L817–L820
33. Karwowski J, Pestka G (2007) Harmonic oscillators in relativistic quantum mechanics. Theor Chem Acc 118:519–525
34. Brown GE, Ravenhall DG (1951) On the interaction of two electrons. Proc R Soc A 208:552–559
35. Pestka G, Bylicki M, Karwowski J (2012) Geminals in Dirac-coulomb eigenvalue problem. J Math Chem 50:510–533

Nuclear Charge Density and Magnetization Distributions

2

Dirk Andrae

Contents

Introduction 52
Nucleons and Nuclei 52
Nuclear Charge Density and Magnetization Distributions 58
  Spherically Symmetric Charge Density Distributions 59
  Dipolar Magnetization Distributions 66
Finite Nucleus Models in Electronic Structure Calculations 70
Summary 77
Appendix 77
References 79

Abstract

In the study of electronic structure of matter, the atomic nuclei play the role of centers of force to which the electrons are bound. Within this context, almost all of the internal details of nuclear structure can be neglected, and the nuclei can be considered as objects with static extended distributions of charge and magnetic moment. This chapter presents a discussion of nuclear charge density and magnetization distributions. The underlying general principles are discussed, and details are given for model distributions that are widely used in relativistic quantum chemistry. Finally, the principal effects of extended nuclear distributions of charge and magnetic moment are pointed out.

Keywords

Nuclear charge density distribution • nuclear magnetization distribution • electrostatic potential energy function • magnetostatic vector potential

D. Andrae (✉)
Institut für Chemie und Biochemie, Freie Universität Berlin, Berlin, Germany
e-mail: dirk.andrae@fu-berlin.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_23

Introduction

In a large part of chemistry and physics, atomic nuclei are considered mainly as massive point-like centers of positive charge, which attract the surrounding electrons and repel other nuclei. The attractive and repulsive Coulomb interactions between the constituents of matter, as well as their kinetic energy, must be balanced if matter shall form stable structures. The application of quantum mechanics to the problem of describing the behavior of electrons and nuclei under these interactions finally leads – in combination with structure optimization according to the energy minimization principle – to the understanding of stable molecular and crystalline structures as representations of local minima on high-dimensional total energy hypersurfaces.

The simplistic description of atomic nuclei just mentioned above requires refinement in various aspects. On the one hand, the nuclei are known to be extended (or finite) objects, composed of smaller constituents (nucleons or quarks) bound together by the strong force (mediated through exchange of mesons or gluons). This strong force outweighs the repulsive electromagnetic interaction between the positively charged nuclear constituents in stable nuclei. Consequently any spatial nuclear property should be represented by an extended distribution, instead of a Dirac delta distribution (finite nucleus case, FNC, versus point-like nucleus case, PNC). On the other hand, the atomic nuclei, which frequently also have a magnetic dipole moment, can serve as most natural and most sensitive probes of the electronic structure of matter. Several experimental techniques, e.g., nuclear magnetic or quadrupolar resonance spectroscopy (NMR, NQR) and Mössbauer spectroscopy, take advantage of this fact.

It follows already from these quite general considerations that a closer examination of models for extended nuclei is of importance in all quantum mechanical studies aiming at an accurate representation of the electronic structure in atoms, molecules, and solids.

The present chapter deals with the representation of extended nuclei in theoretical approaches to the problem of electronic structure. The next section provides a rather general discussion of the properties of atomic nuclei, including a look at the current status of the chart of nuclides. Then follows a section that treats nuclear charge density and magnetization distributions in greater detail, with particular emphasis on spherically symmetric nuclear charge density distributions and purely dipolar nuclear magnetization distributions. Finally, the importance of extended nucleus models in several areas of relativistic quantum chemistry and atomic and molecular physics is demonstrated.

Nucleons and Nuclei

For the purpose of studies of the electronic structure of atoms, molecules or periodic systems (polymers, layers, surfaces, crystals), the atomic nuclei may be regarded as compact objects carrying positive electric charge and possibly a magnetic moment.

Table 1 Experimental values of some properties of the nucleons and other subatomic particles [1][a]

Particle	Charge ratio q/e	Mass ratio m/m_e	g Factor[b]	Magnetic moment μ (in 10^{-26} J T^{-1})
Proton (p)	+1	1836.15267389(17)	5.585694702(17)	1.4106067873(97)
Neutron (n)	0	1838.68366158(90)	−3.82608545(90)	−0.96623650(23)
Electron (e)	−1	1	−2.00231930436182(52)	−928.4764620(57)
Muon (μ)	−1	206.7682826(46)	−2.0023318418(13)	−4.49044826(10)

[a]The numerical value of the standard uncertainty, given in parentheses, refers to the corresponding last digits of the quoted result
[b]See [2] for the adopted convention on sign of g factors

The nuclei can be considered as being composed of protons (p), carrying an electric charge $q_\mathrm{p} = +e$ (e denotes the elementary charge), and neutrons (n) with $q_\mathrm{n} = 0$. These particles are hadrons, i.e., they are subject to the short-ranged attractive strong nuclear force (which overcomes the repulsive Coulomb interaction between the protons). They are the lightest members of this kind of particles, and they have nearly equal mass (see Table 1). Hence, the term "nucleon" has been introduced and is being used to cover both protons and neutrons.

The nucleons are fermions with spin (intrinsic angular momentum) $\boldsymbol{s}$ of squared magnitude $\boldsymbol{s}^2 = s(s+1)\hbar^2$, where $s = 1/2$ is the associated spin quantum number and $\hbar = h/(2\pi)$ is the reduced Planck constant. Both types of nucleons have an intrinsic magnetic dipole moment $\boldsymbol{\mu}_i = \gamma_i \boldsymbol{s} = g_i \mu_\mathrm{N} \boldsymbol{s}/\hbar$ (i = n, p), where $\mu_\mathrm{N} = e\hbar/(2m_\mathrm{p})$ denotes the nuclear magneton (m_p is the mass of the proton). These relations define two scalar quantities, the nuclear magnetogyric ratio γ_i and the nuclear g factor g_i. Their sign shows whether the two vectors $\boldsymbol{s}$ and $\boldsymbol{\mu}_i$ are oriented in parallel or antiparallel. The scalar quantity $\mu_i = s g_i \mu_\mathrm{N}$ is frequently simply called the "magnetic moment." Recommended values [1] for these quantities are shown in Table 1, which includes also corresponding data for the electron (e) and the muon (μ) for comparison. The nucleons have no other higher electric or magnetic multipole moments beyond the electric monopole moment (charge) and the magnetic dipole moment.

It is known that the nucleons are extended rather than point-like objects (in contrast to the electron, which has neither substructure nor smaller constituting components according to present-day knowledge). As for any other quantum mechanical system, the answer to the question of the size of the nucleons, expressed in terms of some radius, depends on the property or quantity being studied (size of the distribution of mass, of charge, or of magnetic moment) and on the probe being used. There are currently available two significantly different values for the proton rms charge radius [1,3]: $r_\mathrm{p}^\mathrm{ch} = 0.8751(61)$ fm (from H and D transition frequencies and electron-proton scattering) and $r_\mathrm{p}^\mathrm{ch} = 0.84087(39)$ fm (from Lamb shift in muonic hydrogen). The disagreement is not yet understood and should in fact be absent if both electron-proton and muon-proton interaction were purely electromagnetic and of identical strength. The proton magnetic radius is $r_\mathrm{p}^\mathrm{mag} = 0.777(16)$ fm [3].

Also available are values for the neutron mean square charge radius, $\langle (r_n^{ch})^2 \rangle = -0.1161(22)\,\mathrm{fm}^2$, and for the neutron magnetic radius, $r_n^{mag} = 0.862(9)\,\mathrm{fm}$ [3].

Neither the proton nor the neutron is an "elementary particle"; they are both composed of smaller constituents called quarks: p = (uud) and n = (udd), where u denotes the up quark ($q_u = +2/3\ e$) and d denotes the down-quark and e denoting elementary charge ($q_d = -1/3\ e$). These quarks, spin-1/2 fermions themselves, are susceptible to elektroweak and strong interactions. They are bound together through exchange of gluons, the massless spin-1 gauge bosons of the strong nuclear force. The detailed study of this strong interaction is the domain of quantum chromodynamics (QCD) and is beyond the scope of this chapter.

The two integers proton number (nuclear charge number), Z, and neutron number, N, completely specify a nuclide, i.e., a type of atomic nucleus. Their sum $A = Z + N$, the nucleon (or mass) number of the nuclide, is used together with the element symbol X to denote a particular isotope of that element (^{A}X, e.g., ^{2}H, ^{7}Li, ^{12}C, ^{40}K, ^{200}Hg, ^{238}U). An atomic ion with n electrons can then be denoted as $^A\mathrm{X}^{(Z-n)+}$ ($0 \leq n < Z$), but this notation does not allow to distinguish between the bare nucleus ($n = 0$, ^{A}X denotes the nuclide) and the neutral atom ($n = Z$, ^{A}X denotes the atom). Recent editions of the chart of nuclides [4,5] provide data for more than 3000 experimentally observed nuclides (see also Fig. 1). The nuclides can

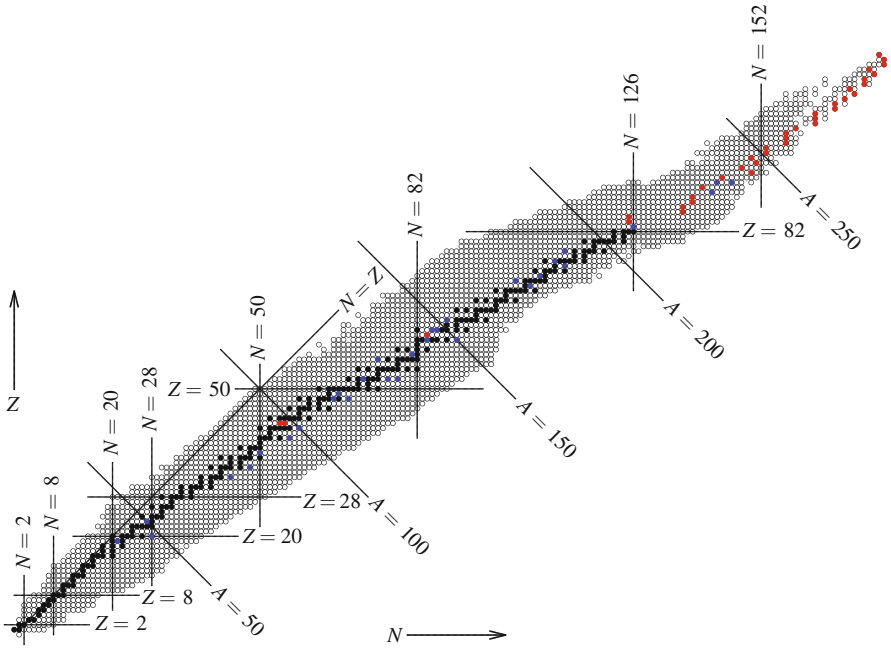


Fig. 1 Chart of nuclides showing stable nuclides (*black dots*) and primordial radionuclides (*blue dots*) tracing out the bottom of the "valley of stability," as well as artificial radionuclides with longest confirmed half-life (for $Z = 43$, $Z = 61$ and $Z > 83$, *red dots*). "Magic" proton/neutron numbers are indicated by *horizontal/vertical lines* (All data taken from [4])

Table 2 Some ground-state properties of selected nuclides ^{A}X (nuclear spin quantum number I and parity label π, magnetic moment μ, electric quadrupole moment Q, rms charge radius a)

Nuclide	I^{π} [a]	μ [b]/μ_N	Q [c]/mb	a [d]/fm
^{1}n	$1/2^{+}$	−1.91304273(45)	–	–
^{1}H	$1/2^{+}$	+2.7928473508(85)	–	0.8751
^{2}H	1^{+}	+0.8574382311(48)	2.860(15)	2.1421
^{7}Li	$3/2^{-}$	+3.256427(2)	−40.1	2.4440
^{12}C	0^{+}	–	–	2.4702
^{13}C	$1/2^{-}$	+0.7024118(14)	–	2.4612
^{40}K	4^{-}	−1.298100(3)	−73	3.4381
^{200}Hg	0^{+}	–	–	5.4551
^{201}Hg	$3/2^{-}$	−0.5602257(14)	387(6)	5.4581
^{235}U	$7/2^{-}$	−0.38(3)	4936(6)	5.8337
^{238}U	0^{+}	–	–	5.8571

[a] From [5]
[b] From [1, 9]
[c] From [10], in millibarn (1 mb $=$ 10^{-31} m^2)
[d] From [1, 20]

be grouped further into four classes, depending on whether the proton and neutron numbers are even or odd.

The full characterization of a stationary state of an atomic nucleus requires, of course, further quantities, in addition to proton and neutron number. These include the nuclear spin quantum number I and the parity π of the state, usually denoted together in shorthand as I^{π} (see Table 2 for some examples). Nuclei with even mass number A have integer spin quantum numbers (they are bosons), whereas those with odd mass number have half-integer spin quantum numbers (they are fermions). The fact that all even-even nuclei in their ground states have $I = 0$ supports the suggestion that the nucleons tend to couple their individual spins pairwise, similar to the way of spin coupling known from the electrons. In consequence, the uncoupled (or valence) nucleon(s) are primarily responsible for a resulting nonzero total nuclear spin.

In general, the nuclear spin quantum number I characterizes the state-specific nuclear angular momentum (nuclear spin) $\boldsymbol{I}$ of squared magnitude $\boldsymbol{I}^2 = I(I+1)\hbar^2$. The existence or nonexistence of nuclear electric and magnetic multipole moments is intimately related to the nuclear spin (and to the assumption that parity is a good quantum number for stationary nuclear states) [6]. There exist then nuclear multipole moments of all orders 2^l with $0 < l \leq 2I$. More specifically, electric multipole moments require l to be even, and magnetic multipole moments require l to be odd. Hence, electric multipole moments of lowest order are the electric quadrupole moment ($l = 2$) and the electric hexadecapole moment ($l = 4$), whereas magnetic multipole moments of lowest order are the magnetic dipole moment ($l = 1$) and the magnetic octupole moment ($l = 3$).

With a nuclear state function at hand, denoted as $|\nu I M_I\rangle$ ($-I \leq M_I < I$, and ν includes any additional quantum numbers required), all properties of the

corresponding nuclear state could be evaluated, in principle. Nuclear properties of interest for (nonrelativistic or relativistic) electronic structure calculations include, in particular:

(1) the particle number density distributions for protons, $n_{\mathrm{p}}(\boldsymbol{r})$, for neutrons, $n_{\mathrm{n}}(\boldsymbol{r})$, and for nucleons, $n_{\mathrm{nuc}}(\boldsymbol{r})$, which are connected through

$$A\, n_{\mathrm{nuc}}(\boldsymbol{r}) = Z\, n_{\mathrm{p}}(\boldsymbol{r}) + N\, n_{\mathrm{n}}(\boldsymbol{r}), \quad \int \mathrm{d}^3 r\, n_i(\boldsymbol{r}) = 1. \tag{1}$$

(2) the nuclear charge density distribution $\rho(\boldsymbol{r})$, normalized to the total nuclear charge

$$\int \mathrm{d}^3 r \rho(\boldsymbol{r}) = Ze. \tag{2}$$

(3) if $I > 0$, the magnetization (or magnetic moment density) distribution $\boldsymbol{m}(\boldsymbol{r})$, normalized to the nuclear magnetic dipole moment $\boldsymbol{\mu}$

$$\int \mathrm{d}^3 r\, \boldsymbol{m}(\boldsymbol{r}) = \boldsymbol{\mu} = \gamma \boldsymbol{I}, \quad \gamma = g\mu_{\mathrm{N}}/\hbar. \tag{3}$$

Here γ and g denote, respectively, the magnetogyric ratio and nuclear g factor of the nucleus under study.

(4) if $I > 1/2$, the nuclear electric quadrupole moment Q

$$\int \mathrm{d}^3 r (3z^2 - r^2)\rho(\boldsymbol{r}) = \sqrt{\frac{16\pi}{5}} \int \mathrm{d}^3 r r^2 Y_{20}(\hat{\boldsymbol{r}})\rho(\boldsymbol{r}) = Qe \tag{4}$$

(more precisely, the nuclear charge density distribution in this integral is the one obtained from the nuclear state function $|\nu I I\rangle$ [7]; $Y_{20}(\hat{\boldsymbol{r}})$ is a standard spherical harmonic; the position unit vector $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$ is equivalent to the two angular coordinates in the spherical coordinates system). A nonzero value of Q indicates a deformed nuclear charge density distribution that is not spherically symmetric. An axially symmetric nucleus with the shape of a prolate ellipsoid has a positive quadrupole moment ($Q > 0$), whereas a negative quadrupole moment ($Q < 0$) occurs when the shape is oblate ellipsoidal.

The nuclear magnetic dipole and electric quadrupole moments give rise to the so-called hyperfine interaction, i.e., the interaction of the electrons with nuclear multipolar moments. Therefore, these multipole moments can be extracted from, e.g., hyperfine splitting (hfs) data derived from optical spectroscopy but also from data obtained from electron paramagnetic resonance (EPR), nuclear magnetic resonance (NMR), and nuclear quadrupolar resonance (NQR) spectroscopy. At present, the combination of accurately measured nuclear quadrupole coupling constants (NQCC) e^2qQ/h with electric field gradients (EFG) eq obtained from sophisticated quantum chemical calculations provides the best way of access to

accurate nuclear electric quadrupole moments Q (see [8] and references therein for further details of this approach). Extensive tables are available that provide critically reviewed data of nuclear magnetic dipole moments [9] and electric quadrupole moments [9–11] (see also [12,13]). A small subset of such data is exemplarily shown in Table 2.

As shown above, the nuclear electric monopole moment (charge) and the nuclear multipole moments are integrated quantities. Hence, they do not provide any detailed information on the nuclear charge density distribution $\rho(\boldsymbol{r})$ and the nuclear magnetization distribution $\boldsymbol{m}(\boldsymbol{r})$ but have instead the role of normalization constants. Details of the nuclear charge density and magnetization distributions are accessible through elastic electron-nucleus or myon-nucleus scattering, and the information gained from these experiments has been parametrized in terms of model distributions [14–16] (see the next section for further details).

The most important nuclear parameters, extracted from these studies, are the size parameters (rms radii a, see also [17–20]) and measures for the diffuseness of the charge density distributions in their peripheral region (skin thickness t, i.e., the radial distance over which the distribution decreases from 90 % to 10 % of its central value). Values for the rms charge radius a are included in Table 2. In comparison to the typical dimensions of atomic and molecular structure ($1\,\text{Å} = 10^{-10}\,\text{m}$), the atomic nuclei are about five orders of magnitude smaller ($1\,\text{fm} = 10^{-15}\,\text{m}$). In the case of absence of experimental data, a good approximate value for the rms charge radius a of nuclides ${}^{A}\text{X}$ close to the bottom of the "valley of stability" (see Fig. 1) can be obtained from the mass number [21]

$$a({}^{A}\text{X})/\text{fm} = 0.836A^{1/3} + 0.570, \quad A > 9. \tag{5}$$

This relation may be augmented with a relation between mass number and proton number, like [22]

$$A(Z) = 4.467 \cdot 10^{-3}Z^2 + 2.163Z - 1.168, \quad 1 \le Z \le 100, \quad A(1) = 1, \tag{6}$$

to obtain rms charge radii changing monotonously and rather smoothly in a sweep through the periodic table, allowing extrapolation even to superheavy elements ($Z > 100$). A more general expression for rms charge radii, dependent on mass number A and neutron excess $N - Z$, is [23]

$$a(Z,N) = r_0^{\text{ch}} A^{1/3}\left(1 + c_1^{\text{ch}}\frac{1}{A} + c_2^{\text{ch}}\frac{N-Z}{A}\right), \quad A = Z + N, \tag{7}$$

with parameters $r_0^{\text{ch}} = 1.240\,\text{fm}$, $c_1^{\text{ch}} = 1.646$, $c_2^{\text{ch}} = -0.191$. Equation (7) describes particularly well the rms charge radii of even-even nuclides and may be used for other nuclides too. The other parameter, the skin thickness t of the nuclear charge density distribution, is practically constant for nuclei with sufficiently large mass [17]:

$$t = 2.3\,\text{fm}, \quad A > 45. \tag{8}$$

Simple, but already quite realistic models for the charge density distribution that depend on just one or two parameters can be (and always should be) standardized to data for rms charge radius a and skin thickness t, whenever such data are available. Parameter values for three widely used extended nuclear charge density distributions (the homogeneous, the Gauss-type, and the Fermi-type distributions; see next section for details), which are in accordance with such a standardization, are available [24].

Nuclear Charge Density and Magnetization Distributions

From static distributions of the nuclear charge ($\rho(\boldsymbol{r})$) and of the nuclear magnetic dipole moment ($\boldsymbol{m}(\boldsymbol{r})$), the electrostatic scalar potential $\Phi(\boldsymbol{r})$ and the vector potential $\boldsymbol{A}(\boldsymbol{r})$ are obtainable. Both are required for (nonrelativistic or relativistic) electronic structure calculations. The details are as follows [25]:

(1) the nuclear charge density distribution $\rho(\boldsymbol{r})$ leads to the scalar potential

$$\Phi(\boldsymbol{r}) = \frac{1}{4\pi\varepsilon_0}\int \mathrm{d}^3 r' \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}, \tag{9}$$

which contributes via the nuclear potential energy function $V_{\mathrm{nuc}}(\boldsymbol{r}) = q_e\Phi(\boldsymbol{r}) = -e\Phi(\boldsymbol{r})$ to the potential energy of the electronic system. The electric field generated by the nucleus is $\boldsymbol{E}(\boldsymbol{r}) = -\nabla\Phi(\boldsymbol{r}) = -\mathrm{grad}\Phi(\boldsymbol{r})$.

(2) the nuclear magnetization distribution $\boldsymbol{m}(\boldsymbol{r})$ gives rise (within Coulomb gauge, $\nabla\cdot\boldsymbol{A} = 0$) to a vector potential

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu_0}{4\pi}\int \mathrm{d}^3 r' \frac{\nabla'\times\boldsymbol{m}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{10}$$

The magnetic field generated by the nucleus is $\boldsymbol{B}(\boldsymbol{r}) = \nabla\times\boldsymbol{A}(\boldsymbol{r}) = \mathrm{curl}\,\boldsymbol{A}(\boldsymbol{r})$.

These relations assume that the charge density and magnetization distributions, as well as the potentials and the fields, vanish at infinity. The scalar potential $\Phi(\boldsymbol{r})$ and the vector potential $\boldsymbol{A}(\boldsymbol{r})$ then enter into the Hamiltonian of the electronic structure calculation at the usual places.

Spherically symmetric charge density distributions and purely dipolar magnetization distributions are of largest importance for the practical work of calculating the electronic structure and other properties of matter. The following subsections provide detailed information on such distributions. The system of units is changed to atomic units ($\hbar = e = m_{\mathrm{e}} = 4\pi\varepsilon_0 = 1$, $\mu_0/(4\pi) = \alpha^2$, $\alpha \approx 1/137$ denotes the Sommerfeld fine structure constant).

Spherically Symmetric Charge Density Distributions

The fundamental working equations for spherically symmetric nuclear charge density distributions are given below. The presentation follows closely [22, 26].

Given is a spherically symmetric nuclear charge density distribution, $\rho(\boldsymbol{r}) = \rho(r)$, that is properly normalized,

$$\int \mathrm{d}^3 r \rho(\boldsymbol{r}) = 4\pi \int_0^\infty \mathrm{d}r\, r^2 \rho(r) = Z. \tag{11}$$

Radial expectation values of this normalized charge density for an arbitrary function of the radius, $f(r)$, are then obtained from

$$\langle f(r) \rangle = \frac{4\pi}{Z} \int_0^\infty \mathrm{d}r\, r^2 f(r) \rho(r). \tag{12}$$

Of particular importance are the expectation values of integral powers of the radius

$$\langle r^k \rangle = \frac{4\pi}{Z} \int_0^\infty \mathrm{d}r\, r^{k+2} \rho(r), \quad k_{\min} \le k \le k_{\max}, \quad \langle r^0 \rangle = 1. \tag{13}$$

For $k < k_{\min}$, the integrand becomes singular at the lower boundary, whereas a finite upper limit $k_{\max}$ ($k_{\max} < \infty$) applies only for charge density distributions that do not decrease rapidly enough as r approaches infinity. Also of interest is

$$\langle \ln(r) \rangle = \frac{4\pi}{Z} \int_0^\infty \mathrm{d}r\, r^2 \ln(r) \rho(r). \tag{14}$$

These expectation values are related to the moment function [27],

$$M(p) = \langle r^p \rangle^{1/p} \; (p \neq 0), \qquad M(0) = \exp\left(\langle \ln(r) \rangle\right). \tag{15}$$

Hence, the rms charge radius of $\rho(r)$ is $a = \langle r^2 \rangle^{1/2} = M(2)$. The nuclear charge form factor, a quantity of importance, e.g., in elastic electron-nucleus scattering, is

$$F(q) = \frac{4\pi}{Z} \int_0^\infty \mathrm{d}r\, r^2 j_0(qr) \rho(r) = \langle j_0(qr) \rangle \tag{16}$$

(see the appendix to this chapter for the definition of the spherical Bessel functions $j_n(x)$). The form factor depends on the magnitude of the transferred linear momentum, $q = |\boldsymbol{p}_\mathrm{f} - \boldsymbol{p}_\mathrm{i}|$ (initial momentum $\boldsymbol{p}_\mathrm{i}$, final momentum $\boldsymbol{p}_\mathrm{f}$, $|\boldsymbol{p}_\mathrm{i}| = |\boldsymbol{p}_\mathrm{f}|$), and is an even function ($F(-q) = F(q)$). Its power series expansion for small q,

$$F(q) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \langle r^{2k} \rangle q^{2k} = 1 - \frac{1}{6} \langle r^2 \rangle q^2 + O(q^4), \tag{17}$$

relates the form factor to the radial expectation values $\langle r^{2k} \rangle$ ($k \geq 0$) of the nuclear charge density distribution. Equation (16) can also be understood as a Fourier-Bessel (or Hankel) integral transform (of order zero). There exists therefore an inverse integral transform relation,

$$\rho(r) = \frac{Z}{2\pi^2} \int_0^\infty \mathrm{d}q \, q^2 \, j_0(qr) F(q), \tag{18}$$

that would allow to obtain the charge density distribution itself directly from the charge form factor if the latter were accurately known from the scattering experiments over the full infinite range of q. This, however, is not the case. The experimentally available data for $F(q)$ always require augmentation for both very small and very large values of the transferred momentum.

The spherically symmetric nuclear charge density distribution $\rho(r)$ leads to a spherically symmetric potential energy function $V_{\mathrm{nuc}}(r)$, given implicitly as

$$-rV_{\mathrm{nuc}}(r) = 4\pi \left(\int_0^r \mathrm{d}s \, s^2 \rho(s) + r \int_r^\infty \mathrm{d}s \, s \rho(s) \right). \tag{19}$$

The first and second derivatives of this potential energy function are

$$V'_{\mathrm{nuc}}(r) = \frac{\mathrm{d}}{\mathrm{d}r} V_{\mathrm{nuc}}(r) = \frac{4\pi}{r^2} \int_0^r \mathrm{d}s \, s^2 \rho(s), \tag{20}$$

$$V''_{\mathrm{nuc}}(r) = \frac{\mathrm{d}^2}{\mathrm{d}r^2} V_{\mathrm{nuc}}(r) = 4\pi \, \rho(r) - \frac{2}{r} V'_{\mathrm{nuc}}(r). \tag{21}$$

The potential energy function $V_{\mathrm{nuc}}(r)$ is, in general, continuous and differentiable. Charge density distributions $\rho(r)$, which have a discontinuity at some radius $r = R$, lead to a discontinuity only in the second and all higher derivatives of $V_{\mathrm{nuc}}(r)$ at that same radius. The power series expansions for small r of the charge density distribution,

$$\rho(r) = \sum_{m=-1}^{\infty} \rho_m r^m = \frac{\rho - 1}{r} + \rho_0 + \rho_1 r + O(r^2), \tag{22}$$

and of the potential energy function,

$$V_{\mathrm{nuc}}(r) = \sum_{k=0}^{\infty} v_k r^k = v_0 + v_1 r + O(r^2), \tag{23}$$

are related through

$$v_0 = -Z \langle r^{-1} \rangle, \qquad v_k = \frac{4\pi}{k(k+1)} \rho_{k-2}, \quad k \geq 1. \tag{24}$$

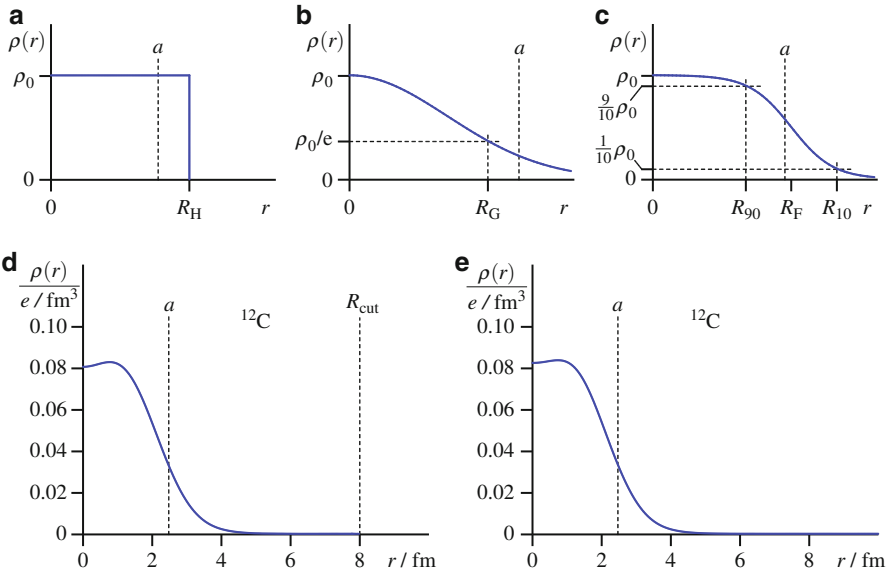


Fig. 2 Spherical charge density distributions $\rho(r)$ (with indication of their rms radius $a = \langle r^2 \rangle^{1/2}$). *Upper row*: (**a**) Homogeneous or uniform distribution, (**b**) Gauss-type distribution, and (**c**) Fermi-type distribution (with $b = 0.15$). *Lower row*: (**d**) Fourier-Bessel distribution for nuclide ^{12}C [18] and (**e**) sum-of-Gaussians distribution for nuclide ^{12}C [16] (See text for further details)

The coefficient $v_0 = V_{\rm nuc}(0)$ represents the depth of the potential well generated by the extended nuclear charge density distribution.

The remaining parts of this subsection provide details for some widely used charge density distributions $\rho(r)$, shown in Fig. 2, including expressions for the expectation values $\langle r^k \rangle$ and $\langle \ln(r) \rangle$ (as required for the moment function), and for the corresponding electrostatic potential energy functions $V_{\rm nuc}(r)$. The expression for the form factor $F(q)$ is included as well in some cases.

The definitions of special mathematical functions and expressions occurring at several places below are given in the appendix to this chapter.

Point-like charge density distribution:

$$\rho(r) = \rho_0 \delta(\boldsymbol{r}) = \frac{\rho_0}{4\pi r^2} \delta(r), \quad \rho_0 = Z. \tag{25}$$

Form factor:

$$F(q) = 1. \tag{26}$$

Potential energy function:

$$V_{\rm nuc}(r) = -\frac{Z}{r} \qquad (r > 0). \tag{27}$$

Homogeneous (or uniform) charge density distribution (see Fig. 2a): This piecewise defined distribution has a single parameter, the model-specific radial size parameter R_H.

$$\rho(r) = \rho_0\,\Theta(R_\mathrm{H} - r), \quad \rho_0 = \frac{3Z}{4\pi R_\mathrm{H}^3}. \tag{28}$$

Expectation values required to evaluate the moment function $M(k)$ (k integer):

$$\langle r^k\rangle = \frac{3}{k+3} R_\mathrm{H}^k \quad (k \geq -2), \qquad \langle \ln(r)\rangle = \ln(R_\mathrm{H}) - \frac{1}{3}. \tag{29}$$

The relation $R_\mathrm{H} = \sqrt{5/3}\ a$ adapts the model-specific radial size parameter R_H to the rms radius a of a given nuclear charge density distribution.
Form factor:

$$F(q) = \frac{3}{qR_\mathrm{H}} j_1(qR_\mathrm{H}). \tag{30}$$

Potential energy function:

$$V_\mathrm{nuc}(r) = \begin{cases} -\dfrac{3Z}{2R_\mathrm{H}}\left(1 - \dfrac{1}{3}\,\dfrac{r^2}{R_\mathrm{H}^2}\right) & 0 \leq r \leq R_\mathrm{H}, \\ -\dfrac{Z}{r} & r > R_\mathrm{H}. \end{cases} \tag{31}$$

Gauss-type charge density distribution (see Fig. 2b): This distribution has a single parameter, the model-specific radial size parameter R_G.

$$\rho(r) = \rho_0 \exp\left(-r^2/R_\mathrm{G}^2\right), \quad \rho_0 = \frac{1}{\Gamma(5/2)}\,\frac{3Z}{4\pi R_\mathrm{G}^3}. \tag{32}$$

Expectation values required to evaluate the moment function $M(k)$ (k integer):

$$\langle r^k\rangle = \frac{3}{k+3}\,\frac{\Gamma((k+5)/2)}{\Gamma(5/2)} R_\mathrm{G}^k \quad (k \geq -2), \quad \langle \ln(r)\rangle = \ln(R_\mathrm{G}/2) + 1 - \frac{\gamma_\mathrm{E}}{2}. \tag{33}$$

The relation $R_\mathrm{G} = \sqrt{2/3}\ a$ adapts the model-specific radial size parameter R_G to the rms radius a of a given nuclear charge density distribution.
Form factor:

$$F(q) = \exp(-q^2 R_\mathrm{G}^2/4). \tag{34}$$

Potential energy function:

$$V_{\rm nuc}(r) = -\frac{Z}{r}\,{\rm erf}(r/R_{\rm G}). \tag{35}$$

Fermi-type charge density distribution (see Fig. 2c): This distribution has two parameters, the model-specific radial size parameter $R_{\rm F}$ and a parameter $b (0 < b \ll 1)$ related to the skin thickness $t = R_{10} - R_{90}$.

$$\rho(r) = C\left\{1 + \exp\left(\frac{r - R_{\rm F}}{bR_{\rm F}}\right)\right\}^{-1} = \frac{C}{2}\left\{1 - \tanh\left(\frac{r - R_{\rm F}}{2bR_{\rm F}}\right)\right\}, \tag{36}$$

$$C = \rho_0 f = \frac{1}{6b^3\mathscr{F}_2(1/b)}\,\frac{3Z}{4\pi R_{\rm F}^3}, \quad \rho(0) = \rho_0 = \frac{C}{f}, \quad f = 1 + {\rm e}^{-1/b}, \tag{37}$$

$$\rho(R_{90}) = \frac{9}{10}\rho_0, \quad \rho(R_{\rm F}) = \frac{C}{2}, \quad \rho(R_{50}) = \frac{\rho_0}{2}, \quad \rho(R_{10}) = \frac{1}{10}\rho_0, \tag{38}$$

$$\frac{R_{50}}{R_{\rm F}} = 1 + b\ln(2f - 1) = 1 + b\ln(1 + {\rm e}^{-1/b}), \tag{39}$$

$$\frac{t}{R_{\rm F}} = b\ln\frac{90f - 9}{10f - 9} = b\ln\frac{81 + 90\,{\rm e}^{-1/b}}{1 + 10\,{\rm e}^{-1/b}}. \tag{40}$$

Expectation values required to evaluate the moment function $M(k)$ (k integer):

$$\langle r^k\rangle = \frac{(k+2)!\mathscr{F}_{k+2}(1/b)}{2\mathscr{F}_2(1/b)}(bR_{\rm F})^k \quad (k \geq -2), \tag{41}$$

$$\langle \ln(r)\rangle = \ln(bR_{\rm F}) + \frac{\mathscr{G}_2(1/b)}{\mathscr{F}_2(1/b)}. \tag{42}$$

From these expressions follows for the ratio of skin thickness t to rms radius a

$$\frac{t}{a} = \left(\frac{\mathscr{F}_2(1/b)}{12\mathscr{F}_4(1/b)}\right)^{1/2}\ln\frac{81 + 90\,{\rm e}^{-1/b}}{1 + 10\,{\rm e}^{-1/b}}. \tag{43}$$

The three ratios $R_{50}/R_{\rm F}$, $t/R_{\rm F}$, and t/a, each of them a function of the parameter b alone, are shown in Fig. 3. For sufficiently small values of b ($b < 1/10$),

$$R_{50} \approx R_{\rm F}, \quad \frac{t}{R_{\rm F}} \approx (4\ln(3))b, \quad \frac{t}{a} \approx \left(4\sqrt{\frac{5}{3}}\ln(3)\right)b. \tag{44}$$

The relations (41) and (43) can serve to adapt the two model-specific parameters to given data for nuclear charge rms radius a and skin thickness t. First, inversion of

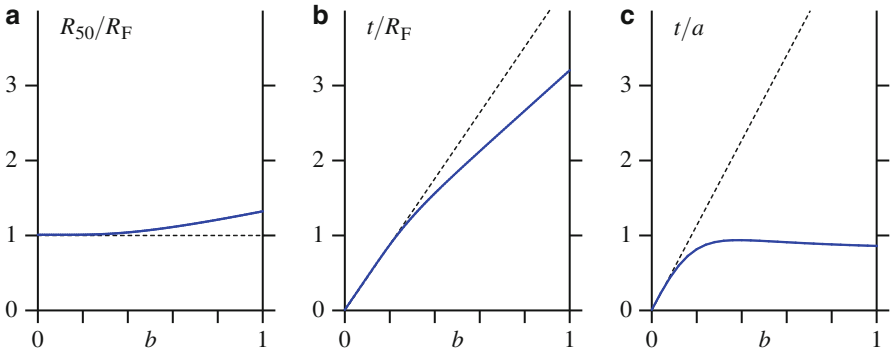


Fig. 3 Important length ratios for the Fermi-type distribution, shown as functions of parameter b. (**a**) R_{50}/R_F (see Eq. (39)), (**b**) t/R_F (see Eq. (40)), and (**c**) t/a (see Eq. (43)). The *dashed lines* represent the simple relations of Eq. (44)

Eq. (43) gives the parameter b, and then the model-specific nuclear size parameter R_F follows from Eq. (41) with $k = 2$.
Potential energy function:

$$V_{\text{nuc}}(r) = -\frac{Z}{r}\left\{1 - \frac{\mathscr{F}_2(1/b, z)}{\mathscr{F}_2(1/b)} + \frac{z}{2}\,\frac{\mathscr{F}_1(1/b, z)}{\mathscr{F}_2(1/b)}\right\}, \quad z = \frac{r}{bR_F}. \tag{45}$$

Fourier-Bessel charge density distribution (see Fig. 2d): This piecewise defined distribution is a "model-independent" representation of the nuclear charge density distribution as derived from electron scattering data [28] (see [16, 18] for sets of parameters). It vanishes exactly for $r \geq R_{\text{cut}}$.

$$\rho(r) = C\,\Theta(R_{\text{cut}} - r)\sum_{v=1}^{K} a_v\, j_0(q_v, r), \quad q_v R_{\text{cut}} = v\pi, \tag{46}$$

$$C = \frac{1}{S_0}\,\frac{Z}{4\pi R_{\text{cut}}^3}, \quad \rho(0) = C\sum_{v=1}^{K} a_v. \tag{47}$$

The values obtained from the finite expansion (46) are not strictly nonnegative. The coefficients a_v are usually given to only five significant digits (at most) and may alter in sign. A global factor C was introduced above to ensure exact normalization. Expectation values required to evaluate the moment function $M(k)$ (k integer):

$$\langle r^k\rangle = \frac{S_k}{S_0} R_{\text{cut}}^k \quad (k \geq -2), \quad \langle \ln(r)\rangle = \ln(R_{\text{cut}}) - \frac{1}{S_0}\sum_{v=1}^{K}\frac{a_v}{(v\pi)^3}\,\text{Si}(v\pi). \tag{48}$$

The auxiliary functions S_k are

$$S_k = \sum_{v=1}^{K} \frac{a_v}{(v\pi)^{k+3}} I_k(v\pi), \qquad I_k(v\pi) = \int_0^{v\pi} \mathrm{d}t\, t^{k+2} j_0(t). \tag{49}$$

The integrals $I_k(v\pi)$ can be evaluated recursively from

$$I_{-2}(v\pi) = \mathrm{Si}\,(v\pi), \qquad I_{-1}(v\pi) = 1 - (-1)^v, \tag{50}$$

$$I_k(v\pi) = -k(k+1) I_{k-2}(v\pi) - (-1)^v (v\pi)^{k+1} \quad (k \geq 0). \tag{51}$$

Potential energy function:

$$V_{\mathrm{nuc}}(r) = \begin{cases} -\dfrac{Z}{R_{\mathrm{cut}}} \left(1 + \dfrac{1}{S_0} \sum\limits_{v=1}^{K} \dfrac{a_v}{(v\pi)^2} j_0(q_v r) \right) & 0 \leq r \leq R_{\mathrm{cut}}, \\ -\dfrac{Z}{r} & r > R_{\mathrm{cut}}. \end{cases} \tag{52}$$

Sum-of-Gaussians charge density distribution (see Fig. 2e): This distribution is a "model-independent" representation of the nuclear charge density distribution as derived from electron scattering data [29] (see [16] for sets of parameters).

$$\rho(r) = C \sum_{v=1}^{K} a_v \left\{ \exp\left(-z_+^2\right) + \exp\left(-z_-^2\right) \right\}, \quad z_\pm = (r \pm R_v)/\gamma, \tag{53}$$

$$\rho(0) = 2C \sum_{v=1}^{K} a_v \exp\left(-\frac{R_v^2}{\gamma^2}\right), \quad a_v = \frac{1}{I_0(R_v/\gamma)} \frac{Q_v Z}{4\pi\gamma^3}, \quad \frac{1}{C} = \sum_{v=1}^{K} Q_v. \tag{54}$$

Each term of the finite expansion (53) is located at a different radius $r = R_v$ and carries a fraction Q_v of the total charge $Z (0 < Q_v < 1)$. A global factor C was introduced above to ensure exact normalization. The parameter γ determines the minimal width of representable structures. It is related to the full width at half-maximum (FWHM) value Γ of a nonsymmetrized Gauss-type function and to the rms radius a_{G} of a symmetrized Gauss function located at the origin ($R_v = 0$) through $\gamma = \Gamma / \left(2\sqrt{\ln(2)}\right) = \sqrt{2/3}\, a_{\mathrm{G}}$. Hence, every term of the sum-of-Gaussians expansion represents the charge density distribution only within a spherical shell of radius $r = R_v$ and of approximate width Γ.

Expectation values required to evaluate the moment function $M(k)$ (k integer):

$$\langle r^k \rangle = C \gamma^k \sum_{v=1}^{K} Q_v \frac{I_k(R_v/\gamma)}{I_0(R_v/\gamma)} \quad (k \geq -2). \tag{55}$$

The auxiliary functions

$$I_k(u) = \Gamma\left(\frac{k+3}{2}\right)\exp(-u^2)\,{}_1F_1\left(\frac{k+3}{2};\frac{1}{2};u^2\right) \tag{56}$$

can be evaluated recursively from

$$I_{-2}(u) = \sqrt{\pi}, \qquad I_{-1}(u) = \sqrt{\pi}\,u\,\mathrm{erf}(u) + \exp(-u^2), \tag{57}$$

$$I_k(u) = \frac{1}{2}(2k+1+2u^2)I_{k-2}(u) - \frac{1}{4}k(k-1)I_{k-4}(u) \quad (k \geq 0), \tag{58}$$

where it is understood that the last term of the recursion relation does not contribute for $k = 0$ and $k = 1$. In addition,

$$\langle \ln(r) \rangle = \ln(\gamma) + C\frac{\sqrt{\pi}}{4}\sum_{v=1}^{K}\frac{Q_v}{I_0(R_v/\gamma)}\exp\left(-\frac{R_v^2}{\gamma^2}\right)L(R_v/\gamma), \tag{59}$$

$$L(x) = \sum_{j=0}^{\infty}\psi(j+3/2)\frac{2j+1}{j!}x^{2j}. \tag{60}$$

Potential energy function:

$$V_{\mathrm{nuc}}(r) = -C\frac{\sqrt{\pi}}{4}\frac{Z}{r}\sum_{v=1}^{K}\frac{Q_v}{I_0(R_v/\gamma)}\left\{\left(1+2\frac{R_v^2}{\gamma^2}\right)\left[\mathrm{erf}(z_+)+\mathrm{erf}(z_-)\right]\right.$$
$$\left.+2\frac{R_v}{\gamma}\frac{r}{\gamma}\left[\mathrm{erf}(z_+)-\mathrm{erf}(z_-)\right]+\frac{2}{\sqrt{\pi}}\frac{R_v}{\gamma}\left[\exp(-z_+^2)-\exp(-z_-^2)\right]\right\}. \tag{61}$$

The last two, rather elaborate representations of the nuclear charge density distribution, the Fourier-Bessel distribution (Eq. (46)) and the sum-of-Gaussians distribution (Eq. (53)), have not been used in electronic structure calculations up to now, despite the fact that they are the most realistic distributions among those presented here. The main reason for this is that electronic structure calculations with finite nucleus models are certainly sensitive to the rms charge radius but rather insensitive to any other further parameters or details of the models. Therefore, the simple models, with just one or two parameters, are usually fully sufficient for the intended purpose. The examples shown below, in the following section, provide further support for these arguments.

Dipolar Magnetization Distributions

Magnetization distributions having just a magnetic dipole moment and vanishing higher magnetic multipolar moments are the simplest form of such distributions.

They are, hence, the magnetostatic equivalent to the spherically symmetric charge density distributions discussed above. The following presentation provides detailed general relations as well as expressions for five model distributions. It is based, in parts, on [30].

Given is a purely dipolar magnetization distribution, $\boldsymbol{m}(\boldsymbol{r}) = \boldsymbol{\mu}\, g(r) = \gamma \boldsymbol{I}\, g(r)$ (γ is the nuclear magnetogyric ratio, $\boldsymbol{I}$ is the nuclear spin), that is properly normalized

$$\int \mathrm{d}^3 r\; g(r) = 4\pi \int_0^\infty \mathrm{d}r\; r^2 g(r) = 1. \tag{62}$$

Radial expectation values of the scalar distribution function $g(r)$ are then defined as

$$\langle f(r) \rangle = 4\pi \int_0^\infty \mathrm{d}r\; r^2 f(r) g(r), \tag{63}$$

in particular

$$\langle r^k \rangle = 4\pi \int_0^\infty \mathrm{d}r\; r^{k+2} g(r), \quad k_{\min} \le k \le k_{\max}, \tag{64}$$

wherefrom an rms radius $\langle r^2 \rangle^{1/2}$ as a measure of the size of the magnetization distribution can be obtained. For subsequent use, a scalar auxiliary function $v(r)$ can be defined through

$$r\, v(r) = 4\pi \left(\int_0^r \mathrm{d}s\; s^2 g(s) + r \int_r^\infty \mathrm{d}s\; s g(s) \right), \tag{65}$$

with first derivative

$$\frac{\mathrm{d}v}{\mathrm{d}r} = -\frac{4\pi}{r^2} \int_0^r \mathrm{d}s\; s^2 g(s). \tag{66}$$

The close similarity between the Eqs. (62), (63), (64), (65), (66) and (11), (12), (13), (19) and (20) deserves to be mentioned.

The vector potential $\boldsymbol{A}(\boldsymbol{r})$, associated with the given magnetization distribution $\boldsymbol{m}(\boldsymbol{r})$, is then

$$\boldsymbol{A}(\boldsymbol{r}) = -\alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r} \frac{\mathrm{d}v}{\mathrm{d}r} = 4\pi \alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r^3} \int_0^r \mathrm{d}s\; s^2 g(s). \tag{67}$$

The remaining parts of this subsection provide details for five dipolar magnetization distributions $\boldsymbol{m}(\boldsymbol{r})$. In addition to the magnetization distribution for a point-like magnetic dipole, which has importance as a reference case, details are also given for four extended distributions, each depending on a single parameter. The exponential and Gauss-type dipolar magnetization distributions are in use

in present-day electronic structure calculations, the latter even more so than the former. Their modified variants may be used to model magnetization distributions exhibiting a maximum in the outer parts of the nucleus, instead of at its center, so that the distribution of the "unpaired" (or "valence") nucleon(s), i.e., the nucleon(s) primarily responsible for the nuclear magnetization, can be described in a more realistic way.

The definitions of special mathematical functions and expressions occurring at several places below are given in the appendix to this chapter.

Magnetization distribution of a point-like magnetic dipole:

$$\boldsymbol{m}(\boldsymbol{r}) = \gamma \boldsymbol{I} \delta(\boldsymbol{r}) = \gamma \boldsymbol{I} \frac{\delta(r)}{4\pi r^2}. \tag{68}$$

Vector potential:

$$\boldsymbol{A}(\boldsymbol{r}) = \alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r^3}. \tag{69}$$

Exponential dipolar magnetization distribution:

$$\boldsymbol{m}(\boldsymbol{r}) = \gamma \boldsymbol{I} \frac{\zeta^3}{8\pi} \exp(-\zeta\, r). \tag{70}$$

Expectation values of r^k of the scalar distribution function:

$$\langle r^k \rangle = \frac{\Gamma(k+3)}{\Gamma(3)} \frac{1}{\zeta^k} \quad (k \geq -2). \tag{71}$$

The relation $\zeta = \sqrt{12/\langle r^2 \rangle}$ adapts the model-specific parameter ζ to the mean square radius $\langle r^2 \rangle$ of a given nuclear magnetization distribution.
Vector potential:

$$\boldsymbol{A}(\boldsymbol{r}) = \alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r^3} \left(1 - \left[1 + x + \frac{x^2}{2} \right] \exp(-x) \right), \quad x = \zeta\, r. \tag{72}$$

Modified exponential dipolar magnetization distributions:

$$\boldsymbol{m}(\boldsymbol{r}) = \gamma \boldsymbol{I} \frac{1}{\Gamma(m+3)} \frac{\zeta^3}{4\pi} (\zeta\, r)^m \exp(-\zeta\, r) \quad (m \geq 0). \tag{73}$$

This expression includes Eq. (70) for $m = 0$.

Expectation values of r^k of the scalar distribution function:

$$\langle r^k \rangle = \frac{\Gamma(k+m+3)}{\Gamma(m+3)} \frac{1}{\zeta^k} \quad (k \geq -m-2). \tag{74}$$

The relation $\zeta = \sqrt{(m+3)\,(m+4)/\langle r^2 \rangle}$ adapts the model-specific parameter ζ to the mean square radius $\langle r^2 \rangle$ of a given nuclear magnetization distribution.
Vector potential:

$$\boldsymbol{A}(\boldsymbol{r}) = \alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r^3} P(m+3, \zeta\, r) = \alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r^3} \left(1 - Q(m+3, \zeta\, r)\right). \tag{75}$$

Gauss-type dipolar magnetization distribution:

$$\boldsymbol{m}(\boldsymbol{r}) = \gamma \boldsymbol{I} \left(\frac{\lambda}{\pi}\right)^{3/2} \exp(-\lambda\, r^2). \tag{76}$$

Expectation values of r^k of the scalar distribution function:

$$\langle r^k \rangle = \frac{\Gamma((k+3)/2)}{\Gamma(3/2)} \frac{1}{\lambda^{k/2}} \quad (k \geq -2). \tag{77}$$

The relation $\lambda = 3/(2\langle r^2 \rangle)$ adapts the model-specific parameter λ to the mean square radius $\langle r^2 \rangle$ of a given nuclear magnetization distribution.
Vector potential:

$$\boldsymbol{A}(\boldsymbol{r}) = \alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r^3} \left(\mathrm{erf}(x) - \frac{2}{\sqrt{\pi}} x\, \exp(-x^2)\right), \quad x = \sqrt{\lambda}\, r. \tag{78}$$

Modified Gauss-type dipolar magnetization distributions:

$$\boldsymbol{m}(\boldsymbol{r}) = \gamma \boldsymbol{I} \frac{1}{\Gamma((m+3)/2)} \frac{\lambda^{3/2}}{2\pi} \left(\sqrt{\lambda}\, r\right)^m \exp(-\lambda\, r^2) \quad (m \geq 0). \tag{79}$$

This expression includes Eq. (76) for $m = 0$.

Expectation values of r^k of the scalar distribution function:

$$\langle r^k \rangle = \frac{\Gamma((k+m+3)/2)}{\Gamma((m+3)/2)} \frac{1}{\lambda^{k/2}} \quad (k \geq -m-2). \tag{80}$$

The relation $\lambda = (m+3)/(2\langle r^2 \rangle)$ adapts the model-specific parameter λ to the mean square radius $\langle r^2 \rangle$ of a given nuclear magnetization distribution.

Vector potential:

$$\boldsymbol{A}(\boldsymbol{r}) = \alpha^2 \gamma \frac{\boldsymbol{I} \times \boldsymbol{r}}{r^3} P((m+3)/2, \lambda\, r^2). \tag{81}$$

The exponential and the Gauss-type dipolar magnetization distributions, Eqs. (70) and (76), are nowadays routinely applicable in electronic structure calculations. Their modified variants, given by Eqs. (73) and (79) with $m > 0$, may be useful for a more realistic modeling of the nuclear magnetization distribution. However, a meaningful comparison of different models requires accurate benchmark studies based on standardized magnetization distributions, possibly requiring more information on the nuclear magnetization distribution than just its mean square radius.

Finite Nucleus Models in Electronic Structure Calculations

The most important consequence for electronic structure calculations, due to the use of a finite nucleus model, is a modified electrostatic potential energy function, or Coulomb potential for short. The change from the Coulomb potential of a point-like nucleus, $V_{\mathrm{nuc}}^{\mathrm{PNC}}(r) = -Z/r$, with its singularity at $r = 0$, to some spherically symmetric potential energy function $V_{\mathrm{nuc}}^{\mathrm{FNC}}(r)$ with finite depth $V_{\mathrm{nuc}}^{\mathrm{FNC}}(0)(V_{\mathrm{nuc}}^{\mathrm{FNC}}(0) > -\infty$, see Table 3 and Fig. 4), represents the most important influence of the finite nucleus model on the problem of electronic structure. In comparison to this, any higher electric or magnetic nuclear multipole moment, if present, is of smaller influence. In terms of changes in total energy, the influence of such higher electric or magnetic nuclear multipole moments is smaller by several orders of magnitude. These multipole moments are of importance, of course, for various kinds of spectroscopy and for the understanding of finer details of electronic structure. They need to be properly modeled in a suitable way, if such details are accurately calculated by a purely theoretical approach.

The short-range behavior of the electrostatic potential energy functions $V_{\mathrm{nuc}}^{\mathrm{FNC}}(r)$, obtained from three different extended nuclear charge density distributions (properly standardized for the nuclide ^{200}Hg), is shown in Fig. 4 together with the function $V_{\mathrm{nuc}}^{\mathrm{PNC}}(r) = -Z/r (Z = 80)$ for the point-like nucleus case. The lower threshold of the upper continuum has been chosen as zero of energy. The position of the rms charge radius a and of the model-specific radial size parameters is indicated by

Table 3 Depth $V_{\mathrm{nuc}}^{\mathrm{FNC}}(0) = v_0 = -f\, Z/a$ ($f = \langle r^{-1}\rangle \langle r^2\rangle^{1/2}$, cf. Eq. (23)) of electrostatic potentials for three widely used finite nucleus models (the corresponding charge density distributions are standardized to a common rms charge radius a, see also Fig. 4)

Type	Homogeneous (See Eq. (28))	Gauss-type (See Eq. (32))	Fermi-type (See Eq. (36))
f	$3\sqrt{15}/10$	$\sqrt{6/\pi}$	$\sqrt{3}\frac{\mathscr{F}_1(1/b)}{\mathscr{F}_2(1/b)}\sqrt{\frac{\mathscr{F}_4(1/b)}{\mathscr{F}_2(1/b)}}$

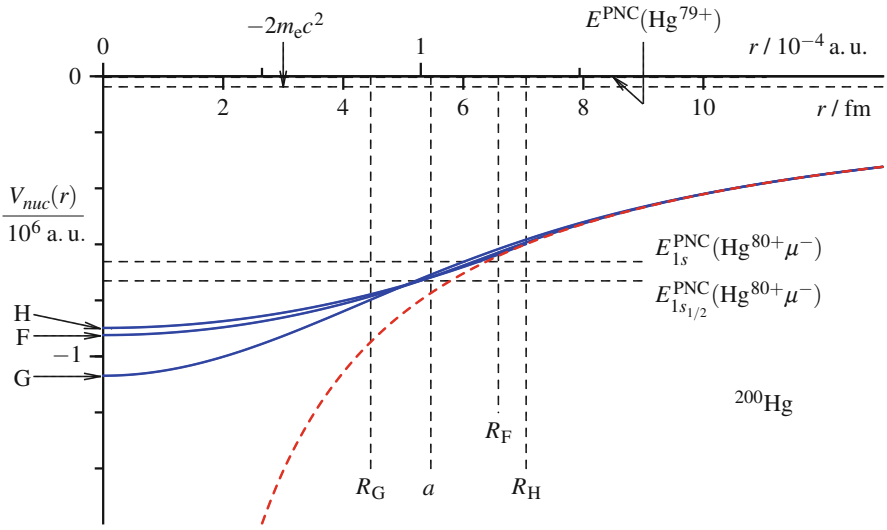


Fig. 4 Behavior of electrostatic potential energy functions $V_{\mathrm{nuc}}(r)$ for small r for the case of ^{200}Hg. Point-like nucleus case (PNC, *dashed red line*) and three different models for the finite nucleus case (FNC, *solid blue lines*): homogeneous distribution (H), Gauss-type distribution (G), and Fermi-type distribution (F). The latter three were standardized to the rms charge radius $a(^{200}\mathrm{Hg}) = 5.4551$ fm [20] (See text for further discussion)

vertical dashed lines. It is clearly seen that the potential energy functions obtained from the homogeneous and the Fermi-type nuclear charge density distributions differ only little, whereas the Gauss-type nuclear charge density distribution leads to a considerably deeper potential energy well. However, in comparison with the singular Coulomb potential of the point-like nucleus, the differences between potential energy functions from different finite nucleus models are always moderate [22]. Four energies are marked in Fig. 4 by horizontal dashed lines with labels. At highest energy, practically indistinguishable from the horizontal axis, is the (nonrelativistic or relativistic) ground-state energy of the hydrogen-like ion Hg^{79+} with a point-like nucleus. At the scale of the figure, any effects of finite nuclear charge density distributions on the total energy of this ion are invisible. Next lower in energy is the level of the upper boundary of the lower continuum of energies for systems composed of ordinary matter (or the negative of the threshold energy for electron-positron pair generation). Finally, two horizontal lines indicate the nonrelativistic and relativistic ground-state energies of the muonic ion $\mathrm{Hg}^{80+}\ \mu^-$ with a point-like nucleus. Even though the discussion of muonic systems is clearly beyond the scope of this chapter, it is instructive to see the huge difference between the ground-state energies of the hydrogen-like ion Hg^{79+} and its muonic analogon. One can easily predict that the change from a point-like nucleus model to a finite nucleus model will have a much larger influence on a muonic system than on the corresponding electronic system. Nevertheless, the change from the PNC to a FNC

has important consequences also in electronic structure calculations. This can be seen clearly already in atomic structure calculations, as will be shown now.

In a nonrelativistic approach [31], based on the Schrödinger equation, solutions (spin orbitals) exist for any proton number Z in both cases, the PNC and the FNC. Within the central field approximation and with equivalence restriction, the spatial part of the spin orbitals

$$\psi_i(\boldsymbol{r}) = r^{-1} P_{nl}(r) Y_{lm}(\hat{\boldsymbol{r}}) \tag{82}$$

(with composite index $i = (nlm)$, n and l are the usual principal and angular momentum quantum numbers, $-l \leq m \leq l$, $Y_{lm}(\hat{\boldsymbol{r}})$ is a standard spherical harmonic, $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$), contains the radial function $P_{nl}(r)$ that fulfills the boundary conditions $P_{nl}(0) = \lim_{r\to\infty} P_{nl}(r) = 0$. The power series expansion of this radial function for small r is

$$P_{nl}(r) = r^{l+1} \sum_{k=0}^{\infty} a_k r^k = r^{l+1}(a_0 + a_1 r + a_2 r^2 + O(r^3)), \qquad a_0 \neq 0. \tag{83}$$

The ratio a_1/a_0 is known as the cusp value of the radial function. Upon change from the PNC to a FNC, it changes from $(a_1/a_0)_{\mathrm{PNC}} = -Z/(l+1)$ to $(a_1/a_0)_{\mathrm{FNC}} = 0$.

In a relativistic treatment [32], based on the Dirac equation, solutions have the form of four-component vector-like functions (spinors),

$$\psi_i(\boldsymbol{r}) = \frac{1}{r} \begin{pmatrix} P_{n\kappa}(r)\, \Omega_{\kappa m}(\hat{\boldsymbol{r}}) \\ \mathrm{i}\, Q_{n\kappa}(r)\, \Omega_{-\kappa m}(\hat{\boldsymbol{r}}) \end{pmatrix}, \quad \mathrm{i}^2 = -1 \tag{84}$$

(with composite index $i = (n\kappa m)$, n is the principal quantum number and the nonzero integer κ is the relativistic symmetry quantum number, $j = l \pm 1/2 = |\kappa| - 1/2$, $-j \leq m \leq j$), where $\Omega_{\kappa m}(\hat{\boldsymbol{r}})$ denotes a standard normalized two-component angular spinor [33]. The pair of radial functions fulfills the conditions

$$\begin{pmatrix} P_{n\kappa}(0) \\ Q_{n\kappa}(0) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \lim_{r\to\infty} \begin{pmatrix} P_{n\kappa}(r) \\ Q_{n\kappa}(r) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{85}$$

In the PNC, the power series expansions of the radial functions $P_{n\kappa}(r)$ and $Q_{n\kappa}(r)$ for small r are

$$\begin{pmatrix} P_{n\kappa}(r) \\ Q_{n\kappa}(r) \end{pmatrix}_{\mathrm{PNC}} = r^{\gamma} \left\{ \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} r + O(r^2) \right\}, \quad \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{86}$$

where $\gamma = \sqrt{\kappa^2 - (\alpha Z)^2}$ is noninteger ($0 < \gamma < |\kappa|$, $0 < \alpha Z < 1$). The ratio $(b_0/a_0)_{\mathrm{PNC}} = -\alpha Z/(\gamma - \kappa) = +(\gamma + \kappa)/(\alpha Z)$ is the analogue of the nonrelativistic cusp value. A point-like nucleus thus induces nonanalytic behavior of the radial functions for small r. As a consequence, the radial parts of the spinors with $|\kappa| = 1$ ($j = 1/2$), i. e., the parts $r^{-1} P_{n,\pm 1}(r)$ and $r^{-1} Q_{n,\pm 1}(r)$ of the $n\mathrm{s}_{1/2}$ and $n\mathrm{p}_{1/2}$

spinors (cf. Eq. (84)) exhibit a mild singularity $\sim r^{\gamma-1}$ at $r = 0$. These solutions eventually cease to exist when the nuclear charge number Z reaches and surpasses the limit imposed by $\alpha Z = 1$. Upon change from the PNC to a FNC, one has, instead of Eq. (86),

– for $\kappa < 0$ ($j = l + 1/2$, i.e., $s_{1/2}$, $p_{3/2}$, $d_{5/2}$, $f_{7/2}$ etc.):

$$\begin{pmatrix} P_{n\kappa}(r) \\ Q_{n\kappa}(r) \end{pmatrix}_{\mathrm{FNC}} = r^{l+1} \left\{ \begin{pmatrix} a_0 \\ 0 \end{pmatrix} + \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} r + O(r^2) \right\}, \qquad a_0 \neq 0, \tag{87}$$

– and for $\kappa > 0$ ($j = l - 1/2$, i.e., $p_{1/2}$, $d_{3/2}$, $f_{5/2}$ etc.):

$$\begin{pmatrix} P_{n\kappa}(r) \\ Q_{n\kappa}(r) \end{pmatrix}_{\mathrm{FNC}} = r^{1} \left\{ \begin{pmatrix} 0 \\ b_0 \end{pmatrix} + \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} r + O(r^2) \right\}, \qquad b_0 \neq 0. \tag{88}$$

The restriction $\alpha Z < 1$ no longer applies in the FNC, and the power series expansion of the radial function $P_{n\kappa}(r)$ starts now with the same lowest power (r^{l+1}) as the one of its nonrelativistic counterpart $P_{nl}(r)$. It also deserves mention that every second term in each row of the power series expansions of Eqs. (87) and (88) vanishes if and only if the power series expansion for $V_{\mathrm{nuc}}(r)$, Eq. (23), contains only even powers.

The relativistic and nonrelativistic radial functions for the ground state ($n = 1$) of hydrogen-like atoms with point-like nucleus (PNC) and $Z = 1$, 80 and 137 are shown in Fig. 5. For $Z = 1$, there is no visible difference between the relativistic radial function $P_{1,-1}(r)$ and its nonrelativistic counterpart $P_{10}(r)$, and the radial function $Q_{1,-1}(r)$ has been enlarged twentyfold to make it clearly visible and thus distinguishable from the horizontal axis. At $Z = 80$, the three radial functions can be clearly identified and distinguished from each other, and the relativistic radial function $P_{1,-1}(r)$ reaches its maximum at slightly smaller radius than the nonrelativistic radial function $P_{10}(r)$. Finally, in the case of the largest allowed proton number, $Z = 137 < 1/\alpha$, the two relativistic radial functions are almost exactly of same magnitude, and the nonanalytic behavior (see Eq. (86)) is very clearly exhibited (the behavior of $P_{1,-1}(r)$ and $Q_{1,-1}(r)$ for very small r, almost vertical increase/decrease, is neither an artifact nor an error of the graphics software). In contrast, the nonrelativistic radial function still exhibits the analytic behavior expressed by Eq. (83). While the conditions $P_{10}(0) = 0$, $P_{1,-1}(0) = 0$, and $Q_{1,-1}(0) = 0$ are always fulfilled, only the nonrelativistic radial function $P_{10}(r)$ retains its linear behavior ($\sim r$) for small r upon increase of Z. In contrast, the relativistic radial function $P_{1,-1}(r)$, which behaves $\sim r^{\gamma}$ for small r, becomes steeper as Z increases ($\gamma \approx 1$ for $Z = 1$, $\gamma \approx 0$ for $Z = 137$). With a finite nucleus (FNC), the behavior of the relativistic radial function $P_{1,-1}(r)$ for small r would resemble more closely its nonrelativistic counterpart $P_{10}(r)$.

An illustrative example for the consequences of the change from the PNC to a FNC in relativistic electronic structure calculations is shown in Fig. 6, where energy differences between low-lying electronic states of lithium-like and beryllium-

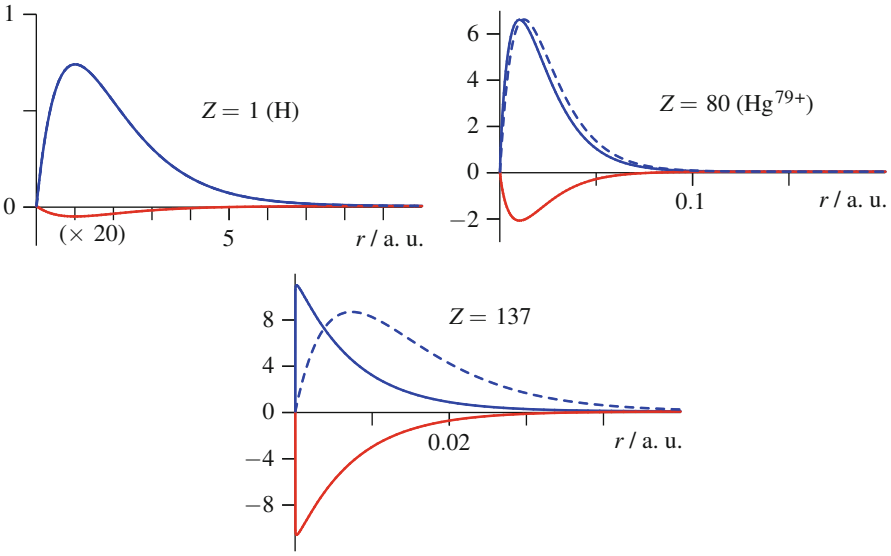


Fig. 5 Normalized radial functions of the electronic ground state ($n = 1$) of hydrogen-like atoms with point-like nucleus and $Z = 1$, 80, 137: $P_{1,-1}(r)$ (*solid blue lines*), $Q_{1,-1}(r)$ (*solid red lines*); $P_{10}(r)$ (*dashed blue lines*). See Eq. (82) for $P_{10}(r)$ and Eqs. (84) and (86) for $P_{1,-1}(r)$ and $Q_{1,-1}(r)$, and see text for further discussion

like high-Z ions ($100 \leq Z \leq 120$) are considered. The data were obtained from numerical finite-difference Dirac-Fock-Coulomb calculations [34]. For both isoelectronic sequences, the studied energy differences are monotonously increasing functions of the nuclear charge number Z, if a point-like nucleus model is used. With a finite nucleus model, these energy differences are smaller than in the PNC at $Z = 100$ and only slowly increasing at first, then they are gradually decreasing, and finally they are being inverted as $Z = 120$ is reached. Reasons for this behavior are, on one hand, the differences in the short-range behavior of the radial functions (r^γ in the PNC versus r^{l+1} in the FNC) and, on the other hand, the changes occurring upon increase of Z in the relative magnitude of the radial functions $P_{n\kappa}(r)$ and $Q_{n\kappa}(r)$ of the 2s and $2p_{1/2}$ spinors in the FNC (for the 2s spinor $P_{2,-1}(r) \sim r$ and $Q_{2,-1}(r) \sim r^2$, whereas for the $2p_{1/2}$ spinor $P_{2,+1}(r) \sim r^2$ and $Q_{2,+1}(r) \sim r$, see Eqs. (87) and (88)). It deserves mention that the differences seen between results obtained with different FNC models, here in particular between Gauss-type and Fermi-type nuclear charge density distributions, are very small. The Gauss-type distribution, which yields a potential well of considerably greater depth than the Fermi-type distribution (see Fig. 4), leads to energy differences that deviate only little from those obtained with the Fermi-type distribution. One can conclude (i) that it is preferable to use a properly parametrized FNC model of the nuclear charge density distribution instead of the PNC, because the latter is not realistic, and (ii) that the details of this FNC model are of lesser importance for electronic structure

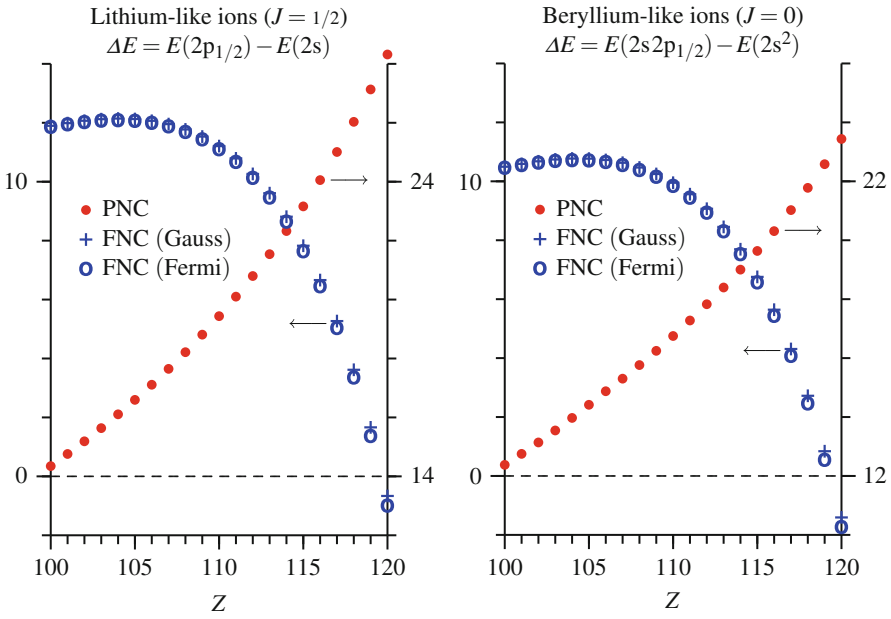


Fig. 6 Total energy differences ΔE (in atomic units) between low-lying electronic states of lithium-like ions (*left*) and beryllium-like ions (*right*) with $100 \leq Z \leq 120$. Data from numerical finite-difference Dirac-Fock-Coulomb calculations with point-like nucleus (PNC, *right axes*) and with two finite nucleus models (FNC, *left axes*) (Adapted from [34])

calculations as long as its parameters (rms charge radius a, skin thickness t) are standardized, e.g., to experimental data.

The use of finite nucleus models is mandatory for realistic relativistic electronic structure calculations on compounds containing heavy or superheavy atoms. It is, therefore, necessary to use such models also in the generation of reference data for the adjustment of relativistic pseudopotentials (or effective core potentials, ECPs), which represent a convenient means for inclusion of scalar relativistic effects in routine electronic structure calculations [35, 36].

Modified Coulomb potentials $V_{\mathrm{nuc}}(r)$ can be used not only in numerical (basis-set-free) electronic structure calculations, as shown by the examples presented above, but also in algebraic approaches using analytic basis functions. This requires the evaluation of appropriate matrix elements (nuclear attraction integrals):

$$\langle \phi_i(\boldsymbol{r}_A) \,|V_{\mathrm{nuc}}(\boldsymbol{r}_C)|\, \phi_j(\boldsymbol{r}_B)\rangle = \int \mathrm{d}^3 r \phi_i^{\dagger}(\boldsymbol{r}_A) V_{\mathrm{nuc}}(\boldsymbol{r}_C)\, \phi_j(\boldsymbol{r}_B), \tag{89}$$

with $\boldsymbol{r}_X = \boldsymbol{r} - \boldsymbol{R}_X$, where $\boldsymbol{R}_X$ usually denotes the positon vector of an atomic nucleus X. This notation covers both the nonrelativistic and relativistic cases (scalar orbitals and four-component spinors, respectively); the indices i and j carry information

to identify the basis functions unambiguously. The integrals in Eq. (89) may involve only a single center ($A = B = C$ ["atomic" integrals]), two centers, or three centers (A, B, C all different). The difficulty of their evaluation increases with the number of centers. In addition, every type of basis functions requires its own implementation of nuclear attraction integrals. The necessary algorithms for the efficient evaluation of these integrals have been developed only for some of the very many combinations of basis function types and Coulomb potentials. However, the implementation of a finite nucleus model requires only little or moderate modification if both the basis functions and the FNC model rely on Gauss-type functions, due to the fact that almost all of the required functions and terms are already available from the previous implementation adapted to the point-like nucleus case. This is the main reason for the widespread and frequent use of the Gauss-type FNC model in electronic structure calculations using Gauss-type basis functions. A large collection of basis sets, including basis set sequences suitable for systematic improvement of results from four-component relativistic all-electron calculations on compounds containing heavy s-, p-, d-, and f-block elements, has been generated by Dyall (see [37] and references cited therein). All these basis sets, which are constructed from Gauss-type functions, were optimized in Dirac-Fock-Coulomb calculations using standardized Gauss-type finite nucleus models.

There are several other effects of finite nucleus models in relativistic electronic structure calculations, beyond their influence via modified Coulomb potentials. The most prominent ones shall be mentioned here at least (see also the recent reviews [38, 39]).

Within a formally nonrelativistic approach, relativistic effects may be included through correction terms, frequently treated as perturbations. Those correction terms that are to be modified when a finite nucleus model is used include the electron-nucleus Darwin term [7] and terms contributing to hyperfine splitting (electron-spin-nuclear-spin dipolar interaction and Fermi contact term) [30].

Within the two- and four-component relativistic approaches to electronic structure, the effect of a finite nucleus model is included most conveniently by the use of the appropriate modified electrostatic potential energy functions and vector potentials, as stated above. The available tools are now so far and so well developed that the influence of extended nuclear charge density and magnetization distributions on, e.g., hyperfine structure and EPR and NMR parameters can be studied. For some recent examples, see [40–42].

The use of a finite nucleus model has also influence on quantum electrodynamic (QED) effects (vacuum polarization, self-energy correction) and on the study of the electron-nucleon weak interaction (parity nonconservation). For example, the lowest-order vacuum polarization contribution, known as Uehling potential, requires modification when an extended nuclear charge density distribution is used [43, 44]. The nuclear number density distributions (see Eq. (1)) are required in studies of the electron-nucleon weak interaction [45, 46].

Summary

In electronic structure calculations for atoms, molecules, or periodic systems, the atomic nuclei play the role of centers of force attracting the electrons and repelling other nuclei. Accurate calculations of energies, electronic wavefunctions, and properties usually require to model the atomic nuclei as extended (or finite) instead of point-like objects (finite nucleus case, FNC, *vs*. point-like nucleus case, PNC). For a nuclide with given proton number Z and neutron number N, the nuclear charge density and magnetization distributions can be modeled in a phenomenological way with suitably parametrized functions (it is not required that some sort of nuclear structure calculation precedes the electronic structure calculation). Most important for practical work are spherically symmetric nuclear charge density distributions and purely dipolar nuclear magnetization distributions, normalized, respectively, to total nuclear charge and nuclear magnetic dipole moment. Several distributions are discussed in detail, including the Gauss-type and Fermi-type distributions for the charge density and the Gauss-type distribution for the magnetization.

The use of an extended instead of a point-like model for the atomic nucleus, properly standardized to the size (and possibly also the shape) of that nucleus, is of much more importance in relativistic than in nonrelativistic approaches. It is shown that the change in the attractive electron-nucleus potential energy function, due to the replacement of the PNC by a FNC, leads to large modifications in the small-r behavior of the relativistic wavefunction, especially for nuclei with high Z. In comparison, the small-r behavior of nonrelativistic wavefunctions is only slightly modified by the change from the PNC to a FNC. It is also shown that differences between different FNC models are almost always of only minor importance. Of course, studies of the influence of the change from the PNC to the FNC on quantum electrodynamic (QED) effects or on electron-nucleon weak interaction may require a more elaborate modeling of atomic nuclei (including, e.g., the nuclear mass density distribution).

Appendix

This appendix provides definitions for mathematical expressions and special functions used in the main part of this chapter. Unless stated otherwise, further details can be found in standard references [47, 48].

Euler-Mascheroni constant γ_{E}:

$$\gamma_{\mathrm{E}} = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln(n)\right) = 0.57721\ldots. \tag{90}$$

Heaviside step function $\Theta(x)$:

$$\Theta(x) = \begin{cases} 0 & x < 0, \\ 1/2 & x = 0, \\ 1 & x > 0. \end{cases} \tag{91}$$

Gamma function $\Gamma(a)$ and its logarithmic derivative:

$$\Gamma(a) = \int_0^\infty \mathrm{d}t\, t^{a-1}\mathrm{e}^{-t} \quad (a > 0), \quad \Gamma(a+1) = a\Gamma(a), \quad \Gamma(1) = 1. \tag{92}$$

$$\psi(a) = \frac{\Gamma'(a)}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_0^\infty \mathrm{d}t\, t^{a-1} \ln(t)\mathrm{e}^{-t}, \qquad \psi(1) = -\gamma_\mathrm{E}. \tag{93}$$

Incomplete gamma functions $P(a,x)$ and $Q(a,x)$ and error function $\mathrm{erf}(x)$:

$$P(a,x) = \frac{1}{\Gamma(a)} \int_0^x \mathrm{d}t\, t^{a-1}\mathrm{e}^{-t}, \quad Q(a,x) = \frac{1}{\Gamma(a)} \int_x^\infty \mathrm{d}t\, t^{a-1}\mathrm{e}^{-t} \quad (a > 0). \tag{94}$$

For a equal to a positive integer ($a = n+1, n \geq 0$):

$$P(n+1,x) = 1 - Q(n+1,x) = 1 - \mathrm{e}^{-x} \sum_{k=0}^{n} \frac{x^k}{k!}. \tag{95}$$

For $2a$ equal to an odd positive integer ($2a = 2k+1, k \geq 0$):

$$P(1/2, x^2) = \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}t\ \exp(-t^2). \tag{96}$$

$$\begin{aligned} P((m+3)/2, x^2) &= \frac{2}{\Gamma((m+3)/2)} \int_0^x \mathrm{d}t\, t^{m+2} \exp(-t^2) \qquad (m = 0, 2, 4, \ldots) \\ &= \mathrm{erf}(x) - \frac{2x}{\sqrt{\pi}} \exp(-x^2) \sum_{k=0}^{m/2} \frac{(2x^2)^k}{(2k+1)!!}. \end{aligned} \tag{97}$$

Confluent hypergeometric function ${}_1F_1(a;c;x)$:

$${}_1F_1(a;c;x) = \sum_{k=0}^{\infty} \frac{(a)_k}{(c)_k} \frac{x^k}{k!}, \quad (a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}. \tag{98}$$

Spherical Bessel functions of the first kind $j_n(x)$:

$$j_n(x) = (-x)^n \left(\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \frac{\sin(x)}{x} \qquad (n \geq 0). \tag{99}$$

The first two members of this sequence are

$$j_0(x) = \frac{\sin(x)}{x}, \qquad j_1(x) = \frac{\sin(x)}{x^2} - \frac{\cos(x)}{x}. \tag{100}$$

Sine integral $\mathrm{Si}(x)$:

$$\mathrm{Si}(x) = \int_0^x \mathrm{d}t \, \frac{\sin(t)}{t}. \tag{101}$$

Fermi-Dirac integrals $\mathscr{F}_j(a)$ and $\mathscr{F}_j(a,x)$ and related functions:

$$\mathscr{F}_j(a) = \frac{1}{\Gamma(j+1)} \int_0^\infty \mathrm{d}t \, \frac{t^j}{1+\exp(t-a)}, \tag{102}$$

$$\mathscr{F}_j(a,x) = \frac{1}{\Gamma(j+1)} \int_x^\infty \mathrm{d}t \, \frac{t^j}{1+\exp(t-a)}, \tag{103}$$

$$\mathscr{G}_j(a) = \frac{1}{\Gamma(j+1)} \int_0^\infty \mathrm{d}t \, \frac{t^j \ln(t)}{1+\exp(t-a)} = \psi(j+1)\mathscr{F}_j(a) + \frac{\partial}{\partial j}\mathscr{F}_j(a). \tag{104}$$

A general algorithm for the evaluation of complete and incomplete Fermi-Dirac integrals has been published [49, 50].

References

1. Mohr PJ, Newell DB, Taylor BN (2015) CODATA recommended values of the fundamental physical constants: 2014. Zenodo. doi:10.5281/zenodo.22826
2. Brown JM, Buenker RJ, Carrington A, Di Lauro C, Dixon RN, Field RW, Hougen JT, Hüttner W, Kuchitsu K, Mehring M, Merer AJ, Miller TA, Quack M, Ramsay DA, Veseth L, Zare RN (2000) Remarks on the signs of g factors in atomic and molecular Zeeman spectroscopy. Mol Phys 98:1597–1601. doi:10.1080/00268970009483366
3. Olive KA et al (Particle Data Group) (2014) Review of particle physics. Chin Phys C 38:090001. doi:10.1088/1674-1137/38/9/090001 (see http://pdg.lbl.gov/ for the 2015 update)
4. Magill J, Pfennig G, Dreher R, Sóti Z (2012) Karlsruher Nuklidkarte / Karlsruhe Chart of the Nuclides, 8th edn. Nucleonica, Eggenstein-Leopoldshafen
5. Brookhaven National Laboratory, National Nuclear Data Center, Chart of Nuclides. http://www.nndc.bnl.gov/chart/. Accessed 11 Oct 2014
6. Ramsey NF (1953) Nuclear moments. Wiley, New York
7. Weissbluth M (1978) Atoms and molecules. Academic, New York

8. Schwerdtfeger P, Pernpointner M, Nazarewicz (2004) Calculation of Nuclear Quadrupole Coupling Constants. In: Kaupp M, Bühl M, Malkin VG (eds) Calculation of NMR and EPR parameters. Theory and application. Wiley-VCH, Weinheim, pp 279–291
9. Stone NJ (2005) Table of nuclear magnetic dipole and electric quadrupole moments. At Data Nucl Data Tables 90:75–176. doi:10.1016/j.adt.2005.04.001
10. Pyykkö P (2008) Year-2008 nuclear quadrupole moments. Mol Phys 106:1965–1974. doi: 10.1080/00268970802018367
11. Stone NJ (2015) New table of recommended nuclear electric quadrupole momnets. Hyperfine Interact 230: 7–16. doi:10.1007/s10751-014-1094-8
12. Cohen ER, Cvitaš T, Frey JG, Holmström B, Kuchitsu K, Marquardt R, Mills I, Pavese F, Quack M, Stohner J, Strauss HL, Takami M, Thor AJ (2008) Quantities, Units and Symbols in Physical Chemistry, 3rd ed., 2nd printing. IUPAC & RSC Publishing, Cambridge, pp 121–128
13. Holden NE (2014) Table of the isotopes. In: Haynes WM (ed) CRC Handbook of chemistry and physics, 95th edn. CRC Press, Boca Raton, pp 11-2–11-174
14. de Jager CW, de Vries H, de Vries C (1974) Nuclear charge-and magnetization-density-distribution parameters from elastic electron scattering. At Data Nucl Data Tables 14:479–508. doi:10.1016/S0092-640X(74)80002-1
15. Engfer R, Schneuwly H, Vuilleumier JL, Walter HK, Zehnder A (1974) Charge-distribution parameters, isotope shifts, isomer shifts, and magnetic hyperfine constants from muonic atoms. At Data Nucl Data Tables 14:509–597. doi:10.1016/S0092-640X(74)80003-3
16. de Vries H, de Jager CW, de Vries C (1987) Nuclear charge-density-distribution parameters from elastic electron scattering. At Data Nucl Data Tables 36:495–536. doi:10.1016/0092-640X(87)90013-1
17. Collard HR, Elton LRB, Hofstadter R (1967) Nuclear Radii. In: Landolt-Börnstein numerical data and functional relationships in science and technology – new series, vol I/2. Springer, Berlin
18. Fricke G, Bernhardt C, Heilig K, Schaller LA, Schellenberg L, Shera EB, de Jager CW (1995) Nuclear ground state charge radii from electromagnetic interactions. At Data Nucl Data Tables 60:177–285. doi:10.1006/adnd.1995.1007
19. Fricke G, Heilig K (2004) Nuclear charge radii. In: Landolt H, Börnstein R (eds) Numerical data and functional relationships in science and technology – new series, vol I/20, Springer, Berlin
20. Angeli I, Marinova KP (2013) Table of experimental nuclear ground state charge radii: an update. At Data Nucl Data Tables 99:69–95. doi:10.1016/j.adt.2011.12.006
21. Johnson WR, Soff G (1985) The Lamb shift in Hydrogen-Like Atoms, $1 \leq Z \leq 110$. At Data Nucl Data Tables 33:405–446. doi:10.1016/0092-640X(85)90010-5
22. Andrae D (2000) Finite nuclear charge density distributions in electronic structure calculations for atoms and molecules. Phys Rep 336:413–525. doi:10.1016/S0370-1573(00)00007-7
23. Nerlo-Pomorska B, Pomorski K (1994) Simple formula for nuclear charge radius. Z Phys A 348:169–172. doi:10.1007/BF01291913
24. Visscher L, Dyall KG (1997) Dirac-Fock atomic electronic structure calculations using different nuclear charge distributions. At Data Nucl Data Tables 67:207–224. doi:10.1006/-adnd.1997.0751
25. Jackson JD (1998) Classical electrodynamics, 3rd edn. Wiley, New York
26. Andrae D (2002) Nuclear charge density distributions in quantum chemistry. In: Schwerdtfeger P (ed) Relativistic electronic structure theory. Part 1: fundamentals. Elsevier, Amsterdam, pp 203–258. doi:10.1016/S1380-7323(02)80030-9
27. Friedrich J, Lenz F (1972) Elastic electron scattering from ^{208}Pb at moderate momentum transfers and model-independent description of the nuclear charge distribution. Nucl Phys A 183:523–544. doi:10.1016/0375-9474(72)90354-5
28. Dreher B, Friedrich J, Merle K, Rothhaas H, Lührs G (1974) The determination of the nuclear ground state and transition charge density from measured electron scattering data. Nucl Phys A 235:219–248. doi:10.1016/0375-9474(74)90189-4

29. Sick I (1974) Model-independent nuclear charge densities from elastic electron scattering. Nucl Phys A 218:509–541. doi:10.1016/0375-9474(74)90039-6
30. Moore EA, Moss RE (1975) Finite nucleus models for molecular calculations. Mol Phys 30:1315–1323. doi:10.1080/00268977500102861
31. Andrae D, Hinze J (1997) Numerical electronic structure calculations for atoms. I. Int J Quantum Chem 63:65–91
32. Andrae D, Reiher M, Hinze J (2000) Numerical electronic structure calculations for atoms. II. Int J Quantum Chem 76:473–499
33. Louck JD (1996) Angular momentum theory. In: Drake GWF (ed) Atomic, molecular & optical physics handbook. American Institute of Physics, Woodbury, pp 6–55
34. Andrae D, Reiher M, Hinze J (2000) A comparative study of finite nucleus models for low-lying states of few-electron high-Z atoms. Chem Phys Lett 320:457–468. doi:10.1016/S0009-2614(00)00068-3
35. Cao X, Dolg M (2010) Relativistic Pseudopotentials. In: Barysz M, Ishikawa Y (eds) Relativistic methods for chemists. Springer, Dordrecht, pp 215–278
36. Dolg M, Cao X (2012) Relativistic pseudopotentials: their development and scope of applications. Chem Rev 112:403–480. doi:10.1021/cr2001383
37. Dyall KG, Relativistic basis sets. http://dirac.chem.sdu.dk/basisarchives/dyall/. Accessed 23 Aug 2014
38. Pyykkö P (2012) The physics behind chemistry and the periodic table. Chem Rev 112:371–384. doi:10.1021/cr200042e
39. Autschbach J (2012) Perspective: relativistic effects. J Chem Phys 136:150902. doi:10.1063/1.3702628
40. Autschbach J (2014) Relativistic calculations of magnetic resonance parameters: background and some recent developments. Philos Trans R Soc A 372:20120489. doi:10.1098/rsta.2012.0489
41. Malkin E, Repiský M, Komorovský S, Mach P, Malkina OL, Malkin VG (2011) Effects of finite size nuclei in relativistic four-component calculations of hyperfine structure. J Chem Phys 134:044111. doi:10.1063/1.3526263
42. Maldonado AF, Giménez CA, Aucar GA (2012) Nuclear charge-distribution effects on the NMR spectroscopy parameters. J Chem Phys 136:224110. doi:10.1063/1.4729253
43. Hylton DJ (1985) Finite-nuclear-size corrections to the Uehling potential. Phys Rev A 32:1303–1309
44. Beier T, Mohr PJ, Persson H, Soff G (1998) Influence of nuclear size on QED corrections in hydrogenlike heavy ions. Phys Rev A 58:954–963
45. Berger R, Quack M (2000) Multiconfiguration linear response approach to the calculation of parity violating potentials in polyatomic molecules. J Chem Phys 112:3148–3158
46. Laerdahl JK, Schwerdtfeger P, Quiney HM (2000) Theoretical analysis of parity-violating energy-differences between enantiomers of chiral molecules. Phys Rev Lett 84:3811–3814
47. Abramowitz M, Stegun IA (eds) (1972) Handbook of mathematical functions with formulas, graphs, and mathematical tables, 9th printing. Dover, New York
48. Olver FWJ, Lozier DW, Boisvert RF, Clark CW (eds) (2010) NIST handbook of mathematical functions. Cambridge University Press, Cambridge
49. Goano M (1993) Series expansion of the Fermi-Dirac integral $\mathscr{F}_j(x)$ over the entire domain of real j and x. Solid-State Electron 36:217–221. doi:10.1016/0038-1101(93)90143-E
50. Goano M (1995) Algorithm 745: Computation of the complete and incomplete Fermi-Dirac integral. ACM Trans Math Softw 21:221–232. See also ACM Trans Math Softw 23:295

One-Particle Basis Sets for Relativistic Calculations

3

Kenneth G. Dyall

Contents

Introduction 84
Kinetic Balance 85
Spin Coupling 90
Radial Behavior 94
Primitive Gaussian Basis Sets 96
Contracted Basis Sets 99
Magnetic Balance 101
References 105

Abstract

Basis sets for relativistic calculations must satisfy conditions beyond those required for nonrelativistic calculations. At the four-component level, they must satisfy the proper symmetry and functional relationships between the large and small components. For calculations in electrostatic fields, they must satisfy the *kinetic balance* condition; when magnetic fields are added, they must satisfy the *magnetic balance* condition. The basis sets must incorporate the effects of the scalar and spin-dependent relativistic operators on the structure, whether in four-component or two-component calculations. The use of contracted basis sets introduces other requirements related to the method of solution of the self-consistent field equations and the issue of linear dependence of the basis functions.

Keywords

Basis set • Kinetic balance • Magnetic balance • Gaussian function • Point nucleus • Finite nucleus • 2-spinor • 4-spinor

K.G. Dyall (✉)
Dirac Solutions, Portland, OR, USA
e-mail: diracsolutions@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_25

Introduction

The basis sets used in quantum chemistry have to satisfy a few conditions in order to be useful. As the basis sets are used to expand the wave function as a linear combination of functions of known form, the boundary conditions on the wave function must also apply in some way to the basis functions. The main boundary conditions to be considered are those for the behavior of the wave function far from the nuclei and those for the behavior close to the nuclei (in addition to requirements of continuity and differentiability).

In nonrelativistic quantum chemistry, the basis functions used are usually atom centered, because the bulk of the energy is obtained from the regions near the atomic nuclei, and a poor description of the atom would seriously affect the accuracy of the molecular calculations. Thus, atom-centered basis sets must provide a good description of the atom, which means satisfying in some degree the boundary conditions of the atom. For the simplest case of a hydrogen atom, the wave function has a cusp at the nucleus due to the use of a point nuclear model and decays exponentially at large distances from the nucleus.

Atom-centered Gaussian basis functions are used in the vast majority of calculations on molecular systems, with great success. Although these functions do not exactly satisfy the boundary conditions that the atomic wave function satisfies – they decay faster at large distances and do not have a cusp at the nucleus – their properties are similar enough to allow a reasonably accurate description of the atom. What is more, it is possible to systematically expand a Gaussian basis set so that it becomes complete, for the purpose of describing an atomic function [1]. The simplicity of these functions and the evaluation of their integrals compensate for their deficiencies, and one can easily add more functions to better describe the nuclear region or the long-range region. The values of the wave function itself are not usually interesting, but the values of properties are, as they are related to observables. Properties are expectation values of some operator, and small errors in the wave function do not usually contribute significantly to the property. At any level of calculation, then, a basis set is chosen that produces property values of a desired accuracy.

To pursue the same sort of approach in relativistic quantum chemistry, several extra conditions are necessary, as the calculations are based on the Dirac Hamiltonian rather than the Schrödinger Hamiltonian. The use of the Dirac Hamiltonian has several consequences that affect basis set choice:

- The spectrum of the Dirac equation has both a positive-energy and a negative-energy branch, and the energy includes the rest mass of the particle, in this case the electron. The negative-energy states are related to the antiparticle (positron) wave functions. This adds an extra degree of freedom to the wave function.
- As the Dirac Hamiltonian is spin dependent, the spin can no longer be separated from the spatial coordinates. This results in coupling of spin and orbital angular momenta in the atoms, and only the total angular momentum is a good quantum number. In molecules this means that the double group must be used. With spin

dependence and the particle-antiparticle behavior, the Dirac wave function has four components to represent the extra degrees of freedom. (The spin is of course included implicitly in nonrelativistic wave functions.)
- For a point nucleus, the solutions of the Dirac equation for the ground state of a hydrogenic atom has a singularity at the nucleus, rather than a cusp. This is related to the relativistic contraction of the wave function. More generally, the wave function contains a non-integer power of r, the radial coordinate. It should be pointed out that the point nuclear model is unphysical: nuclei are extended objects with size and structure.
- Magnetic effects involve the coupling between the positive-energy and negative-energy degrees of freedom (the so-called "large" and "small" components).

The implications of these issues for basis set development and use are discussed in the sections below.

It is also possible to use approximations to the full four-component approach, by decoupling the positive-energy and negative-energy states to some degree or by treating the spin-dependent effects as a perturbation based on a spin-free approximate Hamiltonian. The treatment below also includes comments appropriate to such approximations. As the fundamental issues are those that arise out of the four-component approach, the primary discussion centers on this approach.

In the developments below, SI-based Hartree atomic units are generally used, $\hbar = e = m = 4\pi\varepsilon_0 = 1$, but e and m are kept where it is useful, or SI units are used.

Kinetic Balance

To examine the requirements that are imposed on the basis set due to the existence of the negative-energy states, the Dirac wave function is written in terms of its components. However, as the spin dependence will not be explicitly examined at this point, the wave function is represented in terms of two-component functions (or *2-spinors*), each of which has the spin part coupled to the spatial part in the appropriate manner. The wave function in terms of these 2-spinors is

$$\psi = \begin{pmatrix} \psi^L \\ \psi^S \end{pmatrix} \tag{1}$$

where ψ^L is the 2-spinor for the large component and ψ^S is the 2-spinor for the small component [1]. These components are expanded in terms of a basis set of 2-spinor functions,

[1]This nomenclature is appropriate for positive-energy or electron states, which we are mostly concerned with here, where the large-component integrated density is larger than that of the small component. Other nomenclature in use for these components is "upper" and "lower" components, which refer to their placement in the standard representation of the Dirac wave function.

$$\psi^L = \sum_\mu a_\mu \chi^L_\mu; \qquad \psi^S = \sum_\mu b_\mu \chi^S_\mu \tag{2}$$

and the conditions that these basis functions must satisfy are then determined.

The Dirac equation is written in two-component form with the rest mass subtracted [2] as two coupled equations,

$$(V - E)\psi^L + c\boldsymbol{\sigma} \cdot \mathbf{p}\psi^S = 0 \tag{3}$$

$$c\boldsymbol{\sigma} \cdot \mathbf{p}\psi^L + (V - 2mc^2 - E)\psi^S = 0, \tag{4}$$

where $\boldsymbol{\sigma}$ is the vector of Pauli matrices. The second equation is rearranged to make ψ^S the subject,

$$\psi^S = (2mc^2 + E - V)^{-1} c\boldsymbol{\sigma} \cdot \mathbf{p}\psi^L \tag{5}$$

This equation defines the relationship between the large component and the small component. As the denominator in this expression can go to zero if $E < -2mc^2$, depending somewhat on V, one should be cautious in applying it for the negative-energy states. But as the positive-energy states have a positive value for the denominator (for most potentials in applications of interest to chemistry and physics), this should not be a problem for determining properties relating to electronic states.

The next step is to eliminate the small component from the first line,

$$(V - E)\psi^L + c\boldsymbol{\sigma} \cdot \mathbf{p}(2mc^2 + E - V)^{-1} c\boldsymbol{\sigma} \cdot \mathbf{p}\psi^L = 0. \tag{6}$$

This equation should reduce to the nonrelativistic equation in the limit $c \to \infty$. Taking this limit, the result is

$$(V - E)\psi^L + \frac{1}{2m}\boldsymbol{\sigma} \cdot \mathbf{p}\boldsymbol{\sigma} \cdot \mathbf{p}\psi^L = 0 \tag{7}$$

and it can be seen that the nonrelativistic kinetic energy operator T is represented by the operator $(\boldsymbol{\sigma} \cdot \mathbf{p})^2/2m$. From the properties of the Pauli matrices, $(\boldsymbol{\sigma} \cdot \mathbf{p})^2 = \mathbf{p}^2$ which gives the desired result.

When the basis set expansion is substituted in to develop a matrix version of the Dirac equation, the result is

$$\begin{aligned}(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\mathbf{a} + c\boldsymbol{\Pi}^{LS}\mathbf{b} &= 0 \\ c\boldsymbol{\Pi}^{SL}\mathbf{a} + (\mathbf{V}^{SS} - (2mc^2 + E)\mathbf{S}^{SS})\mathbf{b} &= 0,\end{aligned} \tag{8}$$

[2] On this energy scale, the "positive-energy" bound states all have $E < 0$.

where the matrices are

$$V_{\mu\nu}^{LL} = \langle \chi_\mu^L | V | \chi_\nu^L \rangle, \tag{9}$$

$$V_{\mu\nu}^{SS} = \langle \chi_\mu^S | V | \chi_\nu^S \rangle, \tag{10}$$

$$S_{\mu\nu}^{LL} = \langle \chi_\mu^L | \chi_\nu^L \rangle, \tag{11}$$

$$S_{\mu\nu}^{SS} = \langle \chi_\mu^S | \chi_\nu^S \rangle, \tag{12}$$

$$\Pi_{\mu\nu}^{LS} = \langle \chi_\mu^L | \boldsymbol{\sigma} \cdot \mathbf{p} | \chi_\nu^S \rangle = (\Pi_{\nu\mu}^{SL})^\dagger. \tag{13}$$

Eliminating **b** from the first equation using the second, and taking the nonrelativistic limit,

$$(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\mathbf{a}^L + \frac{1}{2m}\boldsymbol{\Pi}^{LS}(\mathbf{S}^{SS})^{-1}\boldsymbol{\Pi}^{SL}\mathbf{a}^L = 0. \tag{14}$$

The nonrelativistic kinetic energy matrix $\mathbf{T}^{LL}$ is represented in Eq. (14) by a matrix product. The presence of the inverse of the small-component overlap indicates that a projection operator has been inserted into the product of the $\boldsymbol{\sigma}\cdot\mathbf{p}$ operators in Eq. (7). If a poor choice of the small-component basis set is made, a poor approximation to the kinetic energy from the matrix product is the likely result.

The critical piece is the matrix of the operator $\boldsymbol{\sigma} \cdot \mathbf{p}$. The basis functions χ_μ^S must be chosen so that they span the range of $\boldsymbol{\sigma} \cdot \mathbf{p}\chi_\mu^L$; if not, the kinetic energy is underestimated, and the energies will be too low [2–5]. The exact result for the nonrelativistic kinetic energy is guaranteed by making the choice

$$\chi_\mu^S = \boldsymbol{\sigma} \cdot \mathbf{p}\chi_\mu^L. \tag{15}$$

With this choice,

$$\Pi_{\mu\nu}^{LS} = \langle \chi_\mu^L | (\boldsymbol{\sigma} \cdot \mathbf{p})^2 | \chi_\nu^L \rangle = 2mT_{\mu\nu}^{LL} \tag{16}$$

$$S_{\mu\nu}^{SS} = \langle \chi_\mu^L | (\boldsymbol{\sigma} \cdot \mathbf{p})^2 | \chi_\nu^L \rangle = 2mT_{\mu\nu}^{LL}. \tag{17}$$

and therefore

$$\boldsymbol{\Pi}^{LS}(\mathbf{S}^{SS})^{-1} = \mathbf{I} \tag{18}$$

and

$$\frac{1}{2m}\boldsymbol{\Pi}^{SL} = \mathbf{T}^{LL}. \tag{19}$$

so that the correct nonrelativistic limit is obtained. The condition in Eq. (15) is known as *kinetic balance* [5]. Scaling the small component by a constant (that may depend on μ) makes no difference, as the constants cancel in Eq. (14).

Some convenient scaling constants are $1/2mc$, which makes $c\boldsymbol{\Pi}^{LS} = \mathbf{T}^{LL}$, and a normalization constant, to make $S^{SS}_{\mu\mu} = 1$. If a scaling of $1/2mc$ is chosen, for example, then Eq. (8) becomes

$$\begin{aligned}\mathbf{V}^{LL}\mathbf{a} + \mathbf{T}^{LL}\mathbf{b} &= E\mathbf{S}^{LL} \\ \mathbf{T}^{LL}\mathbf{a} + (\mathbf{W}^{LL}/4m^2c^2 - \mathbf{T}^{LL})\mathbf{b} &= E(\mathbf{T}^{LL}/2mc^2)\mathbf{b}\end{aligned} \tag{20}$$

where

$$W^{LL}_{\mu\nu} = \langle\chi^L_\mu|\boldsymbol{\sigma}\cdot\mathbf{p}V\boldsymbol{\sigma}\cdot\mathbf{p}|\chi^L_\nu\rangle. \tag{21}$$

In this formulation, it is only necessary to define the large-component basis set, as kinetic balance is folded into the integrals. These are the usual nonrelativistic integrals (though evaluated over 2-spinor functions), except for **W**, which is a relativistic integral. Further relativistic integrals arise from the electron-electron interaction. In essence, the kinetic balance condition is used to transform the operators.

Kinetic balance is an approximation: it does not guarantee that the full relation between the large-component and the small-component (Eq. 5) wave function is satisfied. However, it does ensure that the correct symmetry relationships between the components are satisfied. The operator $(2mc^2 + E - V)$ is totally symmetric, in any point group of a molecule. Its neglect in the kinetic balance relation therefore does not matter, from the point of view of symmetry. What is important is that kinetic balance is applied to the definition of the basis set, not to the wave function itself. The relation between the large and small components of the wave function is then carried by the coefficients **a** and **b**.

The use of kinetic balance means that, for the spatial functions used in the 2-spinor basis, the small-component functions transform as the derivatives of the large-component functions, due to the appearance of the momentum operator. This means that they have the opposite parity under inversion to the large-component functions: where the large component is even, the small component is odd, and vice versa.

The kinetic balance requirement also helps to explain why relativistic four-component calculations are more expensive than nonrelativistic calculations. With no approximations to the integrals in the four-component calculation, it is necessary to calculate derivatives of the usual nonrelativistic two-electron integrals up to fourth derivatives. And as the derivatives are present regardless of the spin, the same degree of differentiation is required if the spin-dependent operators are removed: the derivatives appear in the scalar relativistic integrals.

Kinetic balance as outlined above only applies to the positive-energy, or electron, states. The small component was eliminated and the large component was kept. The large component becomes the nonrelativistic electronic wave function in the limit $c \to \infty$. If the large component is eliminated, and instead of subtracting the rest mass, it is added, equations are obtained for the negative-energy states:

$$\psi^L = -(V - E + 2mc^2)^{-1} c\boldsymbol{\sigma} \cdot \mathbf{p}\psi^S, \tag{22}$$

which on substitution to eliminate the large component gives

$$(V - E)\psi^S - c\boldsymbol{\sigma} \cdot \mathbf{p}(V - E + 2mc^2)^{-1} c\boldsymbol{\sigma} \cdot \mathbf{p}\psi^S = 0. \tag{23}$$

The kinetic balance condition for the negative-energy states, which is needed to give an accurate representation of the kinetic energy in the nonrelativistic limit for positrons, is

$$\chi_\mu^L = \boldsymbol{\sigma} \cdot \mathbf{p}\chi_\mu^S. \tag{24}$$

When the Dirac Hamiltonian matrix is diagonalized, solutions for both the positive-energy and negative-energy states are obtained. So the basis used applies to both sets of states. If kinetic balance is to apply for both electrons and positrons, then both Eqs. (15) and (24) must apply, and substituting one into the other yields

$$\chi_\mu^L = (\boldsymbol{\sigma} \cdot \mathbf{p})^2 \chi_\mu^L = \mathrm{p}^2 \chi_\mu^L \tag{25}$$

This equation is satisfied by plane waves (allowing for a scaling factor in the equations), but not by the usual basis sets of quantum chemistry. It is possible to use *dual kinetic balance* [6, 7], in which four-component basis functions are used, and the basis includes functions that have kinetic balance for electrons and for positrons.

The kinetic balance relations imply a one-to-one correspondence of the large-component basis set and the small-component basis set. In fact, the accurate reproduction of the nonrelativistic kinetic energy for electrons only requires that the small-component set contains the kinetically balanced functions,

$$\{\chi_\mu^S\} \supseteq \{\boldsymbol{\sigma} \cdot \mathbf{p}\chi_\mu^L\}. \tag{26}$$

If functions are added to the small-component basis set, these functions produce extra negative-energy states. But as there are not enough large-component functions, the extra states will have zero kinetic energy and essentially be eigenfunctions of the potential energy. (In actuality, the extra negative-energy states will be mixed in with the "real" negative-energy states, but they can always be rotated to produce a set of states that have no large component and a set of states that do have a large component.) These states are spurious, nonphysical states. If the negative-energy states are not of interest, this might not matter much, as the description of the positive-energy states will still be accurate. On the other hand, if there are more large-component basis functions than small-component basis functions, spurious positive-energy states would be obtained, which could adversely affect the calculation of electronic states. As they are eigenfunctions of the potential energy, they could be lower in energy than the genuine electronic states and cause some kind of variational collapse. So maintaining a 1:1 correspondence of the large- and small-component basis sets is generally to be preferred.

Spin Coupling

The solutions of the Dirac equation are spin coupled, as a consequence of the inclusion of spin operators in the Dirac equation – the Dirac matrices, which are represented in terms of the Pauli matrices. The spin is represented here as a two-component vector, with the first row representing alpha spin and the second row representing beta spin. For an atom, the coupling means that the solutions are eigenfunctions of the total angular momentum, rather than being eigenfunctions of the spin and orbital angular momenta separately.

In the previous section the wave function was expanded in two-component spinors. These are now examined more closely, in the context of an atom, as atom-centered basis functions are desirable in relativistic calculations. The Dirac equation for an atom separates into a radial equation and a spin-orbital equation, and hence the solutions can also be factorized into radial and spin-orbital functions. So for an atom the one-particle wave functions, or 4-spinors, can be written as

$$\psi = \frac{1}{r}\begin{pmatrix} P(r)\,\xi^L(\theta,\phi,\tau) \\ iQ(r)\,\xi^S(\theta,\phi,\tau) \end{pmatrix} \tag{27}$$

where r, θ, and ϕ are the usual spherical coordinates and τ is the spin coordinate (which represents m_s). The wave function is written in terms of a large component and a small component, with $P(r)$ the *radial* large component and $Q(r)$ the *radial* small component. The spin-angular functions ξ are coupled products of spin functions and spherical harmonics,

$$\xi_{\ell jm} = \sum_{m_\ell, m_s} \langle \ell m_\ell s m_s | jm \rangle \; Y_{\ell m_\ell}(\theta,\phi)\; \eta(m_s) \tag{28}$$

where $\langle \ell m_\ell s m_s | jm \rangle$ is a Clebsch-Gordan coefficient, $Y_{\ell m_\ell}(\theta,\phi)$ is a spherical harmonic (with the usual Condon and Shortley phase conventions), and $\eta(m_s)$ is a spin function. If the spin functions are written as two component vectors,

$$\eta(\alpha) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \qquad \eta(\beta) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{29}$$

then the spin-angular functions can be written as follows,

$$\xi_{\ell jm} = \frac{1}{\sqrt{2\ell+1}}\begin{pmatrix} \sqrt{\ell + m_\ell + 1}\; Y_{\ell m_\ell} \\ \\ \sqrt{\ell - m_\ell}\; Y_{\ell m_\ell + 1} \end{pmatrix}, \qquad j = \ell + 1/2, \tag{30}$$

$$\xi_{\ell jm} = \frac{1}{\sqrt{2\ell+1}}\begin{pmatrix} -\sqrt{\ell - m_\ell + 1}\; Y_{\ell m_\ell} \\ \\ \sqrt{\ell + m_\ell + 1}\; Y_{\ell m_\ell + 1} \end{pmatrix}, \qquad j = \ell - 1/2, \tag{31}$$

Table 1 Quantum numbers for large and small components

Large				Small			
Label	ℓ	j	κ	Label	ℓ	j	κ
$s_{1/2}$	0	1/2	-1	$p_{1/2}$	1	1/2	$+1$
$p_{1/2}$	1	1/2	$+1$	$s_{1/2}$	0	1/2	-1
$p_{3/2}$	1	3/2	-2	$d_{3/2}$	2	3/2	$+2$
$d_{3/2}$	2	3/2	$+2$	$p_{3/2}$	1	3/2	-2
$d_{5/2}$	2	5/2	-3	$f_{5/2}$	3	5/2	$+3$

where the Clebsch-Gordan coefficients have been expanded.

The spin-angular functions are eigenfunctions of a spin-orbit-type operator, whose eigenvalues $\kappa = (\ell - j)(2j + 1)$ can be used to characterize the functions instead of ℓ and j. The large component and the small component have opposite signs of κ, so they can be labeled as follows:

$$\psi^L = \frac{1}{r} P(r)\, \xi_{\kappa m}(\theta, \phi, \tau); \qquad \psi^S = \frac{i}{r} Q(r)\, \xi_{-\kappa m}(\theta, \phi, \tau). \tag{32}$$

As the large and small components are connected by the operator $\boldsymbol{\sigma} \cdot \mathbf{p}$, the spherical harmonics must differ by one unit of angular momentum. If κ is negative for the large component, the small component must have one more unit of angular momentum than the large component, whereas if κ is positive for the large component, the small component must have one less unit of angular momentum than the large component. So for example, for a large-component s function, which has $j = 1/2$ and $\kappa = -1$, the small component must be a p function with $j = 1/2$ and $\kappa = +1$. For a large-component p function, there are two possible values of j, corresponding to the spin-orbit components. If $j = 1/2$ for the large component, $\kappa = +1$, and the small component is an s function with $\kappa = -1$, whereas for $j = 3/2$ for the large component, $\kappa = -2$, and the small component is a d function with $\kappa = +2$. The relation between the components and the quantum numbers for several shells is given in Table 1.

The two-component functions that have been considered so far have the spin coupled to the spatial part, to represent the eigenfunctions of the Dirac Hamiltonian. The spin can be decoupled in the basis set, and each of the four components can be expanded independently. Or, equivalently, the large and small components can be expanded in terms of a set of spin orbitals, rather than a set of coupled 2-spinors:

$$\psi^L = \sum_\mu \left(a_{\mu\alpha}\, \bar{\chi}^L_\mu\, \eta(\alpha) + a_{\mu\beta}\, \bar{\chi}^L_\mu\, \eta(\beta) \right) \tag{33}$$

$$\psi^S = \sum_\mu \left(b_{\mu\alpha}\, \bar{\chi}^S_\mu\, \eta(\alpha) + b_{\mu\beta}\, \bar{\chi}^S_\mu\, \eta(\beta) \right) \tag{34}$$

where the basis functions $\bar{\chi}^{L,S}_\mu$ are now pure spatial functions. Note that the same spatial functions are used for both alpha and beta spin: what differs in the expansion

is the spin-dependent coefficient. This choice of basis is often called a scalar basis, because each component of the wave function is expanded in a basis of scalar functions.

The matrix Dirac equation now has to be represented in terms of integrals over the scalar basis functions.

$$\begin{pmatrix} \bar{\mathbf{V}}^{LL} - E\bar{\mathbf{S}}^{LL} & 0 & c\bar{\boldsymbol{\Pi}}_z^{LS} & c\bar{\boldsymbol{\Pi}}_-^{LS} \\ 0 & \bar{\mathbf{V}}^{LL} - E\bar{\mathbf{S}}^{LL} & c\bar{\boldsymbol{\Pi}}_+^{LS} & -c\bar{\boldsymbol{\Pi}}_z^{LS} \\ c\bar{\boldsymbol{\Pi}}_z^{SL} & c\bar{\boldsymbol{\Pi}}^{SL-} & \bar{\mathbf{V}}^{SS} - E\bar{\mathbf{S}}^{SS} & 0 \\ c\bar{\boldsymbol{\Pi}}_+^{SL} & c\bar{\boldsymbol{\Pi}}_{-z}^{SL} & 0 & \bar{\mathbf{V}}^{SS} - E\bar{\mathbf{S}}^{SS} \end{pmatrix} \begin{pmatrix} \mathbf{a}^\alpha \\ \mathbf{a}^\beta \\ \mathbf{b}^\alpha \\ \mathbf{b}^\beta \end{pmatrix} = 0. \tag{35}$$

where the matrix elements are

$$\bar{V}_{\mu\nu}^{LL} = \langle \bar{\chi}_\mu^L | V | \bar{\chi}_\nu^L \rangle, \tag{36}$$

$$\bar{V}_{\mu\nu}^{SS} = \langle \bar{\chi}_\mu^S | V | \bar{\chi}_\nu^S \rangle, \tag{37}$$

$$\bar{S}_{\mu\nu}^{LL} = \langle \bar{\chi}_\mu^L | \bar{\chi}_\nu^L \rangle, \tag{38}$$

$$\bar{S}_{\mu\nu}^{SS} = \langle \bar{\chi}_\mu^S | \bar{\chi}_\nu^S \rangle, \tag{39}$$

$$(\bar{\Pi}_z^{LS})_{\mu\nu} = \langle \bar{\chi}_\mu^L | p_z | \bar{\chi}_\nu^S \rangle \tag{40}$$

$$(\bar{\Pi}_\pm^{LS})_{\mu\nu} = \langle \bar{\chi}_\mu^L | p_x \pm i p_y | \bar{\chi}_\nu^S \rangle. \tag{41}$$

It might be thought that the rank of the Dirac matrix in the scalar basis is twice that of the Dirac matrix in the 2-spinor basis, but this is not necessarily the case. The number of 2-spinors is the same as the number of spin orbitals for any given ℓ value. In the 2-spinor case, the Dirac matrix for an atom can be placed in block-diagonal form, as both j and m are good quantum numbers. In the scalar basis this is not the case, as neither ℓ nor s are good quantum numbers.

The sets of large- and small-component functions in the scalar basis must still satisfy Eq. (26). Applying the kinetic balance relation Eq. (15) directly to the large-component spin-orbital functions gives the following result:

$$\boldsymbol{\sigma} \cdot \mathbf{p}\, \bar{\chi}_\mu^L\, \eta(\alpha) = -i\hbar \left(\frac{\mathrm{d}}{\mathrm{d}x} \bar{\chi}_\mu^L\, \eta(\beta) + i \frac{\mathrm{d}}{\mathrm{d}y} \bar{\chi}_\mu^L\, \eta(\beta) + \frac{\mathrm{d}}{\mathrm{d}z} \bar{\chi}_\mu^L\, \eta(\alpha) \right), \tag{42}$$

$$\boldsymbol{\sigma} \cdot \mathbf{p}\, \bar{\chi}_\mu^L\, \eta(\beta) = -i\hbar \left(\frac{\mathrm{d}}{\mathrm{d}x} \bar{\chi}_\mu^L\, \eta(\alpha) + -i \frac{\mathrm{d}}{\mathrm{d}y} \bar{\chi}_\mu^L\, \eta(\alpha) - \frac{\mathrm{d}}{\mathrm{d}z} \bar{\chi}_\mu^L\, \eta(\beta) \right). \tag{43}$$

It can be seen that using a spin-orbital basis for the large component still leads to spin-coupled functions in the small component, when the kinetic balance relation is strictly applied, and not to pure spin orbitals. To obtain pure spin orbitals for

the small component, extra functions must be introduced into the small-component basis that do not arise from the kinetic balance relation, such as

$$-i\hbar\left(\frac{\mathrm{d}}{\mathrm{d}x}\bar{\chi}_{\mu}^{L}\,\eta(\beta)-i\frac{\mathrm{d}}{\mathrm{d}y}\bar{\chi}_{\mu}^{L}\,\eta(\beta)\right), \tag{44}$$

$$-i\hbar\left(\frac{\mathrm{d}}{\mathrm{d}x}\bar{\chi}_{\mu}^{L}\,\eta(\beta)-i\frac{\mathrm{d}}{\mathrm{d}y}\bar{\chi}_{\mu}^{L}\,\eta(\beta)\right), \tag{45}$$

which permits the elimination of either the x derivative or the y derivative, and

$$-i\hbar\left(\frac{\mathrm{d}}{\mathrm{d}x}\bar{\chi}_{\mu}^{L}\,\eta(\beta)+i\frac{\mathrm{d}}{\mathrm{d}y}\bar{\chi}_{\mu}^{L}\,\eta(\beta)-2\frac{\mathrm{d}}{\mathrm{d}z}\bar{\chi}_{\mu}^{L}\,\eta(\alpha)\right), \tag{46}$$

$$-i\hbar\left(\frac{\mathrm{d}}{\mathrm{d}x}\bar{\chi}_{\mu}^{L}\,\eta(\alpha)+-i\frac{\mathrm{d}}{\mathrm{d}y}\bar{\chi}_{\mu}^{L}\,\eta(\alpha)+2\frac{\mathrm{d}}{\mathrm{d}z}\bar{\chi}_{\mu}^{L}\,\eta(\beta)\right), \tag{47}$$

which permits the elimination of the z component and thus completes the decoupling into spin orbitals. This formal demonstration of the need to add functions to obtain a spin-orbital basis shows that there is no longer a 1:1 correspondence between the large-component and small-component functions and that the small-component basis is larger than the large-component basis.[3]

In practice, the small-component basis is obtained by simply differentiating the spatial functions and taking combinations with the spin functions to form spin-orbital functions. Kinetic balance in the context of a scalar basis set usually means simple differentiation of the spatial functions, and this is termed *scalar kinetic balance* in this article.

Using scalar kinetic balance on the large-component scalar basis set to generate the small-component scalar basis set always generates more functions in the small-component set than in the large-component set. This means that use of a scalar basis set throughout the calculations will always result in extra negative-energy states, as discussed in the previous section. In a 2-spinor basis set, applying kinetic balance strictly to the large-component 2-spinors will result in the same number of small-component basis functions as large-component basis functions.

To illustrate, consider the ground state of the hydrogen atom. In a 2-spinor basis, applying kinetic balance to the $s_{1/2}$ large-component basis functions generates only a set of $p_{1/2}$ basis functions. The $p_{3/2}$ basis functions are not present. In a scalar basis, applying scalar kinetic balance (as a simple differentiation) to the s functions of the large component generates p functions for the small component, and there are

[3]It would of course be possible to use the mixed basis, with spin orbitals for the large component and the spin-coupled functions derived here for the small component, and this is done implicitly when $\boldsymbol{\sigma}\cdot\mathbf{p}$ is incorporated into the operators instead of the small-component basis. In practice, it is simpler to use spin orbitals, but see below for issues related to doing this.

three times as many functions in the small-component set as in the large-component set. Combined with the spin functions, the small-component spin-orbital basis spans both the $p_{1/2}$ and the $p_{3/2}$ spinor space. The $p_{3/2}$ functions are the ones that must be added for spin separation of the small-component basis into spin orbitals. Solving the Dirac equation in this basis for the hydrogen atom will give the correct solutions, just as in the 2-spinor basis, but it will also give a set of unphysical negative-energy states with only a $p_{3/2}$ small component and no large component.

While for an atom, the extra states can be separated out by symmetry, this is not the case in molecules. In a water molecule, for example, there is only one irreducible representation (irrep) in the double group, so all spinor functions must belong to this irrep, and the $p_{1/2}$ functions cannot be easily separated from the $p_{3/2}$ functions.

There is another issue with the use of a scalar basis. The application of kinetic balance to large-component s functions and to large-component d functions produces a set of small-component p functions in both cases. In the 2-spinor basis, this is not a problem, as the p functions generated from the s functions are $p_{1/2}$ functions, and the p functions generated from the d functions are $p_{3/2}$ functions, and these functions are orthogonal due to their angular symmetry. In the scalar basis, there is no spin coupling to eliminate the extra functions: the small-component p functions generated from both the large-component s and d functions span the same space. This may result in linear dependence of the small-component basis, which will have to be taken care of numerically. Other solutions to alleviate this problem are discussed below, when the use of Gaussian basis functions is considered.

From a practical point of view, however, there are advantages to working in the scalar basis. The 2-spinor basis is a complex, spin-coupled basis, and the one-electron and two-electron integrals would have to be calculated in or transformed into this basis. This would require development of integral codes designed to work with 2-spinors or code that transformed the scalar integrals into 2-spinor integrals. The scalar basis can be represented in terms of real, spin-independent functions (as the spherical harmonics can be represented in real form rather than complex). The integrals can then be taken from standard nonrelativistic integral codes, with some adaptions for the small component, such as the use of derivative technology.

To address the problem of the extra negative-energy states, it is possible to transform between the scalar and the 2-spinor basis. This means that for the critical parts of the calculation where linear dependence matters, such as the diagonalization of the Fock matrix, the 2-spinor basis is used, but in other parts of the calculation, such as accumulation of the Fock matrix, the scalar basis is used.

Radial Behavior

By far the majority of nonrelativistic molecular calculations use Gaussian basis sets, due to the ease of integral evaluation. Of the two deficiencies mentioned above – the behavior at the nucleus and the behavior at large distances – the

behavior at the nucleus is more of an issue in relativistic calculations than in nonrelativistic calculations. The reason is that the Dirac wave function for an s orbital at the nucleus has a singularity rather than a cusp, if the nucleus is treated as a point charge. Although it is still possible to use Gaussian functions for a point nucleus in relativistic calculations, the size of the exponents required to approximate the singularity increases significantly as the atomic number increases. For heavy elements, the largest exponent can be orders of magnitude larger than the corresponding nonrelativistic exponents [8,9].

However, as pointed out above, real nuclei are not point charges, but instead have a spatial distribution. Most four-component relativistic calculations use a finite nuclear charge distribution, which provides both a more physical picture and avoids the mathematical difficulties of the unphysical point nucleus. The nucleus is usually represented by a single Gaussian function [10]. With any finite-sized nuclear model, the wave function near the nucleus for the large component or the small component behaves as r^{ℓ}. For a Gaussian charge distribution, the wave function is an even or odd function of r. Gaussian functions are therefore well suited to describing the wave function near the nucleus with this model, and in addition, the nuclear attraction integrals can be evaluated with the same methods as for the two-electron integrals, as they reduce to integrals of the type $(\mathrm{ss}|ab)$.

A further benefit is gained because the nuclear size increases with atomic number Z (or, rather, atomic mass), and therefore the largest exponent needed in the basis set starts to decrease at some point. For example, in the basis sets of Dyall [11–20], the largest exponent in the quadruple-zeta basis set increases with Z to a maximum of about 10^8 in the 3d block and thereafter slowly decreases as the effect of the nuclear size becomes more significant. In the triple-zeta basis sets, the maximum is in the 4d block, whereas in the double-zeta basis sets, the maximum is in the 5f block. The smaller basis sets do not sample the region near the nucleus as much as the larger basis sets, so the effect of the nuclear size is not felt until higher Z. The use of a finite nuclear size limits the size of the basis set for the heavy elements.

The range of exponents needed for relativistic calculations differs from that in nonrelativistic calculations due to the change in the radial extent of the relativistic wave function. In one-electron systems, the wave function contracts due to the effects of relativity. In many-electron systems, the behavior of the wave function is more complex. The direct effect of relativity in contracting the wave function has two consequences: it changes the orthogonality requirements of shells of the same symmetry, and it produces increased screening of the nucleus. The indirect effect of the increase in screening can result in an expansion of the screened outer shells, which varies with angular momentum [21]. For the s shells, all of which penetrate to the nucleus, the direct effect dominates, and all s shells have a smaller radius than their nonrelativistic counterparts. For the p shells, the direct and indirect effects balance to some extent in the valence region, so that the valence p shells differ very little in extent from their nonrelativistic counterparts. The spin-orbit effect changes this balance, as the $\mathrm{p}_{1/2}$ spinors have an s-like small component, and the direct effect is larger. For the outer d and f shells, the indirect effects dominate, and these shells expand.

The most important effect for basis set design is the contraction of the inner shells, which means that more tight functions are required to describe the wave function. This requirement is primarily for the s shells but is important also for the p shells. The largest exponent needed for the p shell actually increases faster as a function of Z than for the s shell, so that by the 5f block, the largest s exponent and the largest p exponent have similar magnitudes. This is not the case for the nonrelativistic basis sets, where the largest p exponent is considerably smaller than the largest s exponent. As an example, the use of nonrelativistic basis sets in relativistic calculations on the 6p elements incurs a considerable error in the core, of several hartrees, and two additional tight p functions are needed to reduce the error to 0.5 hartrees [22].

In the outer region, the changes are smaller: contraction of the s functions may mean that fewer diffuse functions are needed in the s space, and expansion of the d and f space may require more diffuse functions. Usually these changes are limited to only one function.

Spin-orbit splitting also affects the radial behavior of the wave function, and this is particularly important for the p shells, where it is the largest. For nonrelativistic calculations, sets of functions are usually developed for each ℓ value. In relativistic calculations, the spin-orbit components of a shell have different radial extents, and the difference increases with Z. The demands on the basis for the $p_{1/2}$ shell are greater than for the $p_{3/2}$ shell, so much so that by the 7p block, the largest exponents are more than an order of magnitude different. In fact, the observation on the largest p exponent above is largely, though not entirely, due to the spin-orbit splitting. The effect of spin-orbit splitting is not insignificant in the valence region, either, where the smallest exponents differ by a factor of 2 or 3.

In the design of basis sets, some choices must be made about how to handle the differences in requirements between the spin-orbit components. For efficient integral evaluation, it would be better to have the same exponents for both spin-orbit components. This means that the basis set is larger than needed for either of the components separately. With this approach, the p sets become as large as the s sets for the 7p elements and are not much smaller for the 6p elements [14, 20].

It would be a mistake to suppose that the demands of spin-orbit splitting are only relevant for the heavy elements. Even for the F atom, the spin-orbit splitting of the 2p shell is underestimated by a significant fraction when using the nonrelativistic cc-pVDZ and cc-pVTZ basis sets of Dunning, and an extra tight p function was found necessary to be added to improve the results [23].

For higher angular momenta, the spin-orbit splitting is smaller, and so the effect on the exponent range is smaller than for the p shell.

Primitive Gaussian Basis Sets

Having discussed the general effects of relativity on basis set requirements, the use of Gaussian basis sets is now addressed. Gaussian functions can be written in general as

$$\chi_\mu(\boldsymbol{r}) = N_\mu f(\boldsymbol{r})\, e^{-\zeta_\mu r^2} \tag{48}$$

where N_μ is a normalization factor. For spherical Gaussians, the function $f(\boldsymbol{r})$ is usually chosen as a solid spherical harmonic,

$$f(\boldsymbol{r}) = r^\ell\, Y_{\ell m}(\theta, \phi) \tag{49}$$

where $Y_{\ell m}(\theta, \phi)$ is the complex spherical harmonic introduced above for 2-spinor basis sets. However, for scalar basis sets it is often more convenient to use real spherical harmonics so that the basis functions are entirely real. For Cartesian Gaussians, the function $f(\boldsymbol{r})$ is a Cartesian tensor of a given rank,

$$f(\boldsymbol{r}) = x^i y^j z^k, \quad i + j + k = \ell. \tag{50}$$

How these are used in calculations depends on the type of calculation: scalar relativistic, 2-spinor, or 4-spinor. The discussion starts with the 4-spinor case, for which there are the most constraints on the basis.

To work in a 2-spinor basis, the function $f(\boldsymbol{r})$ must written in terms of the spin-angular functions of Eq. (28),

$$f(\boldsymbol{r}) = r^\ell\, \xi_{\kappa m}(\theta, \phi, \tau) \tag{51}$$

Using these 2-spinor basis functions for the large component, the basis functions for the small component must be determined by kinetic balance. The action of the angular part of $\boldsymbol{\sigma} \cdot \mathbf{p}$ on the spin-angular function changes the sign of kappa, so it remains to determine the radial function. Applying the radial part of $\boldsymbol{\sigma} \cdot \mathbf{p}$,

$$\left[\frac{\mathrm{d}}{\mathrm{d}r} + \frac{\kappa + 1}{r}\right] r^\ell e^{-\zeta r^2} = \left[\frac{\ell + \kappa + 1}{r} - 2\zeta r\right] r^\ell e^{-\zeta r^2}. \tag{52}$$

The value of κ is $-\ell - 1$ for $j = \ell + 1/2$, so in this case the first term vanishes, and the radial function is a simple product of a power of r and a Gaussian. But for $j = \ell - 1/2$, the value of κ is ℓ, and the first term does not vanish: there is now a sum of powers of r in the radial function.

As an example, consider a 2p-type Gaussian for the large component, $re^{-\zeta r^2}$. For the small component, the radial function is

$$\begin{aligned} \left[\tfrac{\mathrm{d}}{\mathrm{d}r} + \tfrac{\kappa+1}{r}\right] re^{-\zeta r^2} &= -2\zeta r^2 e^{-\zeta r^2}, && j = 3/2 \\ &= (2\ell + 1)e^{-\zeta r^2} - 2\zeta r^2 e^{-\zeta r^2}, && j = 1/2 \end{aligned} \tag{53}$$

From the discussion above, the small component for a $\mathrm{p}_{3/2}$ function is expected to be a $\mathrm{d}_{3/2}$ function, and in this case the radial function is that of a 3d-type Gaussian. For $j = 1/2$, the small-component function is expected to be an $\mathrm{s}_{1/2}$ function, and in this case it is a linear combination of a 1s-type Gaussian and a 3s-type Gaussian.

This small-component function could be split into two basis functions: a 1s function and a 3s function. Doing this has been called *unrestricted* kinetic balance (UKB), also called *extended* kinetic balance [24]; keeping the linear combination of the two functions has been called *restricted* kinetic balance [25]. From a computational standpoint, this can be useful if the same exponents are used for both spin-orbit components, as the 3s function can be obtained by transformation of the Cartesian set of six d functions to a spherical basis.

For a scalar basis, in which scalar kinetic balance is applied to a Cartesian Gaussian, the result is

$$\frac{\mathrm{d}}{\mathrm{d}x} x^i y^j z^k e^{-\zeta r^2} = i x^{i-1} y^j z^k e^{-\zeta r^2} - 2\zeta x^{i+1} y^j z^k e^{-\zeta r^2} \tag{54}$$

for the x derivative and similar expressions for the y and z derivatives. All of these small-component functions consist of two radial functions, unless any of i, j, or k is zero. However, from the set of all the derivatives, linear combinations can be taken that reduce to single functions of Cartesian Gaussian type. (The exception is the derivatives of the p functions, which result in six nonredundant functions, but there are seven single functions: six Cartesian 3d functions and one Cartesian 1s function. In this case the seventh function is split off, just as for UKB.) These functions span the required space: one unit higher and one unit lower in angular momentum than the large component.

The next issue is the relationships between the exponents. The kinetic balance requirement already imposes the condition that the exponents for the small component basis functions must be the same as for the corresponding large-component basis functions. From a practical standpoint, it is desirable to minimize the computational cost and maximize the numerical stability. If a 2-spinor basis is used, the exponents of the large-component functions can be chosen at will, as there is never a linear dependence problem in the small component arising from kinetic balance. If a scalar basis is used, kinetic balance produces a set of p functions in the small component arising from the large-component s functions and another set of p functions from the large- component d functions. As these two p sets are highly linearly dependent, it makes sense to use a single set of p functions to eliminate the linear dependence. This imposes the requirement that the d exponents for the large component are a subset of the s exponents. A similar consideration applies for the large-component p and f sets: both generate d functions for the small component, so making the large-component f set a subset of the p set eliminates the linear dependence. Two interleaving sets of exponents are obtained, which could be described as "large s, small p, large d, small f ..." and "small s, large p, small d, large f ...." Sets of this type are termed "dual family" basis sets [26]. In addition to the removal of some linear dependence problems, the use of common exponents results in more efficient integral evaluation, as it reduces the number of integrals to be evaluated.

Even more efficiency could be obtained by using a common set of exponents for all angular momenta, and the same integrals would be used for both the large and the small components. These basis sets are termed "family" basis sets [26].

Dual family basis sets do not entirely remove linear dependence from the small component, however. The derivatives of the 3d Gaussians include both 2p-type and 4p-type Gaussians. The set of 4p-type Gaussians has a fair degree of linear dependence on the set of 2p-type Gaussians, so taking the two sets as independent functions will require some approach for removing or alleviating the linear dependence. Retaining the linear combinations does not do this, as the 2p functions would still be needed as well as the combined 2p + 4p functions obtained from scalar kinetic balance. The usual strategy of using a cutoff on the eigenvalues of the overlap matrix to remove linear dependencies is one approach; another is to perform the integral work and Fock matrix accumulation in a scalar basis and transform the Fock and overlap matrices to a 2-spinor basis for solution of the SCF equations.

These considerations only apply to four-component methods. In two-component methods, the relation between the large and small components is folded into the operators, and the basis set is essentially that of the large component. In this context, the basis sets will differ from nonrelativistic basis sets only so far as the relativistic changes in the radial wave function affect the exponent range.

Finally, the exponents themselves must be chosen. Nonrelativistic basis sets are usually developed by minimizing the total energy with respect to exponent variations. The same approach can be used in relativistic calculations but with the caveat that the total energy is not bounded from below by the exact energy. Instead, it can deviate from the exact energy by an amount of order c^{-4}, as kinetic balance guarantees that the nonrelativistic energy is an upper bound to the exact nonrelativistic energy, and the first-order relativistic correction (of order c^{-2}) is obtained from the nonrelativistic wave function. Due to the relativistic contraction of the core, more basis functions are needed in the core than in nonrelativistic basis sets, particularly in the p space: for example, relativistic triple-zeta basis sets for the 6p block have one more s function and three more p functions than the corresponding nonrelativistic basis sets [12].

Another way of choosing the exponents is to use an even-tempered set, $\zeta_i = \alpha\beta^{i-1}$. Here, there are only two parameters to optimize or to choose, α and β, for a given number of exponents. These sets tend to be larger than the energy-optimized sets for a given valence quality, because the energy-optimized exponents spread out in the high end of the exponent range and to some extent also at the low end of the exponent range.

Contracted Basis Sets

The use of uncontracted basis sets is relatively uncommon in nonrelativistic calculations: most calculations are performed with contracted basis sets. Contraction reduces the number of basis functions and hence the work done in SCF and post-SCF methods. For heavy elements, flexibility in the core is rarely required in molecular calculations, so it makes sense to contract the core orbitals or spinors.

Scalar relativistic calculations are functionally the same as nonrelativistic calculations, differing only in the operators used to construct the kinetic and potential energy matrix elements. Consequently, contraction can be done in the same way and presents no additional problems beyond those that may be encountered in nonrelativistic calculations. Both generally contracted basis sets and segmented basis sets can be used without any particular issues.

Two-component calculations are a different matter. Here, the spin cannot be factored out, due to the presence of the spin-dependent operators. The natural basis set is a 2-spinor basis, as presented in the section "Spin Coupling". In an uncontracted basis set, the spin-orbital basis spans the same space as a 2-spinor basis, so it can be used instead. With contracted basis sets, the contraction coefficients must be determined separately for each spin-orbit component. Using these contractions in a 2-spinor basis presents no linear dependence problems, as the components are associated with angular functions that are eigenfunctions of the spin-orbit operator. Using the contractions in a spin-orbital basis produces a doubling of the basis for each spin-orbit split shell, as there are two radial functions to associate with each spherical harmonic and spin function. This leads to a large degree of linear dependence, as the radial overlap of the two spin-orbit radial functions is usually quite large, and almost unity in the valence region. If the spin-orbital basis is used throughout the calculation, this linear dependence must be removed. An alternative is to use a spin-orbital basis that is contracted in a scalar relativistic calculation. Such a basis provides for first-order spin-orbit effects, but the higher-order effects due to relaxation of the spin-orbit components are largely absent (depending on how the contraction is done).

Four-component calculations present additional issues. Consider first the case where the contraction coefficients are derived from an atomic calculation. Applying kinetic balance directly to the contracted large component yields small-component functions that deviate considerably from the actual small component, particularly in the core region where the potential is large. What is needed here is not kinetic balance, as given by Eq. (15), but *atomic balance* [24], as given by Eq. (5), or its equivalent from an SCF calculation. In other words, the atomic large- and small-component 2-spinors are used as contracted basis functions in the molecular calculation. This choice builds in the information from the atomic calculation, which is most important in the core.

As for two-component calculations, the issue of linear dependence is a major one if a spin-orbital basis is used and applies to both the large and the small components. Dual family primitive sets do not help here, as the contracted small components of the same angular momentum that arise from different large components are not identical, even if the primitives in the contraction are.

There is a further problem with contraction in four-component calculations. In order to separate the positive-energy and negative-energy states, a reasonable representation must be obtained for both. Otherwise, it is possible to produce intruder states, where, for example, a negative-energy state is pushed up into the positive-energy region and mixes with the positive-energy states. In an uncontracted basis, it is possible to obtain a fairly good representation of the negative-energy

states even with kinetic balance for the positive-energy states only, as the kinetically balanced basis for the negative-energy states is fairly well represented by the kinetically balanced basis for the positive-energy states. In a contracted basis, the heavily contracted small components do not provide a good representation of the negative-energy states, as they are contracted for an electronic spinor, not for a positronic spinor. The presence of intruder states with heavily contracted basis sets was observed in calculations on molecules containing 7p elements [20]. Contracted basis sets for four-component calculations should therefore be used with some caution, and it may be better to use uncontracted basis sets instead.

The same issue could arise if segmented contractions are used. Segmented contractions are usually developed in atomic calculations on electronic states, just as general contractions are. As the contractions are adapted for the electronic (positive-energy) states, they do not represent the negative-energy states very well.

Magnetic Balance

While it is often possible for basis sets developed for SCF and correlated calculations to be used for calculating properties, these basis sets may need to be extended for accurate property calculations. For example, calculation of polarizabilities usually needs functions that are more diffuse than in standard basis sets, as the first-order wave function must be able to represent the effect of the dipole operator on the zeroth-order wave function. This requirement implies that functions of higher angular momentum are also needed.

On the whole, for one-component and two-component relativistic calculations, the basis set requirements for properties follow the nonrelativistic requirements. In four-component calculations, the basis set requirements for electric properties are not much different from the nonrelativistic requirements, as the electric field appears in the Dirac equation in much the same way as in the Schrödinger equation. No extension beyond kinetic balance is required to meet the nonrelativistic limit, as the electric perturbation appears as a part of the potential V in the denominator of Eq. (5) and is eliminated in the nonrelativistic limit.

Magnetic properties are a different matter and require a more detailed treatment. The reason is that the vector potential, which is related to the magnetic field, appears along with the momentum operator in the terms that couple the large and small components. The Dirac equation in a magnetic field is written in two-component form with the rest mass subtracted, as two coupled equations,

$$\begin{aligned}(V - E)\psi^L + c\boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})\psi^S &= 0 \\ c\boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})\psi^L + (V - 2mc^2 - E)\psi^S &= 0,\end{aligned} \tag{55}$$

where e is the elementary charge, and the derivation is based on SI units. The relation between the small and large components now involves the magnetic field as well as the momentum,

$$\psi^S = (2mc^2 + E - V)^{-1} c\boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})\psi^L. \tag{56}$$

Eliminating the small component and taking the nonrelativistic limit, the result is

$$(V - E)\psi^L + \frac{1}{2m}\boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})\boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})\psi^L = 0 \tag{57}$$

which, with the use of the Dirac relation

$$(\boldsymbol{\sigma} \cdot \mathbf{u})(\boldsymbol{\sigma} \cdot \mathbf{v}) = \mathbf{u} \cdot \mathbf{v} + i\boldsymbol{\sigma} \cdot (\mathbf{u} \times \mathbf{v}) \tag{58}$$

yields the familiar nonrelativistic magnetic operators,

$$\hat{H}_{\text{mag}} = \frac{e}{2m}(\mathbf{p} \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}) + \frac{e^2}{2m}\mathbf{A}^2 + \frac{e\hbar}{2m}\boldsymbol{\sigma} \cdot \mathbf{B}. \tag{59}$$

The presence of the vector potential in the relation between the large and small components necessitates a rederivation of the nonrelativistic limit conditions on the basis set. The matrix Dirac equation with the vector potential included is written as

$$(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\mathbf{a} + c(\boldsymbol{\Pi}^{LS} + e\mathbf{A}^{LS})\mathbf{b} = 0 \tag{60}$$

$$c(\boldsymbol{\Pi}^{SL} + e\mathbf{A}^{SL})\mathbf{a} + (\mathbf{V}^{SS} - (2mc^2 + E)\mathbf{S}^{SS})\mathbf{b} = 0, \tag{61}$$

where the matrix of the vector potential is

$$A^{LS}_{\mu\nu} = \langle \chi^L_\mu | \boldsymbol{\sigma} \cdot \mathbf{A} | \chi^S_\nu \rangle = (A^{SL}_{\nu\mu})^\dagger. \tag{62}$$

As before, $\mathbf{b}$ is eliminated from the first equation using the second, with the following result when the nonrelativistic limit is taken:

$$(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\mathbf{a}^L + \frac{1}{2m}(\boldsymbol{\Pi}^{LS} + e\mathbf{A}^{LS})(\mathbf{S}^{SS})^{-1}(\boldsymbol{\Pi}^{SL} + e\mathbf{A}^{SL})\mathbf{a}^L = 0, \tag{63}$$

To ensure that the nonrelativistic magnetic operators are obtained, the following choice is made for the relation between the small- and large-component basis functions:

$$\chi^S = \boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})\chi^L. \tag{64}$$

With this choice,

$$\Pi^{LS}_{\mu\nu} = \langle \chi^L_\mu | \boldsymbol{\sigma} \cdot \mathbf{p}\, \boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A}) | \chi^L_\nu \rangle \tag{65}$$

$$A^{LS}_{\mu\nu} = \langle \chi^L_\mu | \boldsymbol{\sigma} \cdot \mathbf{A}\, \boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A}) | \chi^L_\nu \rangle \tag{66}$$

$$S^{SS}_{\mu\nu} = \langle \chi^L_\mu | [\boldsymbol{\sigma} \cdot (\mathbf{p} + e\mathbf{A})]^2 | \chi^L_\nu \rangle \tag{67}$$

so that

$$\boldsymbol{\Pi}^{LS} + e\mathbf{A}^{LS} = \mathbf{S}^{SS} = \boldsymbol{\Pi}^{SL} + e\mathbf{A}^{SL} \tag{68}$$

The nonrelativistic limit reduces to

$$(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\mathbf{a}^{L} + \frac{1}{2m}\mathbf{S}^{SS}\mathbf{a}^{L} = 0, \tag{69}$$

and from the definition of the small-small overlap above, it is clear that the expectation of the nonrelativistic magnetic operator is indeed obtained. The condition in Eq. (64) is called *magnetic balance* [27, 28].

In most applications, the magnetic field is small, so it is treated as a perturbation. The zeroth-order Hamiltonian is field-free, so kinetic balance can be used for the basis of the zeroth-order wave function, and magnetic balance is only needed in the perturbed wave function.

The two most important cases of magnetic fields are an applied uniform field, $\mathbf{A} = \frac{1}{2}\mathbf{B} \times \mathbf{r}$, and a nuclear magnetic field, $\mathbf{A} = \boldsymbol{\mu} \times \mathbf{r}/r^3$, for a point nucleus. In both cases, the critical part of the operator for magnetic balance is the vector product. When combined with the spin operator, the result is

$$\boldsymbol{\sigma} \cdot \mathbf{A} = f(r)\, \boldsymbol{\sigma} \cdot \mathbf{D} \times \mathbf{r} = -rf(r)\, \mathbf{D} \cdot \boldsymbol{\sigma} \times \hat{\mathbf{r}} \tag{70}$$

where $\mathbf{D}$ is a constant vector that relates to the field strength. The basis functions themselves will not include this constant vector, so the magnetic component of the basis functions can be written as

$$\chi_{\mu}^{S} = -rf(r)\, \boldsymbol{\sigma} \times \hat{\mathbf{r}}\, \chi_{\mu}^{L}. \tag{71}$$

The vector product is now a pure spin-angular operator and gives the relation between the angular functions. As a tensor of rank 1, it can change the total angular momentum by one unit, and therefore it couples large-component functions with small-component functions of different angular momentum (as well as those of the same angular momentum). This is in contrast to kinetic balance, where the tensor is of rank 0 and couples only large- and small-component functions of the same angular momentum. Due to the fact that the angular operator $\hat{\mathbf{r}}$ is a tensor of rank 1, the orbital angular momentum of the small component must differ from that of the large component by ± 1, just as for kinetic balance. This means that, as for kinetic balance, the small-component magnetic basis can be expressed in terms of functions of one unit higher and one unit lower in orbital angular momentum, but both j values must be included.

For example, magnetic balance for an $s_{1/2}$ spinor generates both $p_{1/2}$ and $p_{3/2}$ spinors; for a $p_{1/2}$ spinor, both $s_{1/2}$ and $d_{3/2}$ spinors are generated; and for a $p_{3/2}$ spinor, $s_{1/2}$, $d_{3/2}$, and $d_{5/2}$ spinors are generated. As the space of these spinors is spanned by the set of Cartesian spin orbitals, the use of a Cartesian basis in this

application has some advantages. These functions must still be combined with the field strength operator in the construction of the matrix, to ensure that the proper balance is kept.

The radial functions differ from those generated by kinetic balance and obviously depend on the type of magnetic field. For a uniform magnetic field, the small components behave as one power of r higher than the large components. There is no linear combination of $r^{\ell-1}$ and $r^{\ell+1}$ in the radial part $\kappa > 0$ as there is for kinetic balance. So, for example, magnetic balance generates a $3s_{1/2}$ spinor from a $2p_{1/2}$ spinor, whereas kinetic balance generates a linear combination of $1s_{1/2}$ and $3s_{1/2}$ spinors.

The main difficulty in the radial part lies with the nuclear magnetic field, where $f(r) = 1/r^2$. Operating on a $1s_{1/2}$ spinor would result in a $\mathrm{p}_{1/2}$ spinor in the small component with a prefactor of $1/r^2$, which is singular. However, the difficulty is no greater than in the nonrelativistic case. If magnetic balance as defined by Eq. (64) is applied in Eq. (60), with a prefactor of $1/2mc$, the result is

$$\begin{aligned}
&(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\,\mathbf{a} + (\mathbf{T}^{LL} + \mathbf{H}^{LL}_{\mathrm{mag}})\,\mathbf{b} = 0,\\
&(\mathbf{T}^{LL}+\mathbf{H}^{LL}_{\mathrm{mag}})\,\mathbf{a}+\left((\mathbf{W}^{LL}+\mathbf{Y}^{LL})/4m^2c^2-(1+E/2mc^2)(\mathbf{T}^{LL}+\mathbf{H}^{LL}_{\mathrm{mag}})\right)\,\mathbf{b}=0.
\end{aligned} \tag{72}$$

where $\mathbf{Y}^{LL}$ is the matrix of the operator

$$\hat{Y} = \frac{e}{2m}(\mathbf{p}\cdot\mathbf{A}V + V\mathbf{A}\cdot\mathbf{p}) + \frac{e^2}{2m}\mathbf{A}^2V + \frac{e\hbar}{2m}\boldsymbol{\sigma}\cdot\mathbf{B}V + \frac{e\hbar}{2m}\boldsymbol{\sigma}\cdot(\nabla V)\times\mathbf{A}. \tag{73}$$

This operator encompasses relativistic corrections to the (nonrelativistic) magnetic interaction, and there will also be two-electron corrections.

The magnetic integrals $\mathbf{H}^{LL}_{\mathrm{mag}}$ are the same as in the nonrelativistic case, and any problems with the nuclear terms can in principle be addressed in the same way. The main issue is the behavior at the nucleus. If the nucleus is treated as a point, then the resultant singularities in the relativistic wave function must be dealt with. However, if a finite nuclear charge distribution is used, there is no singularity. For the magnetic properties, a finite nuclear current distribution would be required, and the interaction of the nuclear magnetic field with the electronic magnetic field would be evaluated as a kind of Gaunt interaction,

$$\hat{H}_{\mathrm{int}} = \alpha_e \cdot \alpha_n / r_{en} \tag{74}$$

where the subscripts e and n refer to the electronic and nucleonic coordinates, respectively. It would be necessary to integrate over the nuclear wave function to obtain the vector potential of the nucleus. In the absence of such a wave function, a model vector potential could be used, just as for the nuclear charge distribution. This should then eliminate the singularities.

The relativistic correction integrals $\mathbf{Y}^{LL}$ pose similar problems, as they also include the nuclear potential. Again, a finite nucleus should eliminate the singularities, along with the behavior of the wave function near the nucleus imposed by angular symmetry considerations.

References

1. Kutzelnigg W (1994) Theory of the expansion of wave functions in a Gaussian basis. Int J Quantum Chem 51:447
2. Mark F, Schwarz WHE (1981) New representation of the [α] . p operator in the solution of Dirac-type equations by the linear-expansion method. Phys Rev Lett 48:673
3. Schwarz WHE, Wallmeier H (1982) Basis set expansions of relativistic molecular wave equations. Mol Phys 46:1045
4. Dyall KG, Grant IP, Wilson S (1984) Matrix representation of operator products. J Phys B 17:493
5. Stanton RE, Havriliak S (1984) Kinetic balance: a partial solution to the problem of variational safety in Dirac calculations. J Chem Phys 81:1910
6. Shabaev VM, Tupitsyn II, Yerokhin VA, Plunien G, Soff G (2004) Dual kinetic balance approach to basis-set expansions for the Dirac equation. Phys Rev Lett 93:130405
7. Sun Q, Liu W, Kutzelnigg W (2011) Comparison of restricted, unrestricted, inverse, and dual kinetic balances for four-component relativistic calculations. Theor Chem Acc 129:423
8. Ishikawa Y, Baretty R, Binning RC Jr (1985) Relativistic Gaussian basis set calculations on one-electron ions with a nucleus of finite extent. Chem Phys Lett 121:130
9. Visser O, Aerts PJC, Hegarty D, Nieuwpoort WC (1987) The use of Gaussian nuclear charge distributions for the calculation of relativistic electronic wavefunctions using basis set expansions. Chem Phys Lett 134:34
10. Visscher L, Dyall KG (1997) Dirac-Fock atomic electronic structure calculations using different nuclear charge distributions. At Data Nucl Data Tables 67:207
11. Dyall KG (1998) Relativistic and nonrelativistic energy-optimized polarized double zeta basis sets for the 4p, 5p, and 6p elements. Theor Chem Acc 99:366
12. Dyall KG (2002) Relativistic and nonrelativistic energy-optimized polarized triple-zeta basis sets for the 4p, 5p, and 6p elements. Theor Chem Acc 108:335
13. Dyall KG (2004) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the 5d elements Hf-Hg. Theor Chem Acc 112:403
14. Dyall KG (2006) Relativistic quadruple-zeta basis sets and revised triple-zeta and double-zeta basis sets for the 4p, 5p, and 6p elements. Theor Chem Acc 115:441
15. Dyall KG (2007) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the 4d elements Y-Cd. Theor Chem Acc 117:483
16. Dyall KG (2007) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the actinide elements Ac–Lr. Theor Chem Acc 117:491
17. Dyall KG (2009) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the 4s, 5s, 6s, and 7s elements. J Phys Chem A 113:12638
18. Gomes ASP, Visscher L, Dyall KG (2010) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the lanthanide elements La–Lu. Theor Chem Acc 127:369
19. Dyall KG (2011) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the 6d elements Rf–Cn. Theor Chem Acc 129:603
20. Dyall KG (2012) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the 7p elements, with atomic and molecular applications. Theor Chem Acc 131:1172
21. Rose SJ, Grant IP, Pyper NC (1978) The direct and indirect effects in the relativistic modification of atomic valence orbitals. J Phys B: At Mol Phys 11:1171

22. Matsuoka O, Okada S (1989) Dirac-Fock-Roothaan calculations on the sixth-row elements Tl–Rn. Chem Phys Lett 155:547
23. Visscher L, Dyall KG (1996) Relativistic and correlation effects on molecular properties. I. The dihalogens F_2, Cl_2, Br_2, I_2, and At_2. J Chem Phys 104:9040
24. Visscher L, Aerts PJC, Visser O, Nieuwpoort WC (1991) Kinetic balance in contracted basis sets for relativistic calculations. Int J Quantum Chem Symp 25:131
25. Dyall KG, Fægri K Jr (1990) Kinetic balance and variational bounds failures in the solution of the Dirac equation in a finite Gaussian basis set. Chem Phys Lett 174:25
26. Fægri K Jr (2001) Relativistic Gaussian basis sets for the elements K–Uuo. Theor Chem Acc 105:252
27. Aucar GA, Saue T, Visscher L, Jensen HJAa (1999) On the origin and contribution of the diamagnetic term in four-component relativistic calculations of magnetic properties. J Chem Phys 110:6208
28. Kutzelnigg W (1999) Relativistic corrections to magnetic properties. J Comput Chem 20:1199

Relativistic Self-Consistent Fields

4

Christoph van Wüllen

Contents

Introduction . . . 108
Relativistic Hartree-Fock for Atoms . . . 109
Negative-Energy Eigenstates and the "Dirac Sea" . . . 110
Open-Shell Atoms: Multiconfiguration Dirac-Fock . . . 112
Beyond Dirac-Coulomb . . . 113
The Breit and Gaunt Interaction . . . 113
Nuclear Volume and Recoil Corrections . . . 115
Basis Set Expansion Methods . . . 115
Relativistic Hartree-Fock Matrix Equations . . . 116
The "Diseases" and How to Become Immune . . . 119
New Developments . . . 120
Relativistic Kohn-Sham . . . 121
Relativistic Density Functional Theory Basics . . . 121
Relativistic Spin-Density Functional Procedures . . . 123
Relativistic Functionals . . . 123
New Developments . . . 124
Beyond Self-Consistent Field Methods . . . 124
Summary . . . 125
References . . . 126

Abstract

The simplest relativistic computational methods for many-electron systems involve the solution of one-particle Dirac equations for an electron moving in some effective ("mean-field") potential. This potential depends on the one-particle solutions which describe the electron charge distribution; therefore, such mean-field problems are solved iteratively until self-consistency. The two most

C. van Wüllen (✉)
Fachbereich Chemie, Technische Universität Kaiserslautern, Kaiserslautern, Germany
e-mail: vanwullen@chemie.uni-kl.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_24

important relativistic self-consistent field methods are the relativistic variants of the Hartree-Fock and Kohn-Sham methods, whose computational frameworks have large overlap. In this chapter, the development of these methods for atoms and molecules is sketched. While atomic calculations are usually performed solving differential variational equations, molecular calculations rely on basis set expansion methods. These seemed to be problematic initially, but with kinetically balanced basis sets, smooth convergence is obtained. Most relativistic Kohn-Sham calculations performed today combine relativistic kinematics (the use of Dirac spinors) with nonrelativistic exchange-correlation functionals.

Keywords

Relativistic • Mean field • Effective potential • Atomic structure • Multiconfiguration • Hartree-Fock • Dirac-Fock • Kohn-Sham • Kinetic balance • Dirac sea

Introduction

The properties of the Dirac equation for a single spin-$\frac{1}{2}$ particle are known in much detail. Whereas even a hydrogenic atomic ion is a two-particle system (electron and nucleus), the nucleus in a first approximation is just the source of a classical electrostatic field in which the electron moves. Then, hydrogenic ions may be described by the one-particle Dirac equation. This description leads to results that match experimental data fairly well, but some finer details require to go beyond. For example, real nuclei are not point-like but have a finite extension and may be non-spherical which leads to the occurrence of nuclear electric and magnetic multipoles; there are errors introduced by the assumption of an infinitely massive nucleus (these are termed nuclear recoil effects), and differences to a complete field theoretical approach in which both the matter and the radiation fields are quantized are called quantum electrodynamical (QED) corrections. Among the most notable QED corrections is the Lamb shift (which, e.g., removes the degeneracy of the $2s_{\frac{1}{2}}$ and $2p_{\frac{1}{2}}$ energy levels) and a small correction to the electron g value.

In atomic and molecular electronic structure theory, one is however primarily interested in many-electron systems. Compared to the nonrelativistic case, the increase in complexity when going beyond a one-electron system is much more pronounced in a relativistic treatment. The first steps of a complete (or at least convergent) treatment in the framework of bound-state QED are possible for few-electron atoms (or atomic ions), but it is unlikely that this becomes a viable computational approach to real many-electron systems. Designing a practical computational method for a relativistic many-electron system that is firmly based on first principles is a formidable challenge, but things remain relatively straightforward if one uses a mean-field approximation that involves the solution of single-particle Dirac equations for electrons moving in some (effective) potential. Usually, this effective potential depends on the Dirac solutions it generates, which leads to some kind of self-consistency cycle.

While there may (or may not) be a relativistic many-electron wave function present behind the curtain, the computational procedure only deals with single-particle functions. But besides the more or less technical issue of how to calculate self-consistent single-particle functions (Dirac spinors), there are two valid questions: (a) If this is an approximation, what is actually approximated? (b) How can one go beyond this approximation? Such questions were asked when relativistic self-consistent methods were already established, and this caused some confusion (see below) because the reasons for some inconsistencies observed in numerical calculations were not properly understood. But meanwhile relativistic self-consistent field methods, in today's nomenclature subdivided into relativistic Hartree-Fock and relativistic Kohn-Sham methods, are a well-established first approximation to the relativistic many-electron problem.

Relativistic Hartree-Fock for Atoms

The first attempt at a relativistic Hartree-Fock approach was made by Bertha Swirles as early as 1935 [1]. Since a fully relativistic many-electron Hamiltonian does not exist, she introduced an approximation, namely, the Hamiltonian for noninteracting Dirac particles, augmented with the expression for the electron interaction as known from nonrelativistic theory:

$$\hat{H}_{DC} = \sum_{i=1}^{n} \hat{h}_D(i) + \sum_{i<j} \frac{1}{r_{ij}} \tag{1}$$

Here, n is the number of electrons and $\hat{h}_D$ is the Dirac operator for a single electron moving in the electric field of the nuclear framework, described by its electrostatic potential. In standard notation, the Dirac operator reads (with a shift that places the positive continuum at zero energy)

$$\hat{h}_D = \mu \begin{pmatrix} V_N & c\boldsymbol{\sigma} \cdot \mathbf{p} \\ c\boldsymbol{\sigma} \cdot \mathbf{p} & V_N - 2mc^2 \end{pmatrix} \tag{2}$$

Atomic Hartree units are used throughout, which means that the elementary charge e and the electron rest mass m are set to unity from now on, c is the speed of light ($\approx$137 in atomic units), $\boldsymbol{\sigma}$ is the vector of the Pauli spin matrices, and $\mathbf{p}$ is the momentum operator. Note that the nuclear framework is fixed and at rest, so a particular Lorentz frame has been singled out. This Hamiltonian, today called the Dirac-Coulomb Hamiltonian, is still a cornerstone of relativistic quantum chemistry although it has some awkward properties if understood as a Hamiltonian in configuration space. Swirles then considered a Slater determinant Ψ,

$$\Psi = |\phi_1 \phi_2 \ldots \phi_n| \tag{3}$$

an antisymmetrized product built from orthonormal Dirac spinors ϕ_i, and demanded that these spinors are optimized as to produce the lowest possible energy expectation value $E = <\Psi|\hat{H}_{DC}\Psi>$ of the Slater determinant with the Dirac-Coulomb operator. The condition of variational stability then implies that the spinors are eigenfunctions of a one-particle Dirac operator with an effective potential that in turn depends on these solutions. This generates a self-consistency condition in complete analogy to the nonrelativistic (Hartree-Fock) case, and the self-consistent effective potential is termed the self-consistent field.

Swirles herself did not present numerical calculations; subsequently calculations were presented by other workers either neglecting the exchange terms altogether [2] or approximating them by "statistical exchange" [3]. The latter approximation, also called Dirac-Fock-Slater or Dirac-Slater, is today viewed as a Dirac-Kohn-Sham type method, but at that pre-DFT time, Slater exchange was considered as an approximation to Fock exchange. Finally, Grant [4, 5] mastered all the angular momentum algebra necessary to set up a computational procedure for atomic calculations including Fock exchange. In 1973, Desclaux published Dirac-Fock results for all atoms with $Z = 1$ through $Z = 120$ [6]. This work is essentially a compilation of numerical results, but turned out to be highly influential. This is so, because relativistic effects on atomic shell structures, and its consequence for chemical bonding, were often qualitatively discussed based on Desclaux' tables.

Negative-Energy Eigenstates and the "Dirac Sea"

The approach pioneered by Swirles does not look valid at first sight. For example, possible problems associated with the existence of negative-energy Dirac solutions were not even mentioned. As pointed out later in detail by Grant and Quiney [7], variational stability problems cannot arise here because the numerical procedure generates Dirac spinors that satisfy the correct boundary conditions. The angular and spin parts of the eigenfunctions of a Dirac operator with a spherical (effective) potential are known analytically, and the radial part is obtained by numerically solving radial variational differential equations imposing the proper boundary conditions at $r \to \infty$ and $r \to 0$. Scattering eigenfunctions (both with positive or negative energy) cannot be obtained this way; therefore, there is an implicit projection to the positive-energy subspace, and the one-particle eigenvalue obtained this way has a lower bound.

The numerical procedure for the calculation of the spinors thus implicitly contains a projector to the bound positive-energy subspace, which makes the method variationally stable. As a matter of fact, negative-energy eigenfunctions of the Fock operator *do* exist, but the early work on relativistic Hartree-Fock obviously did not see any reason to consider them – although they had irritated Dirac and others so much just a decade before. Why were they disregarded when choosing the occupied spinors? An argument that already goes back to Dirac is that the negative-energy spinors are all occupied, but that this cannot be seen, as the "normal state of electrification," as Dirac termed it, corresponds to all but only the negative-energy

states occupied and that everything one measures is just the departure from this normal state. The Pauli exclusion principle then enforces that additional electrons have to occupy positive-energy spinors. The charge density of the filled Dirac sea of the free particle Dirac equation, in today's terminology the vacuum, will be uniform but not detectable although it is formally infinitely large. The early relativistic Hartree-Fock procedures can thus be justified as defining vacuum expectation values to be zero, such that only the positive-energy occupied spinors contribute.

There is however a problem here, namely, that in the presence of the atomic nucleus, the negative eigenstates of the Dirac equation, if completely filled, do not produce a homogeneous density, because the electric field of the nucleus induces an asymmetry between positive and negative-energy one-particle states. This remains true if the effective Dirac equation containing the Fock potential is considered. In other words, the vacuum becomes polarized. Note that there are additional contributions from pair creation not contained in this picture. The vacuum polarization charge density constitutes a correction to the electron-nucleus interaction and contributes to the removal of the degeneracy of the $2s_{\frac{1}{2}}$ and $2p_{\frac{1}{2}}$ levels already in one-electron atoms (the so-called Lamb shift). From a conceptual point of view, the vacuum polarization charge density is important in relativistic density functional theory, where a charge density need be defined independently of the external potential. However, if one attempts to calculate the interaction with the vacuum polarization charge density, one finds that it becomes infinite.

The Dirac sea picture is a nice starting point to discuss what negative-energy states are all about, but must be considered obsolete. For example, there are massive bosons with negative-energy eigenstates, and here the Pauli exclusion principle does not prevent them all falling into the sea. Quantum field theory – or quantum electrodynamics (QED), as it is called for interacting electrons – is able to treat the negative-energy eigenstates consistently which are then related to positive-energy states of an antielectron (positron). Nevertheless, the problems of the infinities of the Dirac sea remain present, and it requires normal ordering and a proper renormalization procedure to arrive at finite results. A complete description of interacting electrons requires a quantization of both the matter and the radiation field, and this then adds a second term to the Lamb shift, the so-called electron self-energy. This self-energy contribution is actually as important as the vacuum polarization contribution and of opposite sign (in light atoms, it is actually much larger), both together constitute the leading-order QED correction to one-electron atoms. One should also not forget that for heavy nuclei, the finite nuclear volume effect is also an equally important contributor to the Lamb shift. Of course, in many-electron atoms, there are also QED corrections to the electron-electron interaction. For group I atomic ionization energies, the corrections due to the Lamb shift have been found to be as large as those from relativistic corrections to the electron-electron interaction [8].

Another quantum field correction, which can quite easily be calculated for a free electron, is the deviation of its g value from the Dirac result $g_D = 2$ (for a free electron, $g_f \approx 2.002319$). The QED correction to the g value is more difficult

to calculate for a bound electron. When calculating magnetic properties, the Dirac operator is often modified by scaling magnetic interactions with an external field by $g_f/2$ and magnetic interactions between electrons by $g_f^2/4$. For the calculation of electron paramagnetic resonance (EPR) parameters of molecules containing only light elements, this is certainly a useful first estimate of QED effects.

Until today, the largest part of relativistic self-consistent field methods uses the "no sea" (today mostly called "no-pair") approximation tacitly introduced by Swirles, where negative-energy states just remain unoccupied and any contributions from quantum field theory are ignored. For few-electron ions and molecules, such as the helium or lithium isoelectronic series and the hydrogen molecule, such contributions can be calculated to high accuracy. It still seems state of the art to use semiempirical estimates to assess QED effects in many-electron atoms and molecules [9]. There is currently on-going work on how to include QED corrections systematically in atomic and molecular calculations [10–12].

Open-Shell Atoms: Multiconfiguration Dirac-Fock

Early atomic relativistic self-consistent field calculations could only calculate total energies of closed-shell atomic states or an energy average (over all states arising from a given configuration) for open-shell situations. Although such results were important for the development of our qualitative understanding of relativistic effects in atoms and molecules, they were useless for the simulation of atomic spectra. For this purpose, the treatment of partially filled atomic shells had to be improved.

The Dirac spinors used to construct the atomic wave functions were, by construction, eigenfunctions of the one-particle total angular momentum operator $\mathbf{j}^2$ and its z component $\hat{j}_z$. While Slater determinants constructed this way are still eigenfunctions of the many-electron operator $\hat{J}_z$, one generally needs linear combinations of Slater determinants of a given configuration to construct atomic wave functions of proper symmetry, which are eigenfunctions of $\mathbf{J}^2$. A textbook example is the atomic p^2 configuration. In the relativistic case, the p shell is split into the $p_{\frac{1}{2}}$ and $p_{\frac{3}{2}}$ subshells, such that there are three subshell configurations:

$$\begin{aligned} \left(p_{\frac{1}{2}}\right)^2 &: N_{Det} = 1;\ J = 0 \\ \left(p_{\frac{1}{2}}\right)^1\left(p_{\frac{3}{2}}\right)^1 &: N_{Det} = 8;\ J = 1, J = 2 \\ \left(p_{\frac{3}{2}}\right)^2 &: N_{Det} = 6;\ j = 0, J = 2 \end{aligned}$$

For the $(p_{\frac{1}{2}})^2$ configuration, there is only a single Slater determinant (closed-shell case) which is eigenfunction of $\mathbf{J}^2$ with $J = 0$, and the radial part of the spinors is optimized for this $J = 0$ energy. For the $(p_{\frac{3}{2}})^2$ configuration on the other hand, one can construct six Slater determinants by occupying two of the four $p_{\frac{3}{2}}$ spinors. There are fixed linear combinations of these Determinants which are eigenfunctions of $\mathbf{J}^2$, a single one with $J = 0$ and a fivefold degenerate set that forms a multiplet

with $J = 2$. One can optimize the radial parts of the Dirac spinors for either of the two resulting energy levels. This procedure is known as jj-coupling. The problem is now that for most atoms of interest, the interaction between the two $J = 0$ functions arising from the $\left(p_{\frac{1}{2}}\right)^2$ and $\left(p_{\frac{3}{2}}\right)^2$ configurations, mediated by exchange, is larger than their energy difference which depends on the spin-orbit splitting of the p shell. It must therefore be expected that a realistic description involves a strong mixing between the two $J = 0$ functions obtained from jj-coupling. Describing this mixing was not possible in the first atomic structure programs; the radial parts of the Dirac spinors were instead optimized for an average of all these energies (weighted with the degeneracies). The average energy thus obtained roughly corresponds to an energy average of the nonrelativistic LS configurations (in this case, ^3P, ^1D, ^1S). Note that in group 14 atoms with a valence p^2 configuration, these LS-multiplets are separated by about 10,000 cm^{-1}; therefore, the calculation of such averaged energies is of little significance for the interpretation of atomic spectra.

The configuration mixing necessary for a realistic description of atomic energy levels is performed in the multiconfiguration Dirac-Fock (MCDF) method. First implementations were presented by Desclaux [13] and Grant [14]. This method introduces a second self-consistency cycle: for a given linear combination of Slater determinants, the variational radial differential equations are solved until self-consistency. Then, the Hamiltonian matrix is constructed and diagonalized in the space of these determinants, which generates new coefficients for these. This work is largely based on the achievements by Froese Fischer in the nonrelativistic case [15]. Even if a good wave function has been obtained as a linear combination of jj-coupled determinants, it is of interest how this function reads if expanded in LS-terms. For example, in a light atom with a p^2 configuration, one wants to know to which extent the excited $^1S_0$ singlet mixes with the $^3P_0$ component from the lowest triplet. Such transformations between jj and LS coupled descriptions (of the same wave function) can be done using angular momentum algebra [16]. The developments of Grant resulted in the first general-purpose relativistic atomic structure program (GRASP) that has been developed further ever since [17].

Beyond Dirac-Coulomb

There are a number of effects that are not contained in the methods described so far. They include relativistic corrections to the electron interaction, the QED (also called "radiative") corrections mentioned before, and the effect of nuclei not being infinitely heavy point particles.

The Breit and Gaunt Interaction

There are relativistic corrections to the electron interaction, such as magnetic interactions and the effect of retardation, which are not contained in the Dirac-Coulomb Hamiltonian. However, no closed-form configuration space Hamiltonian

exists for a fully relativistic description of the electron interaction. Starting from QED, one can derive the (frequency-independent) Breit interaction as a lowest-order approximation (single virtual photon exchange, low velocities):

$$b(i,j) = -\frac{\boldsymbol{\alpha}(i)\cdot\boldsymbol{\alpha}(j)}{r_{ij}} + \frac{1}{2}\left(\frac{\boldsymbol{\alpha}(i)\cdot\boldsymbol{\alpha}(j)}{r_{ij}} - \frac{(\boldsymbol{\alpha}(i)\cdot\mathbf{r}_{ij})(\boldsymbol{\alpha}(j)\cdot\mathbf{r}_{ij})}{r_{ij}^3}\right) \tag{4}$$

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} \tag{5}$$

such that the Dirac-Coulomb-Breit Hamiltonian reads

$$\hat{H}_{DCB} = \sum_{i=1}^{n} \hat{h}_D(i) + \sum_{i<j} \left\{ \frac{1}{r_{ij}} + b(i,j) \right\} \tag{6}$$

The first term in the Breit operator describes the instantaneous magnetic interaction (Gaunt term); the second one stems from the retardation of the Coulomb interaction. The retardation term leads to greater numerical complexity and is sometimes neglected (then only the Gaunt term is used).

From early on, it has been argued that the Breit term should not be included in the iterative self-consistent procedure, but rather only its expectation value should be added to the Dirac-Coulomb result after convergence has been achieved. In numerical atomic Dirac-Fock programs, it is all but trivial to include the Breit interaction self-consistently [18]. This is (technically) easier in algebraic (basis set expansion) methods (see below). With such methods, to include the Breit interaction self-consistently [19] has been tried and the results suggested that this is a valid procedure at the no-pair level, in contrast to what had been assumed for 30 years. Note that this statement only refers to relativistic Hartree-Fock calculations. What happens if one includes the Breit interaction in a variational procedure that goes beyond Dirac-Fock (or MCDF) seems still to be under debate.

In high-Z, few-electron ions, it is certainly mandatory to include the Breit interaction to obtain accurate results. More relevant is the question whether one needs to include the Breit interaction in the calculation of valence properties of (nearly) neutral atoms or molecules. It is often said that the effect of the Breit interaction in this case is $\sim$1 % of the effects due to relativistic kinematics. However, the spin-other-orbit interaction (which is part of the magnetic interaction) is important for spin-orbit splittings and spin-orbit-dependent properties such as zero field splittings. The spin-dipolar interaction, which is the spin-spin part of the magnetic interaction, can be important especially in light elements. For example, it dominates the zero field splitting in small organic biradicals [20]. Of course, it is the Breit (or rather, the Gaunt) term that is generally responsible for long-range magnetic interactions and thus is important for macroscopic magnetic phenomena.

Nuclear Volume and Recoil Corrections

Nuclei are not point-like objects, but have finite dimensions, such that the electric field generated by a nucleus nowhere goes to infinity. The difference between calculations with a point-like and finite nucleus significantly increases when going from a nonrelativistic to a relativistic description, because in the latter, there is a weak singularity of the wave functions if there is a point-like nucleus. For point-like nuclei, this even implies that there are no Dirac solutions for $Z \geq c \approx 137$. To include finite nuclear size effects in atomic and molecular calculations, one needs a simplified model for the nuclear charge distribution. Initially a uniformly charged sphere with some nuclear radius has been assumed, but a Fermi-type charge distribution is probably much more realistic [21]. In molecular basis set expansion methods, a Gaussian nuclear charge model is still mostly used. Whereas this model is certainly highly unphysical, it is rather popular because molecular integrals of the resulting nuclear potential with Gaussian basis functions can very easily be calculated. For the nuclear radius, there exists an empirical dependence on the number of nucleons [22]. If one considers some electronic property for two different isotopes, the nucleus of the heavier one will have its nuclear charge density more spread out (larger nuclear radius), and this effect will be captured even by a crude nuclear model. It goes without saying that the chosen nuclear model strongly affects the $r \to 0$ behavior of the atomic orbitals. In heavy single-electron ions such as U^{91+}, the finite nuclear size effect on the Lamb shift is as important as the QED corrections [8].

Much more difficult is to account for the finite nuclear mass. In nonrelativistic theory, the center-of-mass motion can be separated off exactly. For one-electron ions, this simply leads to using a reduced mass for the electron, for many-electron atoms mass polarization terms arise which are the origin of the Hughes-Eckart shifts. The relativistic nuclear recoil theory is much more complicated, but the nonrelativistic terms are retrieved to the leading order [23].

Basis Set Expansion Methods

In atomic self-consistent field calculations, the spinor optimization leads to the numerical solution of radial variational differential equations. This separation of spin/angular and radial variable can be taken over to some extent for diatomic molecules. For polyatomic molecules, the algebraic (basis set) expansion is the preferred method to solve a variational problem. This has first been done, still for the atomic case, by Kim [24]. The basis set expansion leads to the relativistic analogue of the Roothaan equation. Atomic basis set expansion Dirac-Fock codes have been further developed [19, 25] and were reasonably successful, but did not substitute numerical atomic codes. For polyatomic molecules, the situation is dramatically different: here basis set expansion methods are probably the best one can do. The reason for this is that in quantum chemical applications, one is rather interested in

energy differences than in total energies, and basis set expansion methods have some sort of built-in error compensation, at least if one uses a fixed basis set for a given problem.

Relativistic Hartree-Fock Matrix Equations

In the relativistic case, one chooses separate basis functions L_μ and S_μ for the upper (large) and lower (small) components of the Dirac spinors

$$L_\mu(\mathbf{r}) = \begin{pmatrix} L_\mu^\alpha(\mathbf{r}) \\ L_\mu^\beta(\mathbf{r}) \\ 0 \\ 0 \end{pmatrix}, \quad S_\mu(\mathbf{r}) = \begin{pmatrix} 0 \\ 0 \\ S_\mu^\alpha(\mathbf{r}) \\ S_\mu^\beta(\mathbf{r}) \end{pmatrix}. \tag{7}$$

For simplicity, the number N of large and small-component basis functions are assumed to match. Basis sets of this general kind are called spinor basis sets. If all basis functions have only a single nonzero component (such basis sets are called scalar basis sets), the number of large and small-component basis functions will be different because of the kinetic balance condition (see below and the following chapter in this section). All basis sets are constructed such that applying the time-reversal operator on a basis function produces a function that can also be expanded in the basis set. Usually, the basis functions are normalized, since this improves numerical stability. Of course, we assume that the basis functions are linearly independent. Then, $2N$ orthonormal Dirac spinors ψ_p can be constructed

$$\psi_p = \sum_{\mu=1}^{N} C_{\mu p}^L L_\mu + C_{\mu p}^S S_\mu \tag{8}$$

with two $N \times 2N$ matrices $\mathbf{C}^L$ and $\mathbf{C}^S$ containing the spinor ("MO") coefficients. If there are n electrons, one chooses n occupied Dirac spinors from the $2N$ linear combinations and forms a Slater determinant. Then, one requires that the energy expectation value with the Dirac-Coulomb operator is stationary with respect to a variation of the MO coefficients under the constraint that the Dirac spinors remain orthonormal. This leads, in full analogy to the nonrelativistic case, to an iterative process, and in each iteration, a generalized (with overlap) eigenvalue problem of the form

$$\begin{pmatrix} \mathbf{F}^{LL} & \mathbf{F}^{LS} \\ \mathbf{F}^{SL} & \mathbf{F}^{SS} \end{pmatrix} \begin{pmatrix} \mathbf{C}^L \\ \mathbf{C}^S \end{pmatrix} = \begin{pmatrix} \mathbf{M}^L & 0 \\ 0 & \mathbf{M}^S \end{pmatrix} \begin{pmatrix} \mathbf{C}^L \\ \mathbf{C}^S \end{pmatrix} \begin{pmatrix} \varepsilon^+ & 0 \\ 0 & \varepsilon^- \end{pmatrix} \tag{9}$$

has to be solved. The submatrices $\mathbf{F}^{XY}$, $\mathbf{M}^Y$, etc. (X, Y is L or S) are $N \times N$ matrices, ε^+ is a diagonal matrix with N positive-energy eigenvalues, and ε^- is a

diagonal matrix with N negative-energy eigenvalues. The overlap submatrices (we have chosen the letter M for *metric* instead of the more conventional S to avoid confusion with the small-component basis functions) are

$$M^L_{\mu\nu} = < L_\mu | L_\nu >, \quad M^S_{\mu\nu} = < S_\mu | S_\nu >, \tag{10}$$

and the Fock matrices contain one-particle ($\mathbf{V}$, $\mathbf{T}$, and $\mathbf{M}$), Hartree ($\mathbf{J}$) and exchange ($\mathbf{K}$) terms

$$\mathbf{F}^{LL} = \mathbf{V}^L + \mathbf{J}^{LL} - \mathbf{K}^{LL} \tag{11}$$

$$\mathbf{F}^{SS} = \mathbf{V}^S - 2c^2\mathbf{M}^S + \mathbf{J}^{SS} - \mathbf{K}^{SS} \tag{12}$$

$$\mathbf{F}^{LS} = \mathbf{T} - \mathbf{K}^{LS} \tag{13}$$

$$\mathbf{F}^{SL} = \mathbf{T}^\dagger - \mathbf{K}^{SL} = \mathbf{T}^\dagger - \left(\mathbf{K}^{LS}\right)^\dagger = (\mathbf{F}^{LS})^\dagger \tag{14}$$

with matrix elements

$$V^L_{\mu\nu} = < L_\mu | V_N L_\nu >, \tag{15}$$

$$V^S_{\mu\nu} = < S_\mu | V_N S_\nu >, \tag{16}$$

$$T_{\mu\nu} = < L_\mu | c\boldsymbol{\sigma} \cdot \mathbf{p} S_\nu >, \tag{17}$$

$$J^{LL}_{\mu\nu} = \sum_{\rho,\tau} D^{LL}_{\tau\rho}(L_\mu L_\nu | L_\rho L_\tau) + D^{SS}_{\rho\tau}(L_\mu L_\nu | S_\rho S_\tau), \tag{18}$$

$$J^{SS}_{\mu\nu} = \sum_{\rho,\tau} D^{LL}_{\tau\rho}(S_\mu S_\nu | L_\rho L_\tau) + D^{SS}_{\rho\tau}(S_\mu S_\nu | S_\rho S_\tau), \tag{19}$$

$$K^{LL}_{\mu\nu} = \sum_{\rho,\tau} D^{LL}_{\tau\rho}(L_\mu L_\tau | L_\rho L_\nu), \tag{20}$$

$$K^{SS}_{\mu\nu} = \sum_{\rho,\tau} D^{SS}_{\tau\rho}(S_\mu S_\tau | S_\rho S_\nu), \tag{21}$$

$$K^{LS}_{\mu\nu} = \sum_{\rho,\tau} D^{LS}_{\tau\rho}(L_\mu L_\tau | S_\rho S_\nu), \tag{22}$$

$$K^{SL}_{\mu\nu} = \sum_{\rho,\tau} D^{SL}_{\tau\rho}(S_\mu S_\tau | L_\rho L_\nu) = \left(K^{LS}_{\nu\mu}\right)^\star. \tag{23}$$

The $\mathbf{J}$ and $\mathbf{K}$ matrices are defined through the two-electron integrals $(\bullet\,\bullet \mid \bullet\,\bullet)$ (in Mulliken notation) and the density matrices are defined as

$$D^{XY}_{\mu\nu} = \sum_i C^X_{\mu i} \left(C^Y_{\nu i}\right)^\star = \left(D^{YX}_{\nu\mu}\right)^\star \tag{24}$$

where the summation index i goes over the occupied molecular spinors. With these definitions, the Fock matrices can be constructed. Note that only two-electron integrals of type $(LL|LL)$, $(LL|SS)/(SS|LL)$, and $(SS|SS)$ occur in the case of the Dirac-Coulomb operator. Including the Breit interaction leads to rather complicated two-electron integrals, and sometimes the Gaunt term (without retardation) is used, since it covers the most important (spin-other-orbit and spin-dipolar) corrections. This is so because the retardation only alters the orbit-orbit term. In each iteration, the diagonalization of the Fock matrix, Eq. (9) produces $2N$ Dirac spinors, many more than n, as needed to construct the Slater determinant. This is in sharp contrast to fully numerical methods, which solve differential equations and produce only occupied Dirac spinors. So after diagonalization of the Fock matrix, one has to decide which of the solutions become occupied in the next iteration. Only positive-energy spinors are considered (this implements the implicit positive-energy projection); among those, one often chooses the n spinors of lowest orbital energy, but it is also possible to choose those with maximum overlap to the occupied spinors of the preceding iteration. The energy expectation value of the Slater determinant with the Dirac-Coulomb Hamiltonian is given by

$$E = E_{1e} + E_{2e} \tag{25}$$

$$E_{1e} = \mathrm{Tr}(\mathbf{D}^{LL} \cdot \mathbf{V}^{LL}) + \mathrm{Tr}(\mathbf{D}^{SS} \cdot \mathbf{V}^{SS}) - 2c^2\mathrm{Tr}(\mathbf{D}^{SS} \cdot \mathbf{M}^{S}) + \mathrm{Tr}(\mathbf{D}^{LS} \cdot \mathbf{T}) + \mathrm{Tr}(\mathbf{D}^{SL} \cdot \mathbf{T}^{\dagger}) \tag{26}$$

$$E_{2e} = \frac{1}{2}\Big\{\mathrm{Tr}(\mathbf{D}^{LL} \cdot (\mathbf{J}^{LL} - \mathbf{K}^{LL})) + \mathrm{Tr}(\mathbf{D}^{SS} \cdot (\mathbf{J}^{SS} - \mathbf{K}^{SS})) - \mathrm{Tr}(\mathbf{D}^{SL} \cdot \mathbf{K}^{LS}) - \mathrm{Tr}(\mathbf{D}^{LS} \cdot \mathbf{K}^{SL})\Big\}. \tag{27}$$

In this equation, E_{1e} is the one-particle energy and E_{2e} the electron interaction, and $\mathrm{Tr}(\bullet)$ denotes the trace of a matrix. Note that the last two terms both of E_{1e} and E_{2e} form a pair of complex conjugates.

The basis set expansion method for relativistic Hartree-Fock has been developed by Kim [24] for closed-shell atoms and extended by Kagawa [26] to the open-shell case. In these atomic cases, the basis set expansion is applied to the radial part of the Dirac spinors only, while the angular and spin parts are taken from the known central field solutions. A general molecular basis set expansion was already drafted by Synek [27] and later by Malli [28]. Early numerical results for molecules were presented by Matsuoka [29] and Mark [30]. The procedure essentially parallels the nonrelativistic case, but the matrix dimension is larger and the matrices are complex valued, such that the memory demand is considerably larger. A significant increase in computer time comes from the additional two-electron integrals, but also matrix operations (such as the diagonalization) become much more costly because of their N^3 scaling.

The "Diseases" and How to Become Immune

Already in the first applications of relativistic basis set expansion methods to molecules [29], it became obvious that the energy differences between a nonrelativistic and relativistic calculation were too large and that one needs very large basis sets to stabilize this energy difference. Even for basis sets that were already large, the calculated energy showed large oscillations when adding new basis functions, and the calculated energy when performing the limit $c \to \infty$ was lower than the nonrelativistic result. Even worse, it could happen that occupied Dirac spinors were chosen, according to the *low positive-energy eigenvalue* criterion, which were completely unphysical, as indicated by their near-zero kinetic energy expectation value [31]. Initially, these failures were associated with the so-called *Brown-Ravenhall disease*, which has been advocated as a major problem by Sucher [32]. The problem is present for many-electron systems already in the absence of electron interaction: Choosing a Dirac spinor from the negative and one from the positive-energy continuum, one can construct many two-particle Slater determinants which are eigenfunctions of $\hat{H}_{DC}$ in the absence of electron interaction for *any* desired eigenvalue. Sucher's recipe was to put projection operators around the Dirac-Coulomb Hamiltonian such that any spinor in a determinant is projected to some positive-energy space, either defined by the positive-energy solutions of a free particle or by those of a single electron moving in the electric field of the nuclear framework. The resulting *no-pair* Hamiltonian has become very important in the relativistic correlation problem, but the projection is already implicitly present in the relativistic Hartree-Fock procedure when picking only positive-energy solutions of the eigenvalue problem Eq. (9) as occupied orbitals. Sucher discussed two possible scenarios for the positive-energy projection: either the potential used to define the positive-energy space is zero (*free particle projection*) or the potential of the nuclear framework (*external field projection*). The projection implicit in the relativistic Hartree-Fock procedure uses a third option, namely, the actual effective potential of the current iteration. This approach is equally valid [33]. The viewpoint today [34] is that the Dirac-Coulomb operator can only be used in Fock space (and not in configuration space) and that the Brown-Ravenhall disease is an artifact of using an illegitimate Hamiltonian. The no-pair projection is then viewed as a simplification of the Fock-space formalism. Inherent to the Fock-space formalism is that the vacuum state of the Hamiltonian depends on the positive and negative-energy spinor spaces and thus changes from one to the next self-consistent field iteration. This also implies that the Fock-space Dirac-Coulomb Hamiltonian to be used, e.g., in a post-Hartree-Fock correlation treatment, depends on the reference state of the Hartree-Fock calculation. These implications are more relevant to the correlation treatment than to the self-consistent field calculation and thus outside the scope of this chapter.

Here it is sufficient to remind the reader that the positive-energy projection that is performed automatically in the relativistic Hartree-Fock procedure makes it immune against the Brown-Ravenhall disease. The actual problems observed in numerical

calculations have a different origin, as first explicitly pointed out by Schwarz [35], and are caused by the basis set expansion itself (*finite basis set disease*). That this problem is different becomes quite obvious if one realizes that it already occurs in one-electron systems. Consider, e.g., an atomic calculation on a hydrogenic ion, where all basis functions L_μ *and* S_μ are of s type. Then, the matrix $\mathbf{T}$ (see Eqs. (17) and (26)) will be zero, as well as the kinetic energy contribution to the one-particle energy. Any Dirac spinor expanded in this basis will be a bad approximation to the hydrogenic 1s orbital and a mixture of positive and negative-energy solutions. There is a recipe, named *kinetic balance*, to construct relativistic basis sets [31, 36, 37]. In what is usually called *restricted kinetic balance*, there is a 1:1 correspondence between large and small-component basis functions such that

$$S_\mu = t_\mu \begin{pmatrix} 0 & 0 \\ \boldsymbol{\sigma} \cdot \mathbf{p} & 0 \end{pmatrix} L_\mu, \tag{28}$$

where the factor t_μ is used to make S_μ normalized. For all but the heaviest elements, one can choose a scalar large-component basis set, but the restricted kinetic balance recipe requires a spinor basis set for the small component. If one uses heavily contracted Gaussian basis functions for atomic core orbitals, an extension of the concept called *atomic balance* has been advocated [38]. In practice this means that one starts with an uncontracted kinetically balanced basis set and obtains the contraction coefficients of the contracted large and small-component basis functions from the atomic spinor coefficients instead of imposing Eq. (28) for contracted functions. Kinetic balance is discussed in detail in the following chapter; at this place it is sufficient to remark that kinetic balance is a very effective cure of the finite basis set disease.

To summarize, relativistic Hartree-Fock is immune against the Brown-Ravenhall disease by construction, and the finite basis set disease does not harm if kinetically balanced basis sets are used. As an aside, no such "diseases" have infected numerical atomic relativistic Hartree-Fock calculations: because the proper boundary conditions are imposed and the variational differential equations are solved exactly (within numerical accuracy), a proper projection to the positive-energy states is implicitly present, and the numerical solutions automatically exhibit kinetic balance.

New Developments

If one knows how to choose the basis set, all fundamental problems – within the "no-pair" approximation – of Dirac-Hartree-Fock are solved. There have been tremendous advances since the end of the 1980s, but these were mostly technical and targeted at increasing the computational efficiency. This will not be discussed in detail in this overview. A significant enhancement of the efficiency came from new two-electron integral programs, especially those which are tailored to the way relativistic basis sets are constructed (e.g., Ref. [39]). In fact, one of the

major drawbacks of early molecular Dirac-Fock programs came from the fact that they used an integral engine developed for nonrelativistic applications. Another important observation was that the density arising from the small-component basis functions is rather compact, nearly spherical, and centered around the nuclei. Visscher has shown [40] that one can discard the $(SS|SS)$ two-electron integrals involving four small-component basis functions altogether from the calculation if a simple point-charge correction is applied. In routine applications performed today, $(SS|SS)$ integrals are rarely calculated and processed. This speeds up the calculation especially if scalar basis sets are used for the large components (either L_μ^α or L_μ^β are zero) but spinor basis sets for the small components (both S_μ^α and S_μ^β are nonzero), which is the case for kinetically balanced basis sets in which the number of large and small-component basis functions match. Using methods known from nonrelativistic treatments, two-electron integrals can be avoided (or approximated) using pseudospectral [41] or density fitting [42] procedures. Current developments focus on the calculation of molecular properties at Dirac-Fock level rather than on how to calculate the Dirac-Fock energy itself.

Relativistic Kohn-Sham

It is known from the nonrelativistic case that the Kohn-Sham variant of density functional theory is operationally very similar to Hartree-Fock. This largely also holds in the relativistic case. The iterative procedure is very similar: one constructs an effective potential (that depends on the occupied Dirac spinors) in each iteration, diagonalizes a Fock-type matrix, and iterates until self-consistency. Although the computational steps are similar, the theoretical foundation is quite different.

Relativistic Density Functional Theory Basics

A relativistic generalization of the Hohenberg-Kohn theorem has first been presented by Rajagopal and Callaway [43]. The four-current, in other words both the charge density $\rho(\mathbf{r})$ and the charge current density $\mathbf{j}(\mathbf{r})$, serves as the basic variable. This first presentation was formulated in a field theoretical language, but it was not taken into account that certain quantities that occur in such a formulation may become infinite and therefore require proper renormalization. The nontrivial point here is that one has to make sure that the renormalization procedure (and the counterterms to the four-current that arise) does not destroy the one-to-one mapping of external potentials to ground state densities needed for the Hohenberg-Kohn logic. With quite some effort, one can show that the relativistic Hohenberg-Kohn theorem holds [44]. A relativistic DFT procedure rooted in this quantum field theoretical framework is however so complicated that all existing computational schemes involve the no-pair approximation discussed at the beginning of this chapter.

The Kohn-Sham energy in the no-pair approximation is derived from a Slater determinant-built positive-energy spinors; the charge and charge current densities are defined as

$$\rho(\mathbf{r}) = \sum_i \phi_i^\dagger(\mathbf{r})\phi_i(\mathbf{r}), \quad \mathbf{j}(\mathbf{r}) = c\sum_i \phi_i^\dagger(\mathbf{r})\boldsymbol{\alpha}\phi_i(\mathbf{r}) \tag{29}$$

where the ϕ_i are the spinors of the Slater determinant. The kinetic energy of the reference (Kohn-Sham) determinant is

$$T_S = \sum_i < \phi_i \Big| \begin{pmatrix} 0 & c\boldsymbol{\sigma}\mathbf{p} \\ c\boldsymbol{\sigma}\mathbf{p} & 0 \end{pmatrix} \phi_i > \tag{30}$$

and the Hartree energy reads

$$E_H = \int \left\{ \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} - \frac{1}{c^2}\frac{\mathbf{j}(\mathbf{r}_1)\mathbf{j}(\mathbf{r}_2)}{r_{12}} \right\} \mathrm{d}\mathbf{r} \tag{31}$$

All other contributions to the energy are cast into the exchange-correlation energy, just as in the nonrelativistic case. The last (current-current) term of the Hartree energy represents a magnetic interaction and is often neglected, since closed shells (including the atomic core shells) do not contribute to $\mathbf{j}$.

Making the energy stationary with respect to orbital variations generates an eigenvalue equation similar to Eq. (9). The Hartree terms ($\mathbf{J}$ matrices) remain the same, but the exchange terms (involving the $\mathbf{K}$ matrices) are redefined according to

$$-K_{\mu\nu}^{LL} = < L_\mu | V_{xc} L_\nu >, \tag{32}$$

$$-K_{\mu\nu}^{SS} = < S_\mu | V_{xc} S_\nu >, \tag{33}$$

$$-K_{\mu\nu}^{LS} = < L_\mu | c\boldsymbol{\alpha}\cdot\mathbf{A}_{xc} S_\nu >, \tag{34}$$

$$-K_{\mu\nu}^{SL} = < S_\mu | c\boldsymbol{\alpha}\cdot\mathbf{A}_{xc} L_\nu > = -(K_{\nu\mu}^{LS})^\star. \tag{35}$$

involving the scalar and vector exchange-correlation potentials $V_{xc}(\mathbf{r})$ and $\mathbf{A}_{xc}(\mathbf{r})$ that result from a variation of the exchange-correlation energy w.r.t. $\rho(\mathbf{r})$ and $\mathbf{j}(\mathbf{r})$. If one includes the Breit interaction, there is an additional Hartree-type term which is included by augmenting $\mathbf{A}_{xc}$ by

$$\mathbf{A}_H(\mathbf{r}) = -\frac{1}{c^2}\int \frac{\mathbf{j}(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|}\, d\mathbf{s} \tag{36}$$

which directly follows from the Gaunt term (there is no retardation in the Hartree part). If the external four-potential is generated by a fixed nuclear framework, it only has a scalar time-independent part. This is called the *electrostatic limit*, for which the relativistic Hohenberg-Kohn theorem states that the density alone is sufficient to

determine the external potential. In this case, there exists an exchange-correlation functional that only depends on the density, and $\mathbf{A}_{xc}$ vanishes.

Relativistic Spin-Density Functional Procedures

Even in the electrostatic limit, there are internal magnetic fields in the molecule unless it is a closed-shell system. Although the fundamental theorem tells us the density is good enough, it is not possible to find a good enough exchange-correlation functional for open-shell systems. This is already known from nonrelativistic theory, where one needs a functional that also depends on the spin density to describe open-shell systems, no matter whether there are external magnetic fields or not. Unfortunately, there is, until today, no generally applicable relativistic exchange-correlation functional that depends on the four-current. As a resort, one uses spin-density functionals known from the nonrelativistic case in relativistic Dirac-Kohn-Sham calculations. To this end, one performs a Gordon decomposition of the charge current density into an orbital and a spin part and then defines an exchange-correlation energy that depends only on the latter. For details, see the chapter Ref. [45]. Besides the charge density, the *magnetization density* $\mathbf{m}$ becomes the basic variable

$$\mathbf{m}(\mathbf{r}) = \sum_i \phi_i^\dagger(\mathbf{r}) \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & -\boldsymbol{\sigma} \end{pmatrix} \phi_i(\mathbf{r}), \tag{37}$$

$\mathbf{K}^{LS}$ and $\mathbf{K}^{SL}$ disappear, and V_{xc} is augmented by a spinor potential involving the functional derivative of the exchange-correlation energy w.r.t. $\mathbf{m}(\mathbf{r})$. Note that the definition used in Eq. (37) follows from the Gordon decomposition; it seems that sometimes a slightly different definition is used in which the Pauli $\boldsymbol{\sigma}$ matrix in the lower right corner does not carry a minus sign.

Relativistic Functionals

It is known that in a (noninteracting) homogeneous electron gas, relativistic effects grow with the density. The critical density at which they become discernible is exceeded in atomic cores. Therefore, one should expect relativistic corrections to exchange-correlation functionals to be important. The personal view of the present author, however, is that no useful relativistic functional has emerged although much effort has been spent. One reason is that too little is known about the current dependent parts of the functional. In most cases, nonrelativistic functionals are used today. So-called hybrid functionals, which mix density functional and Fock exchange, are also widely used in the relativistic case. Magnetic interactions (Gaunt or Breit) are included in an "100% exact exchange" manner, that is, the expectation value of the Kohn-Sham reference determinant with the two-particle Gaunt or Breit

operator defines the contribution to the total energy. This is relevant if one is interested in the spin-other-orbit contribution to the two-electron spin-orbit terms, or if one wants to include the spin-dipolar interaction. Today, relativistic Kohn-Sham is theoretically well developed, but in practice it is mostly a combination of relativistic kinematics with nonrelativistic density functional methods. The situation is not so bad as it seems: in several studies, relativistic corrections to the exchange-correlation functional showed no large effect on spectroscopic constants [46, 47]. But it has to be kept in mind that these investigations were limited to closed-shell systems and relativistic corrections to a functional depending on the density only.

New Developments

As already mentioned for Dirac-Fock, the largest impact on advancing the Dirac-Kohn-Sham method came from technical improvements that increase computational efficiency. It is clear that many of the performance improvements achieved in Dirac-Fock programs immediately transfer to the Dirac-Kohn-Sham case, especially if hybrid functionals are used. Ref. [48] gives an overview over Dirac-Kohn-Sham programs used today and discusses recent advances.

Beyond Self-Consistent Field Methods

Relativistic Kohn-Sham is a widely used method to target problems of heavy-element chemistry. Its quality and applicability mainly depend on the chosen exchange-correlation functional. The hierarchy of functionals is essentially the same as in the nonrelativistic case: local density functionals are widely considered obsolete today, gradient-corrected functionals are frequently used, but also hybrid functionals which contain a portion of Fock exchange. The so-called τ-dependent functionals (also called *meta GGAs*), which contain the kinetic energy density as an additional variable, have not yet found much use in relativistic calculations. However, the Dirac-Kohn-Sham methods cannot be systematically improved (other than replacing one functional with another). This is the main difference to relativistic Hartree-Fock, which is usually viewed as a necessary (and boring) first step, the post-Hartree-Fock or correlation treatment being the real issue. Because Sucher's warnings about the Dirac-Coulomb Hamiltonian [32] were highly influential, the established relativistic electron correlation methods all start with a Dirac-Fock calculation, after which the negative-energy Dirac spinors are discarded. This implements an implicit no-pair projection defined through the self-consistent (mean) field. Using the positive-energy spinors only, the correlation treatment then proceeds in full analogy to the nonrelativistic case; this defines the relativistic configuration interaction, many-body perturbation theory, or coupled cluster methods. Note that only the conventional methods using a finite number of spinors have a relativistic analogue; explicitly correlated ("r_{12}") methods formally involve an extension to a complete (infinitely large) set, such that relativistic r_{12} methods need to go beyond

the no-pair approximation [10]. The difference of the finite-dimensional methods to the nonrelativistic case is mostly technical: the four-component structure results in a much more demanding transformation of the two-electron integrals to the molecular spinor basis set, and there are less savings due to time-reversal symmetry in the relativistic case as there are savings due to spin symmetry in nonrelativistic calculations. In practice, however, the major obstacle stems from the fact that the electron correlation problem itself is much more demanding in applications to heavy-element chemistry, because heavy metals often have multiple open shells, and the number of strongly interacting electrons that must be included in the correlation treatment is larger.

The how and why of including negative-energy spinors in the correlation treatment is less clear. Continuum dissolution can be avoided by methods that extract narrow resonances out of the Brown-Ravenhall continuum [49], but even if one simply performs a configuration interaction calculation, including the negative-energy states, within a finite basis set, one may not see continuum dissolution and get (stable) numerical results [50]. A consistent treatment of the relativistic electron correlation problem beyond the no-pair approximation starts from a Fock-space formulation [34] but has to go beyond [10, 11]. This is a current research topic and clearly beyond the scope of this introductory chapter on the relativistic treatment of many-electron systems.

Summary

Using a mean-field approach, relativistic self-consistent field methods allow for an approximate treatment of many-electron systems that only involves the solution of one-particle Dirac equations in some effective potential. Formally, the relativistic wave function is a Slater determinant built from positive-energy "occupied" spinors which are varied as to minimize the resulting energy expression. In atomic calculations, the correct boundary conditions as obtained for (numerically) exact solutions of the radial Dirac equations implicitly keep the spinors within the positive-energy space; in atomic or molecular basis set expansion methods, this can approximately be achieved through the kinetic balance condition. Then, relativistic self-consistent methods are immune against the *finite basis set disease* that were so virulent already in the earliest relativistic basis set expansion calculations. The existence of negative-energy spinors is not problematic in relativistic self-consistent field methods, as there is an implicit projection to the positive-energy space: these methods are also immune against the so-called *Brown-Ravenhall disease*. Since these fundamental issues have been solved, relativistic self-consistent field methods have advanced mostly on the technical side by introducing better algorithms that lead to higher computational efficiency. Today, relativistic Dirac-Fock is the starting point for the much more involved relativistic correlation treatment, and relativistic Kohn-Sham is a routine method for obtaining approximate results for fairly large systems containing heavy elements.

References

1. Swirles B (1935) Relativistic self-consistent field. Proc R Soc Ser A 152:625
2. Williams AO (1940) A relativistic self-consistent field for Cu. Phys Rev 58:723
3. Liberman D, Waber JT, Cromer DT (1965) Self-consistent-field Dirac-Slater wave functions for atoms and ions. I. Comparison with previous calculations. Phys Rev 137:A27
4. Grant IP (1961) Relativistic self-consistent fields. Proc R Soc Ser A 262:555
5. Grant IP (1970) Relativistic calculation of atomic structures. Adv Phys 19:747
6. Desclaux JP (1973) Relativistic Dirac-Fock expectation values for atoms with atomic numbers Z = 1–120. At Data Nucl Data Tables 12:311
7. Grant IP, Quiney HM (1987) Foundations of the relativistic theory of atomic and molecular-structure. Adv At Mol Opt Phys 23:37
8. Beier T, Mohr PJ, Persson H, Soff G (1998) Influence of nuclear size on QED corrections in hydrogenlike heavy ions. Phys Rev A 58:954
9. Pyykko P, Dyall KG, Csazar AG, Tarczay G, Polyansky OL, Tennyson J (2001) Estimation of Lamb-shift effects for molecules: application to the rotation-vibration spectra of water. Phys Rev A 63:024502
10. Liu WJ (2012) Perspectives of relativistic quantum chemistry: the negative energy cat smiles. Phys Chem Chem Phys 14:35
11. Liu W, Lindgren I (2013) Going beyond "no-pair relativistic quantum chemistry". J Chem Phys 139:014108
12. Aucar GA (2014) Toward a QFT-based theory of atomic and molecular properties. Phys Chem Chem Phys 16:4420
13. Desclaux JP (1975) Multiconfiguration relativistic Dirac-Fock program. Comput Phys Commun 9:31
14. Mckenzie BJ, Grant IP, Norrington PH (1980) A program to calculate transverse Breit and QED corrections to energy-levels in a multiconfiguration Dirac-Fock environment. Comput Phys Commun 21:233
15. Froese Fischer C (1978) General multi-configuration Hartree-Fock program. Comput Phys Commun 14:145
16. Dyall KG (1986) Transform – a program to calculate transformations between various JJ and LS coupling schemes. Comput Phys Commun 39:141
17. Jonsson P, Gaigalas G, Bieron J, Froese Fischer C, Grant IP (2013) New version: GRASP2K relativistic atomic structure package. Comput Phys Commun 184:2197
18. Reiher M, Hinze J (1999) Self-consistent treatment of the frequency-independent Breit interaction in Dirac-Fock and MCSCF calculations of atomic structures: I. Theoretical considerations. J Phys B-At Mol Opt Phys 32:5489
19. Quiney HM, Grant IP, Wilson S (1987) The Dirac-equation in the algebraic-approximation. 5. Self–consistent field studies including the Breit interaction. J Phys B-At Mol Opt Phys 20:1413
20. Havlas Z, Michl J (1999) Ab initio calculation of zero-field splitting and spin-orbit coupling in ground and excited triplets of m-xylylene. J Chem Soc Perkin Trans 2 1999:2299
21. Parpia FA, Mohanty AK (1992) Relativistic basis-set calculations for atoms with Fermi nuclei. Phys Rev A 46:3735
22. Visscher L, Dyall KG (1997) Dirac-Fock atomic electronic structure calculations using different nuclear charge distributions. At Data Nucl Data Tables 67:207
23. Tupitsyn II, Shabaev VM, Lopez-Urrutia JRC, Draganic I, Orts RS, Ullrich J (2003) Relativistic calculations of isotope shifts in highly charged ions. Phys Rev A 68:022511
24. Kim YK (1967) Relativistic self-consistent-field theory for closed-shell atoms. Phys Rev 154:17
25. Kagawa T (1980) Multiconfiguration relativistic Hartree-Fock-Roothaan theory for atomic systems. Phys Rev A 22:2340
26. Kagawa T (1975) Relativistic Hartree-Fock-Roothaan theory for open-shell atoms. Phys Rev A 12:2245

27. Synek M (1964) Analytical relativistic self-consistent field theory. Phys Rev 136:A1552
28. Malli G, Oreg J (1975) Relativistic self-consistent-field (RSCF) theory for closed-shell molecules. J Chem Phys 63:830
29. Matsuoka O, Suzuki N, Aoyama T, Malli G (1980) Relativistic self-consistent-field methods for molecules. 1. Dirac-Fock multiconfiguration self-consistent-field theory for molecules and a single-determinant Dirac-Fock self-consistent-field method for closed-shell linear-molecules. J Chem Phys 73:1320
30. Mark F, Rosicky F (1980) Analytical relativistic Hartree-Fock equations within scalar basis-sets. Chem Phys Lett 74:562
31. Schwarz WHE, Wallmeier H (1982) Basis set expansions of relativistic molecular wave-equations. Mol Phys 46:1045
32. Sucher J (1987) Relativistic many-electron Hamiltonians. Phys Scr 36:271
33. Mittleman MH (1981) Theory of relativistic effects on atoms – configuration-space Hamiltonian. Phys Rev A 24:1167
34. Kutzelnigg W (2012) Solved and unsolved problems in relativistic quantum chemistry. Chem Phys 395:16
35. Schwarz WHE, Wechseltrakowski E (1982) The 2 problems connected with Dirac-Breit-Roothaan calculations. Chem Phys Lett 85:94
36. Stanton RE, Havriliak S (1984) Kinetic balance – a partial solution to the problem of variational safety in Dirac calculations. J Chem Phys 81:1910
37. Dyall KG, Grant IP, Wilson S (1984) Matrix representation of operator products. J Phys B-At Mol Opt Phys 17:493
38. Visscher L, Aerts PJC, Visser O, Nieuwpoort WC (1991) Kinetic balance in contracted basis-sets for relativistic calculations. Int J Quantum Chem Quantum Chem Symp 25:131
39. Yanai T, Nakajima T, Ishikawa Y, Hirao K (2001) A new computational scheme for the Dirac-Hartree-Fock method employing an efficient integral algorithm. J Chem Phys 114:6526
40. Visscher L (1997) Approximate molecular relativistic Dirac-Coulomb calculations using a simple Coulombic correction. Theor Chem Acc 98:68
41. Nakajima T, Hirao K (2004) Pseudospectral approach to relativistic molecular theory. J Chem Phys 121:3438
42. Belpassi L, Tarantelli F, Sgamellotti A, Quiney HM (2008) Poisson-transformed density fitting in relativistic four-component Dirac-Kohn-Sham theory. J Chem Phys 128:124108
43. Rajagopal AK, Callaway J (1973) Inhomogeneous electron-gas. Phys Rev B 7:1912
44. Engel E (2002) Relativistic density functional theory: foundations and basic formalism. In: Schwerdtfeger P (ed) Relativistic electronic structure theory. Part 1: Fundamentals. Elsevier, Amsterdam, p 523
45. van Wüllen C (2010) Relativistic density functional theory. In: Barysz M (ed) Relativistic methods for chemists. Springer, Dordrecht, p 191
46. Mayer M, Haeberlen OD, Roesch N (1996) Relevance of relativistic exchange-correlation functionals and of finite nuclei in molecular density-functional calculations. Phys Rev A 54:4775
47. Varga S, Engel E, Sepp WD, Fricke B (1999) Systematic study of the Ib diatomic molecules Cu2, Ag2, and Au2 using advanced relativistic density functionals. Phys Rev A 59:4288
48. Belpassi L, Storchi L, Quiney HM, Tarantelli F (2011) Recent advances and perspectives in four-component Dirac-Kohn-Sham calculations. Phys Chem Chem Phys 13:12368
49. Pestka G, Bylicki M, Karwowski J (2007) Complex coordinate rotation and relativistic Hylleraas-CI: helium isoelectronic series. J Phys B-At Mol Opt Phys 40:2249
50. Watanabe Y, Nakano H, Tatewaki H (2010) Effect of removing the no-virtual pair approximation on the correlation energy of the He isoelectronic sequence. II. Point nuclear charge model. J Chem Phys 132:124105

Part II

Introduction to Quantum Electrodynamics

Paul Indelicato

Introduction to Bound-State Quantum Electrodynamics

5

Paul Indelicato and Peter J. Mohr

Contents

Introduction to Quantum Electrodynamics . . . 132
The Dirac Equation for Bound States . . . 133
Dirac Equation for a Particle with a Landé Factor $g \neq 2$ . . . 135
Bound States of the Dirac Equation . . . 137
Other Useful Properties of the Dirac Equation in a Coulomb Field . . . 143
Operators Mean Value for the Dirac Equation in a Coulomb Field . . . 144
Problems Associated with the Predictions of the Dirac Equation . . . 146
Quantum Electrodynamics . . . 146
Introduction . . . 146
Representations . . . 147
Evolution Operators . . . 148
Fundamental Relations Between Operators for the Electron Field . . . 149
Fundamental Operator Relations for the Electromagnetic Field . . . 150
Perturbation Theory, S-Matrix, and Energy Shift for Bound States . . . 150
Wick's Theorem, Contractions, and Propagators . . . 153
Evaluation of $\left\langle S^{(1)}_{\epsilon,1} \right\rangle_c$ . . . 158
Evaluation of $\left\langle S^{(2)}_{\epsilon,1} \right\rangle_c$ . . . 159
Evaluation of the One-Loop Self-Energy . . . 170
Introduction . . . 170
Low-Energy Part . . . 173
Regularization . . . 175
Singular Terms: Renormalization in Coordinate Space . . . 176

P. Indelicato (✉)
Laboratoire Kastler Brossel, UPMC-Sorbonne Universités, Collège de France and CNRS, Paris, France
PSL Research University, Paris, France
e-mail: paul.indelicato@lkb.upmc.fr

P.J. Mohr
National Institute of Standards and Technology, Gaithersburg, MD, USA
e-mail: mohr@nist.gov

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_36

Evaluation of the Singular Terms . . . 177
Evaluation of the Subtraction Terms . . . 183
Results and Discussion . . . 189
Vacuum Polarization . . . 189
Regularization of the Vacuum Polarization . . . 190
Higher-Order Contributions . . . 199
Self-Energy Screening . . . 200
Perturbative Derivation of the Self-Energy Screening Correction . . . 202
Self-Energy Screening Low-Energy Part . . . 204
Self-Energy Screening High-Energy Part . . . 211
Self-Energy Screening: Analytic Terms and Verification of Global Renormalization . . . 216
Two-Photon, Two-Electron Diagrams . . . 219
Conclusion . . . 224
Appendix: Useful Properties . . . 228
Properties of Dirac γ Matrices . . . 228
Basic Relations . . . 228
More Properties of the Dirac Matrices . . . 228
Asymptotic Properties . . . 229
References . . . 231

Abstract

In this chapter we describe the fundamental aspects of bound-state quantum electrodynamics (BSQED). We recall the principal features of the Dirac equation. Then we describe quantum electrodynamics as a field theory. We provide the basic elements about representations, evolution operators and the S matrix. We then proceed to describe perturbation expansion of the S-matrix, and its relations with bound-state energies. We express this expansion in terms of Feynman diagrams. Finally we illustrate on practical examples the concepts of regularization, using the method of Pauli and Villars, and renormalization in coordinate space. We describe in detail the practical ways of doing the calculations, using self-energy, vacuum polarization, self-energy screening and QED corrections to the ladder approximation. Finally we show the quality of the agreement between BSQED and experiment by showing comparison for two- and three-electron ions transitions.

Keywords

Atoms • Bound State Quantum Electrodynamics • Dirac-Equation • Self-energy • Relativistic many-body • Highly charged ions • Quantum-field theory • Vacuum polarization

Introduction to Quantum Electrodynamics

The most general theory of electrons and photons in electromagnetic fields is called quantum electrodynamics (QED) (see, e.g., [1]). It combines special relativity and quantum mechanics. It describes at the same time bound states (with the acronym BSQED) and dynamical processes, like photon emission, and can then influence the transition probability (see, e.g., Ref. [2]). Here we will specialize in the BSQED

aspects. The starting point of QED is the Dirac equation [3], which combines relativistic invariance and relativity. It is described in section "The Dirac Equation for Bound States" together with the formulas and relations that are needed to perform practical QED calculations. Yet the Dirac equation cannot describe all the physics associated with special relativity, as the possibility of particle creation and annihilation are not accounted for. We will thus describe QED, the field theory that enables to treat particle creation and annihilation in section "Quantum Electrodynamics." We will then move to the perturbation theory that allows to do practical calculations in QED in section "Perturbation Theory, S-Matrix, and Energy Shift for Bound States." We will then treat the most important QED corrections to bound-state energies in one- and two-electron atoms in sections "Evaluation of the One-Loop Self-Energy" to "Two-Photon, Two-Electron Diagrams". We describe in this a formalism based on the evolution operator. In the next three chapters, a different formalism developed by the St Petersburg group, based on the two-time Greens function, will be described in details. In the last chapter, a formalism developed in the Göteborg group, which allows to combine BSQED and many-body technique, is described. Thanks to 40 years of intense work, it is now possible to calculate very accurately bound-state energies in hydrogen-like atoms from hydrogen to $Z = 110$ with all first and second order in the fine structure constant corrections [4]. Two-electron systems have also been studied [5], as well as lithium-like [6, 7] and beryllium-like ions [8–10] and both singly and doubly excited states. Finally QED calculations of the fine structure of the boron-like sequence have been evaluated recently [11]. For more electrons, calculations become more and more difficult. Combinations of many-body technique and QED corrections based on a variety of model are used. The self-energy screening in many-electron atoms has been calculated using method ranging from effective-Z based on the mean value of r for the orbital [12], on the Welton method [13, 14], or BSQED calculations based on the Hartree-Fock potential [15, 16] have been used. Recently a new method has been proposed [17, 18] to evaluate the BSQED contribution, which uses most available *ab initio* calculations to build model operators, allowing to evaluate screened self-energy contributions, including non-diagonal ones in many-body calculations.

The Dirac Equation for Bound States

In this section, we define the notations that are usually used in bound-state quantum electrodynamics (BSQED) for the Dirac equation, and we recall useful properties that are needed in practical QED calculations.

The Dirac equation for a free fermion of mass m, with a wave function $\phi(x)$, is given by

$$(\gamma^\mu p_\mu - mc)\phi(x) = 0 \tag{1}$$

where γ represent Dirac matrices, the properties of which are given in section "Properties of Dirac γ Matrices."

In an external electromagnetic field, given by the 4-vector $A^\mu = (\phi/c, \mathbf{A})$, one obtains the Dirac equation in an external field by replacing in (1) the 4-momentum p by $p - qA$ following the principle of *minimum coupling*. Here q is the charge of the particle (NB: q is the charge including the sign, e.g., for an electron $q = -e$ where $e = 1.6019 \times 10^{-19}\,C$.):

$$[\gamma^\mu(p_\mu - qA_\mu) - mc]\psi = 0. \tag{2}$$

Using the definition of γ and p, one obtains

$$\gamma^\mu p_\mu = \gamma^0 p_0 - \boldsymbol{\gamma}\mathbf{p} = i\hbar\gamma^0\frac{1}{c}\frac{\partial}{\partial t} + i\hbar\boldsymbol{\gamma}\boldsymbol{\nabla}. \tag{3}$$

The corresponding Hamiltonian-like equation is obtained by multiplying

$$\left[\gamma^0\gamma^\mu(p_\mu - qA_\mu) - \beta mc\right]\psi = 0, \tag{4}$$

by c and $\gamma_0 = \beta$ on each side, leading to

$$i\hbar\frac{\partial}{\partial t}\psi = [c\boldsymbol{\alpha}(\mathbf{p} - q\mathbf{A}) + q\phi + \beta mc^2]\psi = H\psi, \tag{5}$$

with

$$H_D = [c\boldsymbol{\alpha}(\mathbf{p} - q\mathbf{A}) + q\phi + \beta mc^2] \tag{6}$$

as the Dirac Hamiltonian.

The first-order equation (2) can be transformed into a second-order equation by applying the operator $\gamma^\mu(p_\nu - qA_\nu) + mc$:

$$[\gamma^\mu\gamma^\nu(p_\mu - qA_\mu)(p_\nu - qA_\nu) - m^2c^2]\psi = 0. \tag{7}$$

Replacing the second-order tensor $\gamma^\mu\gamma^\nu$ by its decomposition in a symmetric and antisymmetric operator (see Eq. 441), we obtain

$$\gamma^\mu\gamma^\nu = g^{\mu\nu} + \sigma^{\mu\nu}, \tag{8}$$

In the product of $(p_\mu - qA_\mu)(p_\nu - qA_\nu)$ by $\sigma^{\mu\nu}$only the antisymmetric part will contribute. One can then express

$$(p_\mu - qA_\mu)(p_\nu - qA_\nu) \to \frac{1}{2}[(p_\mu - qA_\mu), (p_\nu - qA_\nu)] \tag{9}$$

and

$$\frac{1}{2}[(p_\mu - qA_\mu), (p_\nu - qA_\nu)] = \frac{1}{2}q(-A_\mu p_\nu + p_\nu A_\mu - p_\mu A_\nu + A_\nu p_\mu) \tag{10}$$

$$= \frac{1}{2}i\hbar q(\partial_\nu A_\mu - \partial_\mu A_\nu) \tag{11}$$

$$= -\frac{1}{2}i\hbar q F_{\mu\nu} \tag{12}$$

and obtain an equation in terms of $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, the tensor for the electromagnetic field. To transform (10) into (11), we have used the fact that $[(p_\mu - qA_\mu), (p_\nu - qA_\nu)]$ is an operator that act both on A_μ and on the following wave function, and thus $p_\nu A_\mu \phi = A_\mu p_\nu \phi + i\hbar \left(\partial_\mu A_\nu\right) \phi$, where we use the convention that p_ν applies on all what follows, while ∂_μ applies only on the wave function that immediately follow. We thus obtain the second-order equation:

$$[(p_\mu - qA_\mu)(p^\mu - qA^\mu) - i\frac{q\hbar}{2}\sigma^{\mu\nu}F_{\mu\nu} - m^2c^2]\psi = 0. \tag{13}$$

Expressing $\sigma^{\mu\nu}F_{\mu\nu}$ in the tridimensional form

$$\sigma^{\mu\nu} = (\boldsymbol{\alpha}, i\boldsymbol{\Sigma}), \quad F^{\mu\nu} = \left(-\frac{\mathbf{E}}{c}, \mathbf{B}\right), \tag{14}$$

one gets

$$\sigma^{\mu\nu}F_{\mu\nu} = 2\left(\frac{\boldsymbol{\alpha}\mathbf{E}}{c} + i\boldsymbol{\Sigma}\mathbf{B}\right). \tag{15}$$

The final second-order Dirac equation is then

$$\left[\left(i\hbar\frac{1}{c}\frac{\partial}{\partial t} - q\frac{\phi}{c}\right)^2 - (i\hbar\boldsymbol{\nabla} + q\mathbf{A})^2 - iq\hbar\frac{\boldsymbol{\alpha}\mathbf{E}}{c} + q\hbar\boldsymbol{\Sigma}\mathbf{B} - m^2c^2\right]\psi = 0. \tag{16}$$

Dirac Equation for a Particle with a Landé Factor $g \neq 2$

The second-order Dirac equation (16) corresponds to a Landé factor $g = 2$. One of the first result of QED, obtained by Schwinger, is that the electron has an anomalous magnetic moment such that $g \neq 2$. Other particles that can be bound to an atom like the antiproton may have a Landé factor very different from 2. The second-order Dirac equation can be corrected in that case, as

$$\left[\left(i\hbar\frac{1}{c}\frac{\partial}{\partial t} - q\frac{\phi}{c}\right)^2 - (i\hbar\boldsymbol{\nabla} + q\mathbf{A})^2 + (1+a)\left(-iq\hbar\frac{\boldsymbol{\alpha}\mathbf{E}}{c} + q\hbar\boldsymbol{\Sigma}\mathbf{B}\right) - m^2c^2\right]\psi = 0, \tag{17}$$

with, for example, $1 + a = g/2 = 2.79278$ for an antiproton or $g - 2 = \frac{\alpha}{\pi}$ for an electron.

To obtain the corresponding first-order equation, let's introduce a new term in the original equation (see, e.g., Refs. [19, 20]) (2)

$$\left(\gamma^{\mu}\pi_{\mu} - mc + \lambda\sigma^{\mu\nu}F_{\mu\nu}\right)\psi = 0, \tag{18}$$

with $\pi_{\mu} = (p_{\mu} - qA_{\mu})$ and λ a constant. To obtain the corresponding equation, let's re-derive the second-order equation corresponding to (18), using the same method as above:

$$\left(\gamma^{\mu}\pi_{\mu} - mc + \lambda\sigma^{\mu\nu}F_{\mu\nu}\right)\left(\gamma^{\nu}\pi_{\nu} + mc - \lambda\sigma^{\alpha\beta}F_{\alpha\beta}\right)\psi = 0, \tag{19}$$

which leads to

$$\begin{aligned}&\left(\gamma^{\mu}\gamma^{\nu}\pi_{\mu}\pi_{\nu} - m^2c^2\right.\\ &- \lambda\gamma^{\mu}\pi_{\mu}\sigma^{\alpha\beta}F_{\alpha\beta} + \lambda\sigma^{\mu\nu}F_{\mu\nu}\gamma^{\nu}\pi_{\nu}\\ &+ \left.2mc\lambda\sigma^{\mu\nu}F_{\mu\nu} - \lambda^2\sigma^{\mu\nu}F_{\mu\nu}\sigma^{\alpha\beta}F_{\alpha\beta}\right)\psi = 0.\end{aligned} \tag{20}$$

Among those terms one can recognize the part with $g = 2$

$$\gamma^{\mu}\gamma^{\nu}\pi_{\mu}\pi_{\nu} - m^2c^2 \tag{21}$$

and the anomalous magnetic moment part

$$2mc\lambda\sigma^{\mu\nu}F_{\mu\nu} \tag{22}$$

Comparing the part corresponding to $g - 2$ and (17), one gets

$$\begin{aligned}4imc\lambda &= a\hbar q\\ \lambda &= -ia\frac{\hbar q}{4mc},\end{aligned} \tag{23}$$

leading to the modified Dirac equation (18):

$$\left(\gamma^{\mu}\pi_{\mu} - mc - ia\frac{\hbar q}{4mc}\sigma^{\mu\nu}F_{\mu\nu}\right)\psi = 0. \tag{24}$$

The corresponding Hamiltonian equation for a particle with an anomalous magnetic moment is then

$$h_{Da} = c\boldsymbol{\alpha}(\mathbf{p} - q\mathbf{A}) + q\phi + \beta mc^2 + a\frac{\hbar q}{2m}\beta\left(i\frac{\boldsymbol{\alpha}\mathbf{E}}{c} - \boldsymbol{\Sigma}\mathbf{B}\right). \tag{25}$$

This can be useful to evaluate lowest-order QED corrections to the g-factors in many-electron atoms [21, 22].

Bound States of the Dirac Equation

Second-Order Dirac Equation in a Coulomb Potential

Solving analytically the Dirac equation in a Coulomb potential can be more easily performed starting from the second-order Dirac equation (16) [23]:

$$\left[\left(i\hbar\frac{1}{c}\frac{\partial}{\partial t}-eA^0\right)^2-(i\hbar\nabla+e\mathbf{A})^2-ie\hbar\frac{\boldsymbol{\alpha}\cdot\mathbf{E}}{c}+e\hbar\boldsymbol{\Sigma}\mathbf{B}-m^2c^2\right]\psi=0, \tag{26}$$

where $\mathbf{B}=\mathbf{0}$, for a pure Coulomb field. Using atomic units $V=-Z/r$, $eE=-\hat{r}\,Z/r^2$, where $\hat{r}=\mathbf{r}/r$ is the unit vector along $\mathbf{r}$, we get

$$\left[\frac{1}{c^2}\left(E+\frac{Z}{r}\right)^2+\frac{1}{r}\frac{\partial^2}{\partial r^2}r-\frac{\mathbf{L}^2}{r^2}+iZ\alpha\frac{\boldsymbol{\alpha}\cdot\hat{r}}{r^2}-m^2c^2\right]\psi=0, \tag{27}$$

which can be rewritten as

$$\left[\frac{E^2-m^2c^4}{c^2}+\frac{2E}{c}\frac{Z\alpha}{r}+\frac{(Z\alpha)^2}{r^2}+\frac{1}{r}\frac{\partial^2}{\partial r^2}r-\frac{\mathbf{L}^2}{r^2}+iZ\alpha\frac{\boldsymbol{\alpha}\cdot\hat{r}}{r^2}\right]\psi=0. \tag{28}$$

Regrouping $1/r^2$ terms it leads to

$$\left[\frac{E^2-m^2c^4}{c^2}+\frac{2E}{c}\frac{Z\alpha}{r}+\frac{1}{r}\frac{\partial^2}{\partial r^2}r-\frac{\left(\mathbf{L}^2-iZ\alpha\boldsymbol{\alpha}\cdot\hat{r}-(Z\alpha)^2\right)}{r^2}\right]\psi=0, \tag{29}$$

We now need to study the angular eigenfunctions for this operator.

Symmetries of the Dirac Equation

The Dirac equation is defined so as to respect Lorentz invariance and the usual discrete symmetries (parity, time reversal, charge invariance). It is also invariant by rotations and translation. The transformations between two different reference frames A and A' are given by

$$ic\gamma^\mu\frac{\partial}{\partial x^\mu}\psi(x)-mc^2\psi(x)=0 \tag{30}$$

$$ic\gamma^\mu\frac{\partial}{\partial x'^\mu}\psi'(x')-mc^2\psi'(x')=0 \tag{31}$$

The 4-momentum expression is easily deduced from the one on x:

$$\frac{\partial}{\partial x'^\mu}=\frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial}{\partial x^\nu}=\left(A^{-1}\right)^\nu_\mu\frac{\partial}{\partial x^\nu}. \tag{32}$$

The observer in the second frame of reference should be able to reconstruct $\psi'(x')$ from $\psi(x)$. This writes

$$\psi'(x') = S(A)\psi(x). \tag{33}$$

The transform $S(A)$ obeys $S\left(A^{-1}\right) = S^{-1}(A)$. An observer in the first frame of reference should be able to reconstruct $\psi(x)$ from $\psi'(x')$, leading to

$$\psi(x) = S^{-1}(A)\psi'(x'). \tag{34}$$

The observer in the A' frame should be able to do the inverse transform and one gets

$$\psi(x) = S\left(A^{-1}\right)\psi'(x'), \tag{35}$$

Applying $S(A)$ and (32) to the Dirac equation (31), and multiplying on the left by $S^{-1}(A)$, one gets

$$icS^{-1}(A)\gamma^{\mu}\left(A^{-1}\right)_{\mu}^{\nu}\frac{\partial}{\partial x^{\nu}}S(A)\psi(x) - mc^2\psi(x) = 0 \tag{36}$$

which must be identical to (30)) for all observers to observe the same physics. This leads to the relation

$$\gamma^{\mu}\left(A^{-1}\right)_{\mu}^{\nu} = S(A)\gamma^{\nu}S^{-1}(A). \tag{37}$$

We apply this formalism to the example of the parity operator. It transforms $\mathbf{x}$ into $-\mathbf{x}$ and x^0 leaves invariant. The corresponding matrix Π is thus

$$\Pi = \Pi^{-1}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \tag{38}$$

i.e., $\Pi_{\mu}^{\nu} = g^{\nu\mu}$. Equation (37) for parity thus writes

$$S^{-1}(\Pi)\gamma^{\mu}\Pi_{\mu}^{\nu}S(\Pi) = \gamma^{\nu}. \tag{39}$$

Using the properties of the γ matrices (see Appendix Properties of Dirac γ Matrices), we find that the operator is thus $S(\Pi) = \exp(i\phi)\gamma^0$. The usual convention is to take $\phi = 0$.

Symmetries of Bilinear Forms

For applications to effects like parity violation or time reversal, it is necessary to know the rules of transformation of the bilinear forms constructed from the Dirac wave function. One can define the $\gamma^5 = \gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ matrix, which

Table 1 Properties of the bilinear forms of the Dirac equation. We denote $\boldsymbol{x}_t = (-x^0, \mathbf{x})$ the transform of the 4-vector $\boldsymbol{x}$ under time reversal and $\boldsymbol{x}_p = (x^0, -\mathbf{x})$ the one transformed under parity

Form	Transforming under Lorentz as	C	P	T
$S(\boldsymbol{x}) = \bar{\Psi}(\boldsymbol{x})\Psi(\boldsymbol{x})$	Scalar	$S(\boldsymbol{x}_p)$	$S(\boldsymbol{x}_t)$	$S(\boldsymbol{x})$
$Ps(\boldsymbol{x}) = \bar{\Psi}(\boldsymbol{x})\gamma^5\Psi(\boldsymbol{x})$	Pseudo-scalar	$Ps(\boldsymbol{x})$	$-Ps(\boldsymbol{x}_p)$	$-Ps(\boldsymbol{x}_p)$
$J^\mu(\boldsymbol{x}) = \bar{\Psi}(\boldsymbol{x})\gamma^\mu\Psi(\boldsymbol{x})$	Vector	$-J_\mu(\boldsymbol{x})$	$J_\mu(\boldsymbol{x}_p)$	$J^\mu(\boldsymbol{x}_t)$
$Pj^\mu(\boldsymbol{x}) = \bar{\Psi}(\boldsymbol{x})\gamma^5\gamma^\mu\Psi(\boldsymbol{x})$	Pseudo-vector	$Pj^\mu(\boldsymbol{x})$	$-Pj_\mu(\boldsymbol{x}_p)$	$Pj_\mu(\boldsymbol{x}_t)$
$T^{\mu\nu}(\boldsymbol{x}) = \bar{\Psi}(\boldsymbol{x})\gamma^\mu\gamma^\nu\Psi(\boldsymbol{x})$	Tensor	$-J_{\mu\nu}(\boldsymbol{x})$	$T_{\mu\nu}(\boldsymbol{x}_p)$	$T^{\mu\nu}(\boldsymbol{x}_t)$

anticommutes with all other γ matrices, and obey $\gamma_5^2 = 1$. In Dirac representation, we find

$$\gamma_5 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \tag{40}$$

Any 4×4 matrix can be written as a linear combination of the 16 matrices I, γ^μ, $\sigma^{\mu\nu}$, $\gamma^5\gamma^\mu$, and γ^5. One can define from those 16 matrices a number of bilinear forms. The transformation rules for the different cases are shown on Table 1. We reminds that the prefix "pseudo" concerns transformation by parity.

Spinors

We now turn to the angular properties of the solutions of the Dirac equation in a central potential. We will make use of Pauli's spinors of dimension 2. Using the addition of angular momentum, we build the tensor products of eigenstate functions of L^2, L_z ($Y^{l,m}$) by one of S^2, S_z ($|+>$, $|->$), to obtain an eigenstate of J^2, J_z, L^2 and S^2. We thus have

$$\begin{cases} j = l + \frac{1}{2} \;\; \chi_{j,m_j,l,s} = a_+ Y^{l,m_j-1/2}(\hat{x})|+> + b_+ Y^{l,m_j+1/2}(\hat{x})|-> \\ j = l - \frac{1}{2} \;\; \chi_{j,m_j,l,s} = a_- Y^{l,m_j-1/2}(\hat{x})|+> + b_- Y^{l,m_j+1/2}(\hat{x})|-> \end{cases} \tag{41}$$

where $\hat{x}$ represents the angular variables θ and ϕ, and $a_\pm$ and $b_\pm$ are Clebsh-Gordan coefficients. Rewriting $|\pm>$ as column 2 vectors, we obtain for $j = l + 1/2$

$$\chi_{j,m_j,l,s} = \frac{1}{\sqrt{2l+1}} \begin{bmatrix} \sqrt{l + m + \frac{1}{2}}\, Y^{l,m-\frac{1}{2}}(\hat{x}) \\ \sqrt{l - m + \frac{1}{2}}\, Y^{l,m+\frac{1}{2}}(\hat{x}) \end{bmatrix}, \tag{42}$$

and for $j = l - 1/2$

$$\chi_{j,m_j,l,s} = \frac{1}{\sqrt{2l+1}} \begin{bmatrix} \sqrt{l - m + \frac{1}{2}}\, Y^{l,m-\frac{1}{2}}(\hat{x}) \\ -\sqrt{l + m + \frac{1}{2}}\, Y^{l,m+\frac{1}{2}}(\hat{x}) \end{bmatrix}. \tag{43}$$

To obtain more compact forms and find possible new symmetries, we study the effect over the χ-functions of the product $\mathbf{L}\cdot\sigma$, which represents the angle between the particle spin and angular momentum. This product can be expressed from $\mathbf{J}^2 = (\mathbf{L}+\mathbf{S})^2$, leading to $\mathbf{L}\cdot\boldsymbol{\sigma} = 2\mathbf{L}\cdot\mathbf{S} = \left(\mathbf{J}^2 - \mathbf{L}^2 - \mathbf{S}^2\right)$. It then leads to

$$\begin{cases} \text{for } j = l + \frac{1}{2}: & \mathbf{L}\cdot\boldsymbol{\sigma}\,\chi_{j,m_j,l,s} = \left(j - \frac{1}{2}\right)\chi_{j,m_j,l,s} \\ \text{for } j = l - \frac{1}{2}: & \mathbf{L}\cdot\boldsymbol{\sigma}\,\chi_{j,m_j,l,s} = \left(-j - \frac{3}{2}\right)\chi_{j,m_j,l,s} \end{cases} \tag{44}$$

It is then clear that the operator $1 + \mathbf{L}\cdot\boldsymbol{\sigma}$ has the eigenvalues $\pm(j+1/2)$ when acting on $\chi_{j,m_j,l,s}$. We denote $k = (1 + \mathbf{L}\cdot\boldsymbol{\sigma})$ and the eigenvalue $-\kappa$. One can write l and j as a function of κ, leading to $j = |\kappa| - 1/2$ and $l = |\kappa + 1/2| - 1/2$. We can thus write $\chi_{j,m_j,l,s}$ as χ^{μ}_{κ} with $\mu = m_j$. We note that χ^{μ}_{κ} and $\chi^{\mu}_{-\kappa}$ are eigenvectors of $\mathbf{L}^2$ with l values differing by 1 and are thus of opposite parity.

We now use these equations to separate radial and angular coordinates in the Dirac equation. We want to build a 4-spinor eigenstate of J^2, J_z and of the parity Π and of the Dirac equation. The most general expression of such a 4-spinor is

$$\phi_{n\kappa}(\mathbf{x}) = \begin{bmatrix} f_1(x)\chi^{\mu}_{\kappa}(\hat{x}) + g_1(x)\chi^{\mu}_{-\kappa}(\hat{x}) \\ i f_2(x)\chi^{\mu}_{-\kappa}(\hat{x}) + i g_2(x)\chi^{\mu}_{\kappa}(\hat{x}) \end{bmatrix}, \tag{45}$$

which is not an eigenstate of parity. The i is added so that the radial parts are real. For $\phi_{n\kappa}(\mathbf{x})$ to be an eigenstate of the parity, it must obey $\Pi\phi_{n\kappa}(\mathbf{x}) = \varpi\phi_{n\kappa}(\mathbf{x})$, with $\varpi = \pm 1$. Using (39), i.e., multiplying the wave function by γ^0 and changing $\mathbf{x} \to -\mathbf{x}$ in the spinor, we must have $\Pi\chi^{\mu}_{\kappa}(\hat{x}) = \varpi\chi^{\mu}_{\kappa}(\hat{x})$ and $\Pi\chi^{\mu}_{-\kappa}(\hat{x}) = -\varpi\chi^{\mu}_{-\kappa}(\hat{x})$ since κ et $-\kappa$ correspond to a difference of 1 in the l values and thus to opposite parities. We thus get

$$\begin{aligned} \Pi\phi_{n\kappa}(\mathbf{x}) &= \gamma^0 \begin{bmatrix} f_1(x)\Pi\chi^{\mu}_{\kappa}(\hat{x}) + g_1(x)\Pi\chi^{\mu}_{-\kappa}(\hat{x}) \\ +i f_2(x)\Pi\chi^{\mu}_{-\kappa}(\hat{x}) + i g_2(x)\Pi\chi^{\mu}_{\kappa}(\hat{x}) \end{bmatrix} \\ &= \varpi \begin{bmatrix} f_1(x)\chi^{\mu}_{\kappa}(\hat{x}) - g_1(x)\chi^{\mu}_{-\kappa}(\hat{x}) \\ i f_2(x)\chi^{\mu}_{-\kappa}(\hat{x}) - i g_2(x)\chi^{\mu}_{\kappa}(\hat{x}) \end{bmatrix} \end{aligned} \tag{46}$$

which must be equal to

$$\varpi\phi_{n\kappa}(\mathbf{x}) = \varpi \begin{bmatrix} f_1(x)\chi^{\mu}_{\kappa}(\hat{x}) + g_1(x)\chi^{\mu}_{-\kappa}(\hat{x}) \\ i f_2(x)\chi^{\mu}_{-\kappa}(\hat{x}) + i g_2(x)\chi^{\mu}_{\kappa}(\hat{x}) \end{bmatrix}. \tag{47}$$

This is only possible if $g_1(x) = g_2(x) = 0$. The most general spinor eigenstate of $\mathbf{J}^2$, J_z and Π, solution to the Dirac equation, is thus of the form [24]

$$\phi_{n\kappa}(\mathbf{x}) = \begin{bmatrix} f_1(x)\chi^{\mu}_{\kappa}(\hat{x}) \\ i f_2(x)\chi^{\mu}_{-\kappa}(\hat{x}) \end{bmatrix}. \tag{48}$$

It is not an eigenstate of $\mathbf{L}^2$ and $\mathbf{S}^2$, but it is one of

$$K = \beta k = \beta(\boldsymbol{\sigma} \cdot \mathbf{L} + 1) \tag{49}$$

with the eigenvalue $-\kappa$. We note that $K^2 = 1 + 2\boldsymbol{\sigma} \cdot \mathbf{L} + (\boldsymbol{\sigma} \cdot \mathbf{L})^2 = 1 + \boldsymbol{\sigma} \cdot \mathbf{L} + \mathbf{L}^2$ using the well-known identity

$$(\boldsymbol{\sigma} \cdot \mathbf{A})(\boldsymbol{\sigma} \cdot \mathbf{B}) = \mathbf{A} \cdot \mathbf{B} + i\boldsymbol{\sigma} \cdot \mathbf{A} \times \mathbf{B} \tag{50}$$

with $\mathbf{A} = \mathbf{B} = \mathbf{L}$. Noting that the commutation rules among the components of $\mathbf{L}$ lead to $\mathbf{L} \times \mathbf{L} = i\mathbf{L}$ and that $\boldsymbol{\sigma} = 2\mathbf{S}$, we have $K^2 = \mathbf{J}^2 + 1/4$, and K thus commutes with $\mathbf{J}^2$ and has the eigenvalues $\kappa = \pm\sqrt{j\,(j\,+1) + 1/4} = \pm(j + 1/2)$ as expected.

Solutions of the Dirac Equation in a Coulomb Potential

One can check that $[K, h_D] = 0$, $[K, \mathbf{J}] = 0$, $[K, \boldsymbol{\alpha}\cdot\hat{r}] = 0$, and $[K, \boldsymbol{\alpha}\cdot\mathbf{p}] = 0$, where h_D is defined in Eq. (6). The solutions must be eigenstate of K and J. They can be obtained in a number of methods, the easiest one being probably the one presented in Ref. [25], using the second-order Dirac equation. The eigenenergy is found to be

$$E_{n\kappa} = \frac{mc^2}{\sqrt{1 + \frac{(Z\alpha)^2}{(n - |\kappa| + \sqrt{\kappa^2 - (Z\alpha)^2})^2}}}, \tag{51}$$

$$= \frac{mc^2}{\sqrt{1 + \frac{(Z\alpha)^2}{(n - j - \frac{1}{2} + \sqrt{(j+\frac{1}{2})^2 - (Z\alpha)^2})^2}}} \tag{52}$$

with a series expansion in the vicinity of $Z\alpha \to 0$

$$E_{nj} = mc^2 - \frac{mc^2\,(Z\alpha)^2}{2\,n^2} + \frac{mc^2\left(\frac{3}{2n^4} - \frac{2}{(\frac{1}{2}+j)\,n^3}\right)(Z\alpha)^4}{4} + \mathrm{O}(Z\alpha)^6. \tag{53}$$

These solutions do not depend on the sign of κ. This is due to an extra symmetry, a remnant of the Lentz vector which commutes with the nonrelativistic Hamiltonian and explains why the energy does not depend on l. Defining $\Lambda = -\beta K - iZ\alpha\boldsymbol{\alpha}\cdot\hat{r}$, it is found that the Dirac Hamiltonian in a Coulomb potential commutes with the operator

$$B = \gamma_5\left(\Lambda + \frac{Kh_D}{mc^2}\right) \tag{54}$$

as discovered by Johnson et Lippmann [26].

The radial wave function $f_{n,i}$ in Eq. (48) obeys the differential equations (see, e.g., [27] and Appendix "Appendix: Useful Properties" in this chapter)

$$\begin{aligned}\frac{d}{dx} f_{n,1}(x) &= -\frac{1+\kappa_n}{x} f_{n,1}(x) + [1 + E_n - V(x)]\, f_{n,2}(x) \\ \frac{d}{dx} f_{n,2}(x) &= [1 - E_n + V(x)]\, f_{n,1}(x) - \frac{1-\kappa_n}{x} f_{n,2}(x)\ . \end{aligned} \tag{55}$$

The function $f_{n,1}$ is called the large component and $f_{n,2}$ the small component. The analytic solution of this equation can be obtained. We define

$$\begin{aligned}\gamma = Z\alpha,\ \ \lambda &= \sqrt{\kappa^2 - (Z\alpha)^2}, \\ n_r = n - |\,\kappa\,|,\ \ a &= \frac{Z\alpha}{\sqrt{(n - |\,\kappa\,| + \lambda)^2 + (Z\alpha)^2}}. \end{aligned} \tag{56}$$

With this notation the energy in Eq. (52) is simply $E_{n,\kappa} = \sqrt{1 - a^2}$ in theoretical units. The radial wave function then writes [28]

$$\begin{aligned} f_{1n}(x) &= N_{n,\kappa}^{1/2}(1 + E_{n,\kappa})^{1/2} \sum_{m=0}^{n_r} \left(C_1^{(m)} + C_2^{(m)}\right) (2ax)^{m+\lambda-1} e^{-ax} \\ f_{2n}(x) &= N_{n,\kappa}^{1/2}(1 - E_{n,\kappa})^{1/2} \sum_{m=0}^{n_r} \left(C_1^{(m)} - C_2^{(m)}\right) (2ax)^{m+\lambda-1} e^{-ax} \end{aligned} \tag{57}$$

where the normalisation factor is given by

$$N_{n,\kappa} = \frac{2a^5}{\gamma(\gamma - \kappa a)} \frac{\Gamma(2\lambda + 1 + n_r)}{\Gamma(2\lambda + 1)^2 \Gamma(n_r + 1)} \tag{58}$$

and the coefficients are defined recursively by

$$\begin{aligned} C_1^{(m+1)} &= \frac{m + 1 - n_r}{(m + 2\lambda + 1)(m + 1)} C_1^{(m)} \\ C_2^{(m+1)} &= \frac{m - n_r}{(m + 2\lambda + 1)(m + 1)} C_2^{(m)} \end{aligned} \tag{59}$$

with the initial condition

$$\begin{aligned} C_1^{(0)} &= -n_r \\ C_2^{(0)} &= \frac{\gamma}{a} - \kappa\ . \end{aligned} \tag{60}$$

A specific expression for the $1s$ wave function is given in Ref. [29], and expressions for the $n = 2$ wave functions are given in Ref. [30].

Other Useful Properties of the Dirac Equation in a Coulomb Field

Notations

Here are some useful relations and properties of the Dirac equation (1) in an external potential $V(\mathbf{x})$. It can be rewritten as

$$K = [-i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + V(\mathbf{x}) + \beta - E_n]\phi_n(\mathbf{x}) = 0, \tag{61}$$

using the theoretical units $\hbar = c = m = 1$ and with 4×4 matrices $\boldsymbol{\alpha}$ et β defined in More Properties of the Dirac Matrices. The spinors defined in section "Spinors" by Eq. (43), which obey

$$(\boldsymbol{\sigma}\cdot\mathbf{L} + 1)\ \chi_\kappa^\mu(\hat{x}) = -\kappa\chi_\kappa^\mu(\hat{x}) \tag{62}$$

also verify

$$\boldsymbol{\sigma}\cdot\hat{x}\ \chi_\kappa^\mu(\hat{x}) = -\chi_{-\kappa}^\mu(\hat{x}), \tag{63}$$

which is useful in many calculations.

In a system of spherical coordinates, one can make the variable separation

$$\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} = \boldsymbol{\alpha}\cdot\hat{x}\left(\frac{1}{x}\frac{\partial}{\partial x}x - \beta\frac{K}{x}\right) \tag{64}$$

for the nabla operator in Eq. (61) using the operator K as defined in (49). It can be shown also that K has the property

$$K\boldsymbol{\alpha}\cdot\hat{x}\phi_n(\mathbf{x}) = -\kappa\boldsymbol{\alpha}\cdot\hat{x}\phi_n(\mathbf{x}), \tag{65}$$

from which we can deduce

$$\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}\phi_n(\mathbf{x}) = \boldsymbol{\alpha}\cdot\hat{x}\left(\frac{1}{x}\frac{\partial}{\partial x}x + \beta\frac{\kappa}{x}\right)\phi_n(\mathbf{x}), \tag{66}$$

and

$$\begin{aligned}\nabla^2\phi_n(\mathbf{x}) &= \left[\boldsymbol{\alpha}\cdot\hat{x}\left(\frac{1}{x}\frac{\partial}{\partial x}x - \beta\frac{K}{x}\right)\right]^2\phi_n(\mathbf{x})\\ &= \boldsymbol{\alpha}\cdot\hat{x}\left(\frac{1}{x}\frac{\partial}{\partial x}x + \beta\frac{\kappa}{x}\right)\boldsymbol{\alpha}\cdot\hat{x}\left(\frac{1}{x}\frac{\partial}{\partial x}x + \beta\frac{\kappa}{x}\right)\phi_n(\mathbf{x})\\ &= \left[\frac{1}{x}\frac{\partial^2}{\partial x^2}x - \frac{\kappa(\kappa+\beta)}{x^2}\right]\phi_n(\mathbf{x}).\end{aligned} \tag{67}$$

Substituting Eq. (64) in the Dirac equation (61) leads to the two-component radial Dirac equation

$$\left[\alpha\beta\frac{1}{x}\frac{\partial}{\partial x}x + \alpha\frac{\kappa}{x} + V(x) + \beta - E_n\right]F_n(x) = 0, \tag{68}$$

where

$$F_n(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \tag{69}$$

is the radial function defined in Eq. (57). The 2×2 matrices in Eq. (68) are given by

$$\alpha = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \beta = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{70}$$

We will write $\boldsymbol{\alpha}$ and β the 4×4 Dirac matrices defined in Eq. (437) in the 4-component expressions $\phi_n(\mathbf{x})$ and α and β the 2×2 matrices defined in Eq. (70) in the expressions containing the radial function $F_n(x)$.

Operators Mean Value for the Dirac Equation in a Coulomb Field

A number of useful operators mean values and order-of-magnitude estimates are given in this section. They will be used when dealing with regularization and renormalization of divergent integrals.

The Dirac Hamiltonian (1)

$$H = \boldsymbol{\alpha} \cdot \mathbf{p} + V + \beta \tag{71}$$

satisfies the operator identity

$$\beta H + H\beta = 2\beta V + 2, \tag{72}$$

since $\beta\boldsymbol{\alpha} + \boldsymbol{\alpha}\beta = 0$. Taking the mean value of (72) on the state with principal quantum number n, one obtains

$$E_n \langle\beta\rangle = \langle\beta V\rangle + 1. \tag{73}$$

One can obtain two other exact identities

$$\left\langle p^2\right\rangle = \left\langle(\boldsymbol{\alpha} \cdot \mathbf{p})^2\right\rangle = \left\langle(E_n - V - \beta)^2\right\rangle = E_n^2 - 1 - 2E_n \langle V\rangle + \left\langle V^2\right\rangle \tag{74}$$

and

$$\begin{aligned}\left\langle \beta p^2 \right\rangle &= -\left\langle \boldsymbol{\alpha}\cdot\mathbf{p}\,\beta\,\boldsymbol{\alpha}\cdot\mathbf{p}\right\rangle - \left\langle (E_n - V - \beta)\beta(E_n - V - \beta)\right\rangle \\ &= \left\langle \beta\right\rangle (E_n^2 - 1) - 2\left\langle V\right\rangle - \left\langle \beta V^2\right\rangle. \end{aligned} \tag{75}$$

Equation (73) was used in both cases to simplify the result.

Useful orders of magnitude in α can be obtained for quantities appearing in calculations. These orders of magnitude estimates make use of the fact that for a Coulomb potential the mean value of the radial coordinates x and the momentum $\mathbf{p}$ have orders of magnitude given by $x \sim \frac{1}{Z\alpha}$ et $|\mathbf{p}\,| \sim Z\alpha$.

As a consequence, for example, one has

$$\left\langle V\right\rangle = \left\langle -\frac{Z\alpha}{x}\right\rangle = \mathcal{O}\big((Z\alpha)^2\big) \tag{76}$$

$$\left\langle p^2\right\rangle = \mathcal{O}\big((Z\alpha)^2\big) \tag{77}$$

$$\left\langle V^2\right\rangle = \mathcal{O}\big((Z\alpha)^4\big). \tag{78}$$

These estimates, combined with the fact that the wave function small component is of order $(Z\alpha)$ compared to the large component, lead to

$$\left\langle \boldsymbol{\alpha}\cdot\mathbf{p}\,\right\rangle = \mathcal{O}\big((Z\alpha)^2\big) \tag{79}$$

$$E_n = 1 + \mathcal{O}\big((Z\alpha)^2\big) \tag{80}$$

$$\left\langle \beta V\right\rangle = \left\langle V\right\rangle + \mathcal{O}\big((Z\alpha)^4\big) \tag{81}$$

$$\left\langle V\boldsymbol{\alpha}\cdot\mathbf{p}\,\right\rangle = \mathcal{O}\big((Z\alpha)^4\big). \tag{82}$$

In the same way, it is possible to deduce

$$\begin{aligned} 1 - E_n^2 &= E_n(\left\langle \beta\right\rangle - E_n) - \left\langle \beta V\right\rangle \\ &= \left\langle \beta\right\rangle - E_n - \left\langle V\right\rangle + \mathcal{O}\big((Z\alpha)^4\big) \end{aligned} \tag{83}$$

$$\begin{aligned} \left\langle p^2\right\rangle &= \left\langle (\boldsymbol{\alpha}\cdot\mathbf{p})^2\right\rangle \\ &= \frac{1}{2}\left\langle \boldsymbol{\alpha}\cdot\mathbf{p}\,(E_n - V - \beta) + (E_n - V - \beta)\boldsymbol{\alpha}\cdot\mathbf{p}\,\right\rangle \\ &= \left\langle \boldsymbol{\alpha}\cdot\mathbf{p}\,\right\rangle + \mathcal{O}\big((Z\alpha)^4\big). \end{aligned} \tag{84}$$

Problems Associated with the Predictions of the Dirac Equation

The introduction of the Dirac equation leads to very important prediction and explanation of experimental results concerning the energy spacing of levels and the fine structure in atoms and the prediction of the existence of the positron. Yet progress in experimental techniques leads to new experimental evidences that questioned the theory based on the Dirac equation. In 1947, Lamb and Retherford [31, 32] showed that the $2s_{1/2}$ and $2p_{1/2}$ levels, which should have exactly degenerate energies, have in fact separate energies.

Moreover, Kush and Foley [33] showed the next year that the spectra of sodium and gallium were leading to a Landé factor g_J incompatible with the prediction of $g_e = 2$ for the electron. Those two experiments lead to the birth of a new theory, quantum electrodynamics, following the seminal work of Bethe [34], Kroll and Lamb [35], Feynman [36–38], Schwinger [39–42], Dyson [43, 44], and others.

Quantum Electrodynamics

Introduction

It is not possible here to give a general account of quantum field theory (QFT). Historically, QED has been the model for QFT. Here we will describe the main features of QED and give practical examples of calculations of importance in atoms. The Dirac equation was meant to replace classical mechanics by relativity in quantum theory. Yet the difficulty related to the possibility of particle creation could be solved only by the ad hoc hypothesis of the Dirac electron sea. Quantum field theory represents then the next step after the Dirac equation, to solve this problem. In this theory, based on second quantification, the classical fields and the wave functions are replaced by linear combinations of annihilation and creation operators. The use of the Dirac equation is particularly difficult in the framework of many-body atomic and molecular theory, as, unless specific precautions are taken in the electron-electron interaction treatment, continuum dissolution (unstable solutions) due to the existence of the negative-energy continuum, also known as the Brown and Ravenhall disease [45], will happen (Fig. 1). These issues have been discussed in many references following the work of Brown and Ravenhall and Sucher [46–50]. QED is then the only proper framework where a consistent treatment of the electron-electron interaction and of radiative corrections can be performed. Yet the complexity of the theory and the very strong limitation to our ability to apply it beyond the simplest cases oblige to have approximations based on standard techniques like relativistic many-body perturbation theory or multiconfiguration Dirac-Fock or similar approaches.

Here we define the field operators, the equations they obey, and standard perturbation theory for QED. Complementary informations can be found in Refs. [51–53].

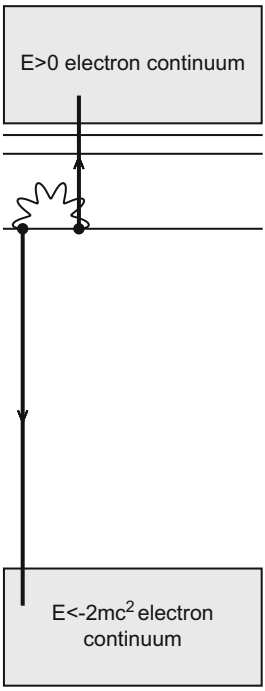


Fig. 1 Continuum dissolution of a two-electron atom. The wavy line represents the electron-electron interaction

Representations

In quantum mechanics, one can choose in which way operators and wave functions depend on time. Each choice corresponds to a so-called representation. Standard quantum mechanics textbooks (see, e.g., [25, 54]) describe in detail the different representations: since the beginning of quantum mechanics, the Schrödinger representation in which the operators are time independent and the Heisenberg one in which the wave functions are time independent. In QED, one uses the intermediate representation, in which both are time dependent. It is often called the Furry representation [55]. From now on, we will suppose that the Hamiltonian is the sum of a main Hamiltonian and of a perturbation, both time independent:

$$H_T = H_0 + H_1. \tag{85}$$

The Schrödinger equation is then

$$i\hbar \frac{d}{dt}\psi = H_T \psi = (H_0 + H_1)\,\psi. \tag{86}$$

We assume that we know the exact solution of the equation without perturbation

$$i\hbar \frac{d}{dt}\psi_0 = H_0 \psi_0, \tag{87}$$

and we want to make explicit the time dependence due to the perturbation. We thus write $\psi_I(t) = e^{(iH_0t/\hbar)}\psi$. The Schrödinger equation (89) for ψ_I then writes

$$\begin{aligned} i\hbar\frac{d}{dt}\psi_I(t) &= -H_0 e^{(iH_0t/\hbar)}\psi + e^{(iH_0t/\hbar)}i\hbar\frac{d}{dt}\psi \\ &= -H_0 e^{(iH_0t/\hbar)}\psi + e^{(iH_0t/\hbar)}(H_0 + H_1)\psi \\ &= e^{(iH_0t/\hbar)}H_1\psi = e^{(iH_0t/\hbar)}H_1 e^{(-iH_0t/\hbar)}\psi_I(t), \end{aligned} \tag{88}$$

using the fact that H_0 commutes with $e^{(iH_0t/\hbar)}$. We thus see that ψ_I obeys

$$i\hbar\frac{d}{dt}\psi(t) = H_I(t)\psi_I(t), \tag{89}$$

where $H_I(t)$, the interaction representation Hamiltonian, is $H_I(t) = e^{(iH_0t/\hbar)} H_1 e^{(-iH_0t/\hbar)}$.

Evolution Operators

We now define the evolution operator, which transform a wave function known at time t into the wave function at time t'. This writes $\psi(t') = U(t',t)\psi(t)$. This operator must obviously be such that $U(t,t) = 1$. Applying the Schrödinger equation to $\psi(t')$, one obtains

$$i\hbar\frac{d}{dt'}\psi(t') = i\hbar\frac{d}{dt'}U(t',t)\psi(t) = H_T U(t',t)\psi(t), \tag{90}$$

from which we can deduce the differential equation for $U(t',t)$,

$$i\hbar\frac{d}{dt'}U(t',t) = H_T U(t',t). \tag{91}$$

In the interaction representation, defining

$$U_I(t',t) = e^{(iH_0(t'-t)/\hbar)}U_T(t',t)e^{(-iH_0(t'-t)/\hbar)} \tag{92}$$

and using a reasoning identical to what was done in section "Representations", we get

$$i\hbar\frac{d}{dt'}U_I(t',t) = H_I U_I(t')(t',t). \tag{93}$$

This equation is solved by integrating in both sides:

$$U_I(t',t) - U_I(t,t) = U_I(t',t) - 1 = \frac{1}{i\hbar}\int_t^{t'} dt_1 H_I(t_1)U_I(t_1,t). \tag{94}$$

A perturbative expansion is obtained by replacing $U_I(t_1,t) = 1$ in the right-hand side

$$U_I(t',t) \approx 1 + \frac{1}{i\hbar}\int_t^{t'} dt_1 H_I(t_1), \tag{95}$$

and by iterating, one obtains

$$U_I(t',t) \approx 1 + \frac{1}{i\hbar}\int_t^{t'} dt_1 H_I(t_1) + \left(\frac{1}{i\hbar}\right)^2 \int_t^{t'}\int_t^{t_1} dt_1 dt_2 H_I(t_1) H_I(t_2) + \cdots \tag{96}$$

the basic equation for QED perturbation expansion.

Fundamental Relations Between Operators for the Electron Field

The next step consists in writing the general solution of the Dirac equation as a linear superposition of all the solutions, including both positive- and negative-energy continua:

$$\begin{aligned}\psi(x) &= \sum_{E_n>0} a_n \phi_n(x) + \sum_{E_m<0} b_m^\dagger \phi_m(x), \\ &= \sum_{E_n>0} a_n e^{-\mathrm{i}E_n t}\phi_n(\mathbf{x}) + \sum_{E_m<0} b_m^\dagger e^{iE_n t}\phi_m(\mathbf{x}).\end{aligned} \tag{97}$$

Here the sums are symbolic notations, which represents the sum over all bound states and integrals over the continua. The a_n et $b_m^\dagger$ are complex numbers and ϕ_n are Dirac Coulomb wave functions defined by $H_D\phi_n = E_n\phi_n$. In order to implement mathematically the idea of the Dirac sea of electrons, we will now assume, as it is currently done, e.g., in solid-state physics for pseudo-particles, that the operator a_n destroys an electron in state n ($E_n > 0$) and that the operator $b_n^\dagger$ creates a positron in state m ($E_m < 0$). To avoid that the electrons will fall into the negative-energy continuum, as represented in Fig. 1, we assume as Dirac that all the $E < 0$ states are filled and that the Pauli principle prevents them to occupy negative-energy states. For that, the operators must obey the following anticommutations rules [56]:

$$\begin{aligned}\left\{a_n, a_{n'}^\dagger\right\} &= \delta_{n,n'}^{+}, \\ \left\{b_n, b_{n'}^\dagger\right\} &= \delta_{n,n'}^{-}, \\ \{a_n, a_{n'}\} &= \left\{a_n^\dagger, a_{n'}^\dagger\right\} = 0, \\ \{b_n, b_{n'}\} &= \left\{b_n^\dagger, b_{n'}^\dagger\right\} = 0,\end{aligned} \tag{98}$$

where

$$\delta^{\pm}_{n,n'} = \theta\left(\pm E_n\right)\delta_{n,n'}\delta_{l,l'}\delta_{j,j'}\delta_{m,m'} \tag{99}$$

if n et n' correspond to bound states,

$$\delta^{\pm}_{n,n'} = \theta\left(\pm E\right)\delta_{E-E'}\delta_{l,l'}\delta_{j,j'}\delta_{m,m'} \tag{100}$$

if n and n' correspond to continuum states, and

$$\delta^{\pm}_{n,n'} = 0 \tag{101}$$

if one of the state is in one of the continua and the other is bound.

Fundamental Operator Relations for the Electromagnetic Field

We now establish the expression equivalent to (97) for the photon field. The most general solution for the wave equation can be written as

$$A_\mu(x) = \int \frac{d\mathbf{k}}{2k_0(2\pi)^3}\sum_{\lambda=0}^{3}\left(a^\lambda(k)\epsilon^\lambda_\mu e^{-ik\cdot x} + a^{\lambda\,\dagger}(k)\epsilon^{\lambda\,*}_\mu e^{ik\cdot x}\right) \tag{102}$$

where $k_0 = |\mathbf{k}|$, ϵ^λ_μ is the polarization vector and where $a^\lambda(k)$ annihilate a photon of polarization λ and momentum $\mathbf{k}$. The integration weight

$$\frac{d\mathbf{k}}{2k_0(2\pi)^3} = \frac{d^4k}{(2\pi)^4}2\pi\delta\left(k^2\right)\theta\left(k^0\right) \tag{103}$$

is invariant by Lorentz transform. One can show that in this case the commutators are given by

$$\left[a^\lambda(k), a^{\lambda'\,\dagger}(k')\right] = -2k^0(2\pi)^3\delta\left(\mathbf{k}-\mathbf{k}'\right)g^{\lambda\lambda'}. \tag{104}$$

Perturbation Theory, *S*-Matrix, and Energy Shift for Bound States

Standard bound-state QED provides [51] a complete description of any atomic system and should enable to calculate its energy to any accuracy. This theory is based on the Furry bound picture [55], described in section "Representations." The unperturbed Dirac Hamiltonian H_D contains the Coulomb field of the nucleus.

All effects of this Coulomb field are thus included to all orders in all the quantities involved in the calculation, which derive from Eq. (1) (electron wave functions and propagators). The electron-electron interaction is treated as a perturbation

$$V_{\epsilon,g} = gH_I e^{-\epsilon|t|}, \tag{105}$$

where

$$H_I = j^\mu A_\mu - \delta M(x). \tag{106}$$

As the electromagnetic interactions can act at an infinite distance, $e^{-\epsilon|t|}$ enables to turn off adiabatically the interaction at $t = \pm\infty$ to recover the unperturbed states before and after the interaction. In Eq. (106),

$$j^\mu = -\frac{e}{2}\left[\bar{\psi}(x)\gamma^\mu, \psi(x)\right] \tag{107}$$

is the 4-current, and

$$\delta M(x) = \frac{\delta m}{2}\left[\bar{\psi}(x), \psi(x)\right] \tag{108}$$

is the mass counter term, which will be needed for a proper definition of perturbation theory, g is a perturbation parameter, e is the electron charge, and A_μ is the photon 4-vector operator defined in Eq. (102). $\psi(x)$, is the electron Dirac field operator defined in Eq. (97). We also define $\bar{\Psi} = \Psi^\dagger\gamma^0$.

Perturbation theory will be dealt with the adiabatic evolution operator

$$U_{\epsilon,g}(t_1, t_2) = Te^{-i\int_{t_1}^{t_2} dt V_{\epsilon,g}(t)}. \tag{109}$$

where T is the time-ordering operator. One can then define the adiabatic S-matrix as $S_{\epsilon,g} = \lim_{t\to\infty} U_{\epsilon,g}(-t, t)$. The energy shift for the p-electron state with no real photons $\left|N_p; 0\right\rangle = \left|n_1, \ldots, n_p; 0\right\rangle$, the unperturbed energy of which is $E^0_{N_p} = \sum_{k=1}^{p} E_{n_k}$, is given by the Gell-Mann and Low theorem [57, 58], as symmetrized by Sucher [59]:

$$\Delta E_{N_p} = \lim_{\substack{\epsilon\to 0 \\ g\to 1}} \frac{i\epsilon g}{2}\frac{\partial}{\partial g}\log\left\langle N_p; 0\right| S_{\epsilon,g}\left|N_p; 0\right\rangle \tag{110}$$

Before moving to perturbation theory, we need to define the disconnected diagrams. We rewrite

$$\left\langle N_p; 0\right| S_{\epsilon,g}\left|N_p; 0\right\rangle = \left\langle N_p; 0\right| S_{\epsilon,g}\left|N_p; 0\right\rangle_C \left\langle 0; 0\right| S_{\epsilon,g}\left|0; 0\right\rangle. \tag{111}$$

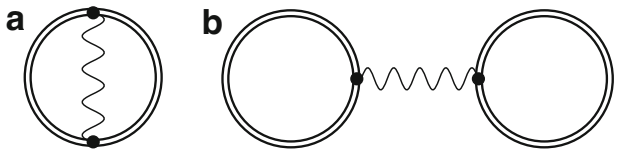


Fig. 2 Disconnected Feynman diagrams of order α. (**a**), for the one-electron case; (**b**), for the two-electron case

Replacing in Eq. (110), we get

$$\begin{aligned}\Delta E_{N_p} &= \lim_{\substack{\epsilon\to 0\\ g\to 1}} \frac{i\epsilon g}{2}\left[\frac{\left\{\frac{\partial}{\partial g}\left\langle N_p;0\right| S_{\epsilon,g}\left|N_p;0\right\rangle_C\right\}\left\langle 0;0\right| S_{\epsilon,g}\left|0;0\right\rangle}{\left\langle N_p;0\right| S_{\epsilon,g}\left|N_p;0\right\rangle_C\left\langle 0;0\right| S_{\epsilon,g}\left|0;0\right\rangle}\right.\\ &\left.+\frac{\left\langle N_p;0\right| S_{\epsilon,g}\left|N_p;0\right\rangle_C\frac{\partial}{\partial g}\left\langle 0;0\right| S_{\epsilon,g}\left|0;0\right\rangle}{\left\langle N_p;0\right| S_{\epsilon,g}\left|N_p;0\right\rangle_C\left\langle 0;0\right| S_{\epsilon,g}\left|0;0\right\rangle}\right]\\ &= \lim_{\substack{\epsilon\to 0\\ g\to 1}} \frac{i\epsilon g}{2}\frac{\partial}{\partial g}\log\left\langle N_p;0\right| S_{\epsilon,g}\left|N_p;0\right\rangle_C + constant \end{aligned} \tag{112}$$

The disconnected diagrams are presented in Fig. 2. They contribute only to the constant in Eq. (112) and affect in an identical manner all levels, since they have no dependence on a specific wave function. They thus shift all level identically and have no effect on transition energies. One can then consider only the connected diagrams in any calculation.

The S-matrix can be expanded in power of g (see, e.g., [51,60])

$$\begin{aligned} g\frac{\partial}{\partial g}\log\left\langle S_{\epsilon,g}\right\rangle_C\Big|_{g=1} &= \frac{\left\langle S^{(1)}_{\epsilon,1}\right\rangle_C + 2\left\langle S^{(2)}_{\epsilon,1}\right\rangle_C + 3\left\langle S^{(3)}_{\epsilon,1}\right\rangle_C + \cdots}{1+\left\langle S^{(1)}_{\epsilon,1}\right\rangle_C + \left\langle S^{(2)}_{\epsilon,1}\right\rangle_C + \left\langle S^{(3)}_{\epsilon,1}\right\rangle_C + \cdots}\\ &= \left\langle S^{(1)}_{\epsilon,1}\right\rangle_C + 2\left\langle S^{(2)}_{\epsilon,1}\right\rangle_C - \left\langle S^{(1)}_{\epsilon,1}\right\rangle^2_C\\ &\quad + 3\left\langle S^{(3)}_{\epsilon,1}\right\rangle_C - 3\left\langle S^{(1)}_{\epsilon,1}\right\rangle_C\left\langle S^{(2)}_{\epsilon,1}\right\rangle_C + \left\langle S^{(1)}_{\epsilon,1}\right\rangle^3_C\\ &\quad + 4\left\langle S^{(4)}_{\epsilon,1}\right\rangle_C - 4\left\langle S^{(1)}_{\epsilon,1}\right\rangle_C\left\langle S^{(3)}_{\epsilon,1}\right\rangle_C - 2\left\langle S^{(2)}_{\epsilon,1}\right\rangle^2_C\\ &\quad + 4\left\langle S^{(1)}_{\epsilon,1}\right\rangle^2_C\left\langle S^{(2)}_{\epsilon,1}\right\rangle_C - \left\langle S^{(1)}_{\epsilon,1}\right\rangle^4_C \end{aligned} \tag{113}$$

where

$$\begin{aligned}\left\langle S^{(j)}_{\epsilon,1}\right\rangle_C &= \left\langle N_p;0\right| S^{(j)}_{\epsilon,1}\left|N_p;0\right\rangle_C\\ \left\langle S_{\epsilon,g}\right\rangle_C &= \left\langle N_p;0\right| S_{\epsilon,g}\left|N_p;0\right\rangle_C\end{aligned} \tag{114}$$

At each order in Eq. (113), there are terms with different powers of $1/\epsilon$. The combination of terms in each order is such that all powers $n > 1$ cancel, so that the energy expression in Eq. (110) has a finite limit when $\epsilon \to 0$.

From the definition of the S-matrix and of the evolution operator (109), one obtains

$$S_{\epsilon,g}^{(j)} = \frac{(-ig)^j}{j!} \int d^4x_j \ldots \int d^4x_1 e^{-\epsilon|t_j|} \ldots e^{-\epsilon|t_1|} T\left[H_I\left(x_j\right) \ldots H_I\left(x_1\right)\right] \tag{115}$$

It should be noted that in the case of QED, where the series expansion parameter is $\alpha = 1/137$, it has been shown by Dyson [61] that the expansion in Eq. (113) is only asymptotic. If one changes α in $-\alpha$, the Dirac equation has no bound states, since it corresponds to two particles of the same charge. This shows that the radius of convergence of the series in α is zero. Yet this has little practical influence, since the divergence will happen only at orders $n \approx 1/\alpha = 137$.

Wick's Theorem, Contractions, and Propagators

Now that we have established the relationship between bound-state energies and field operators, in the framework of the S-matrix formalism, we need to set up the method to evaluate the time-ordered products that appears in Eq. (115). Here we defined the so-called propagators and learned how to manipulate them and relate them to time-ordered products. Wick's theorem [62] relates T-products to normal-ordered field operators, using the commutation rules defined in Eq. (98) and contractions of field operators:

$$T\left[\bar{\psi}\left(x\right)\psi\left(y\right)\right] =: \bar{\psi}\left(x\right)\psi\left(x\right) : + \underbrace{\bar{\psi}\left(x\right)\psi}\left(y\right), \tag{116}$$

where the :: indicates normal ordering, in which all the creation operators have been rearranged to be to the left of all destruction operators and the underbrace the contraction. It is defined as

$$\underbrace{AB} = \langle 0 \mid T[AB] \mid 0 \rangle . \tag{117}$$

More generally Wick has proven that

$$\begin{aligned} T\left[ABCD \ldots X\right] =& : ABC \ldots X : + : \underbrace{AB} CD \ldots X : \\ &+ : \underbrace{AB}\ \underbrace{CD} \ldots X : + \ldots \end{aligned} \tag{118}$$

where A, B, C, $D \ldots$,X are fields operators and the sum is running on all possible contractions.

As it is useful for all two-electron application, we calculate as an example of application the mean value over a two-electron bound state $|i, j\rangle = a_i^\dagger a_j^\dagger |0\rangle$ of a 4-operator normal-ordered product $\langle i, j \,|\, a_k^\dagger a_l^\dagger a_m a_n \,|i, j\rangle$. All indices refer to

positive-energy bound states. We use n as a short for n, l, j, m. There are two possible methods. We can use Wick's theorem with contractions $\underbrace{a_i a_j^\dagger} = \delta_{ij}$, $\underbrace{a_i a_j} = 0$. So all contractions involving two creations or destruction operators vanish. Also vacuum expectation of normal-ordered products vanishes. The only case for which a nonzero value can be obtained must have 4 contractions between a and $a^\dagger$ operators.

$$\begin{aligned}\langle i, j | a_k^\dagger a_l^\dagger a_m a_n | i, j \rangle = & \langle 0 | : a_i a_j a_k^\dagger a_l^\dagger a_m a_n a_i^\dagger a_j^\dagger : | 0 \rangle \\ & + \langle 0 | \underbrace{a_i a_j} : a_k^\dagger a_l^\dagger a_m a_n a_i^\dagger a_j^\dagger : | 0 \rangle \\ & + \langle 0 | \underbrace{a_i a_k} : a_i^\dagger a_l^\dagger a_m a_n a_i^\dagger a_j^\dagger : | 0 \rangle + \cdots \end{aligned} \tag{119}$$

Another method is to make direct use of the anti-commutation relations in Eq 98. One can use the Mathematica [63] package *NCAlgebra* [64] to easily reduce these products.

$$\begin{aligned}\langle i, j | a_k^\dagger a_l^\dagger a_m a_n | i, j \rangle &= \langle 0 | a_i a_j a_k^\dagger a_l^\dagger a_m a_n a_i^\dagger a_j^\dagger | 0 \rangle \\ &= \langle 0 | a_i a_j a_k^\dagger a_l^\dagger a_m (-a_i^\dagger a_n + \delta_{ni}) a_j^\dagger | 0 \rangle \\ &= - \langle 0 | a_i a_j a_k^\dagger a_l^\dagger a_m a_i^\dagger a_n a_j^\dagger | 0 \rangle + \cdots \end{aligned} \tag{120}$$

The result is

$$\begin{aligned}\langle i, j | a_k^\dagger a_l^\dagger a_m a_n | i, j \rangle = & \left(\delta(i,l)\,\delta(j,k) - \delta(i,k)\,\delta(j,l)\right) \\ & \times \left(\delta(m,j)\,\delta(n,i) - \delta(m,i)\,\delta(n,j)\right). \end{aligned} \tag{121}$$

Electron Field Propagator

In order to use efficiently Wick's theorem, one can establish the following relationships:

$$\begin{aligned} T\left[\psi(x)\,\psi(y)\right] &= :\psi(x)\,\psi(y): + \langle 0 | T\left[\psi(x)\,\psi(y)\right] | 0 \rangle \\ &= :\psi(x)\,\psi(y):, \end{aligned} \tag{122}$$

$$\begin{aligned} T\left[\bar{\psi}(x)\,\bar{\psi}(y)\right] &= :\bar{\psi}(x)\,\bar{\psi}(y): + \langle 0 | T\left[\bar{\psi}(x)\,\bar{\psi}(y)\right] | 0 \rangle \\ &= :\bar{\psi}(x)\,\bar{\psi}(y):, \end{aligned} \tag{123}$$

$$T\left[\psi(x)\,\bar{\psi}(y)\right] = :\psi(x)\,\bar{\psi}(y): + \langle 0 | T\left[\psi(x)\,\bar{\psi}(y)\right] | 0 \rangle, \tag{124}$$

$$T\left[\bar{\psi}(x)\,\psi(y)\right] = :\bar{\psi}(x)\,\psi(y): + \langle 0 | T\left[\bar{\psi}(x)\,r\psi(y)\right] | 0 \rangle. \tag{125}$$

The two last expressions are equivalent since

$$T\left[\psi_\alpha(x)\,\bar{\psi}_\beta(y)\right] = -T\left[\bar{\psi}_\beta(y)\,\psi_\alpha(x)\right] \tag{126}$$

The propagator is expressed in terms of the vacuum expectation value of $T\left[\psi(x)\bar{\psi}(y)\right]$ as

$$\begin{aligned} S_F(x,y) &= \langle 0|\, T\left[\psi(x)\bar{\psi}(y)\right]|0\rangle = \underbrace{\psi(x)\bar{\psi}}(y) \\ &= \begin{cases} \sum_{E_n>0}\phi_n(x)\bar{\phi}_n(y) & t_x > t_y \\ -\sum_{E_n<0}\phi_n(x)\bar{\phi}_n(y) & t_x < t_y \end{cases} \\ &= \frac{-i}{2\pi}\int_{-\infty}^{+\infty} dz \sum_n \frac{\phi_n(\mathbf{x})\bar{\phi}_n(\mathbf{y})}{E_n - z(1+i\delta)} e^{-iz(t_2-t_1)} \end{aligned} \tag{127}$$

Contractions of other products are 0:

$$\langle 0|\, T\left[\psi(x)\psi(y)\right]|0\rangle = \langle 0|\, T\left[\bar{\psi}(x)\bar{\psi}(y)\right]|0\rangle = 0 \tag{128}$$

The electron propagator can be written in terms of the Dirac Green function as

$$S_F(x_2,x_1) = \frac{-i}{2\pi}\int_{C_F} dz G\left(\mathbf{x}_2,\mathbf{x}_1,z(1+i\delta)\right)\gamma^0 e^{-iz(t_2-t_1)} \tag{129}$$

where the integration contour C_F is presented in Fig. 3. G obeys

$$\left(-i\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}_2 + V(|\mathbf{x}_2|) + \beta m - z\right) G\left(\mathbf{x}_2,\mathbf{x}_1,z(1+i\delta)\right) = \delta(\mathbf{x}_2-\mathbf{x}_1), \tag{130}$$

and corresponds to the operator $G = (H_D - z)^{-1}$, where H_D is the Dirac Hamiltonian defined in Eq. (6). From Eq. (127), we can write

$$G(\mathbf{x}_2,\mathbf{x}_1,z) = \sum_n \frac{\phi_n(x_2)\phi_n^\dagger(x_1)}{E_n - z}, \tag{131}$$

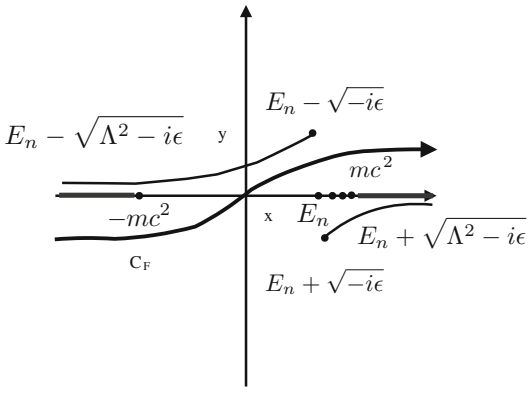


Fig. 3 Integration contour C_F for the Feynman bound propagator. E_n is the ground-state energy. The *black dots* on the x axis represent bound-state poles or $\pm mc^2$. The *gray area* on the x axis corresponds to the positive- and negative-energy continua

where the sum is a symbolic expression that runs over all solutions of the Dirac equation. Replacing by the wave functions as defined in Eq. (48), we obtain

$$G\left(\mathbf{x}_2,\mathbf{x}_1,z\right)=\sum_{n,\kappa,\mu}\frac{1}{E_{n,\kappa}-z}$$
$$\times\begin{bmatrix} f_1(x_2)\chi_\kappa^\mu(\hat{x}_2)f_1(x_1)\chi_\kappa^{\mu\dagger}(\hat{x}_1) & -if_1(x_2)\chi_\kappa^\mu(\hat{x}_2)f_2(x_1)\chi_{-\kappa}^{\mu\dagger}(\hat{x}_1)\\ if_2(x_2)\chi_{-\kappa}^\mu(\hat{x}_2)f_1(x_1)\chi_\kappa^{\mu\dagger}(\hat{x}_1) & f_2(x_2)\chi_{-\kappa}^\mu(\hat{x}_2)f_2(x_1)\chi_{-\kappa}^{\mu\dagger}(\hat{x}_1)\end{bmatrix} \quad (132)$$

Analytic expressions for $G\left(\mathbf{x}_2,\mathbf{x}_1,z\right)$ derived from Eq. (132) can be found in Ref. [29], Appendix A.

When setting the potential $V(x)=0$ in Eq. (130), one obtains the free Green function, which has an exact analytic expression:

$$F(\mathbf{x}_2,\mathbf{x}_1,z)=\left[-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}_2+\beta m+z\right]\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{x}_1|}}{4\pi\,|\mathbf{x}_2-\mathbf{x}_1|}\,, \quad (133)$$

with $c=\sqrt{m^2-z^2}$, $\mathrm{Re}(c)>0$. It obviously satisfies the equation

$$\left[-\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}_2+\beta m-z\right]F(\mathbf{x}_2,\mathbf{x}_1,z)=\delta(\mathbf{x}_2-\mathbf{x}_1)\,. \quad (134)$$

The full Dirac Green function in Eq. (130) can be expanded in powers of the external potential V as presented in Fig. 4:

$$G(\mathbf{x}_2,\mathbf{x}_1,z)=F(\mathbf{x}_2,\mathbf{x}_1,z)-\int d\mathbf{x}_3\,F(\mathbf{x}_2,\mathbf{x}_3,z)V(\mathbf{x}_3)F(\mathbf{x}_3,\mathbf{x}_1,z)+\cdots\,, \quad (135)$$

where $F(\mathbf{x}_2,\mathbf{x}_1,z)$ is the free-electron Dirac Green function, defined in Eq. (133).

A useful relation can be obtained by considering the expansion

$$F\left(\mathbf{x}_2,\mathbf{x}_1,z+\delta z\right)=F\left(\mathbf{x}_2,\mathbf{x}_1,z\right)+\int \mathrm{d}\mathbf{w}\,F\left(\mathbf{x}_2,\mathbf{w},z\right)\delta z\,F\left(\mathbf{w},\mathbf{x}_1,z\right)$$
$$+\int \mathrm{d}\mathbf{v}\int \mathrm{d}\mathbf{w}\,F\left(\mathbf{x}_2,\mathbf{v},z\right)\delta z\,F\left(\mathbf{v},\mathbf{w},z\right)\delta z\,F\left(\mathbf{w},\mathbf{x}_1,z\right)$$
$$+\ldots\,. \quad (136)$$

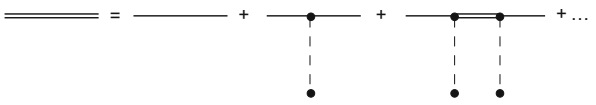


Fig. 4 Bound Coulomb propagator (*double line*) expansion in terms of the free propagator (*single line*). The dashed line corresponds to the Coulomb interaction with the nucleus

This yields

$$\frac{\partial}{\partial\,\delta z}\,F\left(\mathbf{x}_2,\mathbf{x}_1,z+\delta z\right)\Big|_{\delta z=0}=\int \mathrm{d}\mathbf{w}\,F\left(\mathbf{x}_2,\mathbf{w},z\right)F\left(\mathbf{w},\mathbf{x}_1,z\right)\,, \tag{137}$$

so we can write

$$G_{\mathrm{A}}^{(1)}(\mathbf{x}_2,\mathbf{x}_1,z)=-\frac{V(\mathbf{x}_2)+V(\mathbf{x}_1)}{2}\,\frac{\partial}{\partial z}\,F(\mathbf{x}_2,\mathbf{x}_1,z)\,. \tag{138}$$

The Green function can be expressed in terms of radial and angular functions in a way similar to the wave functions as defined in section "Spinors" and Solutions of the Dirac Equation in a Coulomb Potential.

Photon Field Operators

The contraction of two-photon field operators gives the photon propagation function

$$\langle 0|\,A_\mu\left(x_2\right)A_\nu\left(x_1\right)|0\rangle=g_{\mu\nu}D_F\left(x_2-x_1\right) \tag{139}$$

where

$$\begin{aligned}D_F\left(x_2-x_1\right)&=-\frac{i}{(2\pi)^4}\int d^4q\,\frac{e^{-iq\cdot(x_2-x_1)}}{q^2+i\delta}\\&=\frac{1}{(2\pi i)}\int_{-\infty}^{+\infty}dq_0H\left(\mathbf{x}_2-\mathbf{x}_1,q_0\right)e^{-iq_0(t_2-t_1)}\end{aligned} \tag{140}$$

In (140) $H\left(\mathbf{x}_2-\mathbf{x}_1,q_0\right)$ is the photon Green function, which is given by

$$\begin{aligned}H\left(\mathbf{x}_2-\mathbf{x}_1,q_0\right)&=-\frac{e^{-bx_{21}}}{4\pi x_{21}}\\x_{21}=|\mathbf{x}_2-\mathbf{x}_1|\,;\quad b&=-i\left(q_0^2+i\delta\right)^{\frac{1}{2}},\Re(b)>0.\end{aligned} \tag{141}$$

Unperturbed States

With this formalism we can write the state vector for a one-electron state as

$$|N_1\rangle=a_n^\dagger\,|0\rangle \tag{142}$$

and

$$|N_2\rangle=\sum_{n,m}k_{n,m}a_n^\dagger a_m^\dagger\,|0\rangle \tag{143}$$

for a two-electron state. For these states the mean value of the unperturbed Hamiltonian is given by

$$H_0 = \int \mathbf{x} : \psi^\dagger(x) H_D \psi(x) := \sum_{E_n>0} E_n a_n^\dagger a_n - \sum_{E_m<0} E_m b_m^\dagger b_m. \tag{144}$$

The $-$ sign is what distinguishes the treatment of negative-energy states in a quantum field theory from standard Dirac theory.

Evaluation of $\left\langle S_{\epsilon,1}^{(1)} \right\rangle_c$

As an example, we use the formalism setup above to evaluate the first term in Eq. (115). It reads

$$\begin{aligned}\left\langle S_{\epsilon,1}^{(1)} \right\rangle_c &= -i \left\langle N_p; 0 \right| \int d^4x e^{-\epsilon|t|} T \left[H_I(x) \right] \left| N_p; 0 \right\rangle_c \\ &= -i \left\langle N_p; 0 \right| \int d^4x e^{-\epsilon|t|} T \left[j^\mu(x) A_\mu(x) - \delta M(x) \right] \left| N_p; 0 \right\rangle_c \end{aligned} \tag{145}$$

Since there are no external photon lines (i.e., $\langle 0| A_\mu |0\rangle = 0$), we get

$$\begin{aligned}\left\langle S_{\epsilon,1}^{(1)} \right\rangle_c &= i \left\langle N_p; 0 \right| \int d^4x e^{-\epsilon|t|} T \left[\delta M(x) \right] \left| N_p; 0 \right\rangle_c \\ &= \mathrm{i} \left\langle N_p; 0 \right| \int d^4x e^{-\epsilon|t|} T \left[\delta m \bar{\psi}(x) \psi(x) \right] \left| N_p; 0 \right\rangle_c \end{aligned} \tag{146}$$

and we then apply Wick's theorem

$$T \left[\bar{\psi}(x) \psi(x) \right] =: \bar{\psi}(x) \psi(x) : + \underbrace{\bar{\psi}(x) \psi}(x) \tag{147}$$

The second term is a disconnected loop and can be discarded and we are left with

$$\begin{aligned}\left\langle S_{\epsilon,1}^{(1)} \right\rangle_c &= i\delta m \int d^4x e^{-\epsilon|t|} \left\langle N_p; 0 \right| : \bar{\psi}(x) \psi(x) : \left| N_p; 0 \right\rangle \\ &= i\delta m \int d^4x e^{-\epsilon|t|} \left\langle N_p; 0 \right| : \sum_{nm} a_n^\dagger \phi_n^\dagger(x) \gamma^0 a_m \phi_m(x) : \left| N_p; 0 \right\rangle \\ &= i\delta m \int d^4x e^{-\epsilon|t|} \phi_n^\dagger(\mathbf{x}) \gamma^0 \phi_m(\mathbf{x}) \\ &\quad \times \left\langle N_p; 0 \right| : \sum_{nm} e^{i(E_n - E_m)t} a_n^\dagger a_m : \left| N_p; 0 \right\rangle \end{aligned}$$

$$
\begin{aligned}
&= i\delta m \sum_{nm} \int_{-\infty}^{+\infty} dt e^{i(E_n - E_m)t - \epsilon|t|} \int d\mathbf{x} \phi_n^\dagger(\mathbf{x}) \gamma^0 \phi_m(\mathbf{x}) \\
&\quad \times \left\langle N_p; 0 \right| : a_n^\dagger a_m : \left| N_p; 0 \right\rangle \\
&= i\delta m \sum_{nm} \frac{2\epsilon}{(E_n - E_m)^2 + \epsilon^2} \int d\mathbf{x} \phi_n^\dagger(\mathbf{x}) \gamma^0 \phi_m(\mathbf{x}) \\
&\quad \times \left\langle N_p; 0 \right| : a_n^\dagger a_m : \left| N_p; 0 \right\rangle \\
&= i\delta m \sum_{nm} \frac{2}{\epsilon} \delta(E_n, E_m) \int d\mathbf{x} \phi_n^\dagger(\mathbf{x}) \gamma^0 \phi_m(\mathbf{x}) \\
&\quad \times \left\langle N_p; 0 \right| : a_n^\dagger a_m : \left| N_p; 0 \right\rangle
\end{aligned} \tag{148}
$$

using the definition of the field operator (97) and the fact that we consider bound states made only of electrons. Finally we use the fact that $\left\langle N_p; 0 \right| a_n^\dagger a_m \left| N_p; 0 \right\rangle = \delta_{n,n_1}\delta_{m,n_1}$ and (110) to get the first-order energy shift

$$
\Delta E_n^{(1)} = \lim_{\epsilon \to 0} \frac{1}{2} i\epsilon \left\langle S_{\epsilon,1}^{(1)} \right\rangle_c = -\delta m \int d\mathbf{x} \phi_n^\dagger(\mathbf{x}) \gamma^0 \phi_n(\mathbf{x}) . \tag{149}
$$

To this order the energy shift is thus only the mean value of the mass counter term on the bound state considered.

Evaluation of $\left\langle S_{\epsilon,1}^{(2)} \right\rangle_c$

We now turn to the second order of Eq. (115). It reads

$$
\begin{aligned}
\left\langle S_{\epsilon,1}^{(2)} \right\rangle_c &= \frac{(-i)^2}{2!} \left\langle N_p; 0 \right| \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} T\left[H_I(x_2) H_I(x_1)\right] \left| N_p; 0 \right\rangle_c \\
&= -\frac{1}{2} \left\langle N_p; 0 \right| \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} T\left[\left(j^\mu(x_2) A_\mu(x_2) - \delta M(x_2)\right)\right. \\
&\quad \left. \times \left(j^\mu(x_1) A_\mu(x_1) - \delta M(x_1)\right)\right] \left| N_p; 0 \right\rangle_c \\
&= -\frac{1}{2} \left\langle N_p; 0 \right| \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} T\left[j^\mu(x_2) j^\mu(x_1) A_\mu(x_2) A_\mu(x_1)\right. \\
&\quad - j^\mu(x_2) A_\mu(x_2) \delta M(x_1) \\
&\quad - \delta M(x_2) j^\mu(x_1) A_\mu(x_1) \\
&\quad \left. + \delta M(x_2) \delta M(x_1)\right] \left| N_p; 0 \right\rangle_c
\end{aligned} \tag{150}
$$

The terms with a single photon field operator annihilate $|0>$ and we are thus left with

$$\left\langle S^{(2)}_{\epsilon,1}\right\rangle_c = -\frac{1}{2}\left\langle N_p;0\right| \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} T\left[j^\mu(x_2)\, j^\mu(x_1)\, A_\mu(x_2)\, A_\mu(x_1)\right.$$
$$\left. + \delta M(x_2)\, \delta M(x_1)\right] \left|N_p;0\right\rangle_c$$
$$= \left\langle S^{(2a)}_{\epsilon,1}\right\rangle_c + \left\langle S^{(2b)}_{\epsilon,1}\right\rangle_c \tag{151}$$

We first evaluate

$$\left\langle S^{(2b)}_{\epsilon,1}\right\rangle_c = -\frac{1}{2}\left\langle N_p;0\right| \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} T\left[\delta M(x_2)\, \delta M(x_1)\right] \left|N_p;0\right\rangle_c$$
$$= -\frac{(\delta m)^2}{2} \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)}$$
$$\times \left\langle N_p;0\right| T\left[\bar\psi(x_2)\,\psi(x_2)\,\bar\psi(x_1)\,\psi(x_1)\right] \left|N_p;0\right\rangle_c \tag{152}$$

We now apply Wick's theorem.

$$T\left[\bar\psi(x_2)\,\psi(x_2)\,\bar\psi(x_1)\,\psi(x_1)\right] = T\left[\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}(x_1)\,\psi_{\alpha_1}(x_1)\right]$$
$$= :\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}(x_1)\,\psi_{\alpha_1}(x_1):$$
$$+ \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_2}}(x_2) :\bar\psi_{\alpha_1}(x_1)\,\psi_{\alpha_1}(x_1):$$
$$+ \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}}(x_1) :\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}(x_1):$$
$$- \underbrace{\bar\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}}(x_1) :\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}:$$
$$- \underbrace{\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}}(x_1) :\bar\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}(x_1):$$
$$+ \underbrace{\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}}(x_1) :\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}(x_1):$$
$$+ \underbrace{\bar\psi_{\alpha_1}(x_1)\,\psi_{\alpha_1}}(x_1) :\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_2}(x_2):$$
$$+ \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_2}}(x_2)\,\underbrace{\bar\psi_{\alpha_1}(x_1)\,\psi_{\alpha_1}}(x_1)$$
$$- \underbrace{\bar\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}}(x_1)\,\underbrace{\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}}(x_1)$$
$$+ \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}}(x_1)\,\underbrace{\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}}(x_1)\,. \tag{153}$$

We now use Eqs. (122), (123), (124), (125) and the definition of the Green function (127) and of other vacuum expectations (128) to get

$$
\begin{aligned}
&T\left[\bar{\psi}\left(x_2\right)\psi\left(x_2\right)\bar{\psi}\left(x_1\right)\psi\left(x_1\right)\right] \\
&\quad = T\left[\bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_2}\left(x_2\right)\bar{\psi}_{\alpha_1}\left(x_1\right)\psi_{\alpha_1}\left(x_1\right)\right] \\
&\quad =: \bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_2}\left(x_2\right)\bar{\psi}_{\alpha_1}\left(x_1\right)\psi_{\alpha_1}\left(x_1\right): && (154)\\
&\qquad -S_F\left(x_2,x_2\right)_{\alpha_2,\alpha_2} : \bar{\psi}_{\alpha_1}\left(x_1\right)\psi_{\alpha_1}\left(x_1\right): && (155)\\
&\qquad -S_F\left(x_1,x_2\right)_{\alpha_1,\alpha_2} : \psi_{\alpha_2}\left(x_2\right)\bar{\psi}_{\alpha_1}\left(x_1\right): && (156)\\
&\qquad +S_F\left(x_2,x_1\right)_{\alpha_2,\alpha_1} : \bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_1}\left(x_1\right): && (157)\\
&\qquad -S_F\left(x_1,x_1\right)_{\alpha_1,\alpha_1} : \bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_2}\left(x_2\right): && (158)\\
&\qquad +S_F\left(x_2,x_2\right)_{\alpha_2,\alpha_2} S_F\left(x_1,x_1\right)_{\alpha_1,\alpha_1} && (159)\\
&\qquad -S_F\left(x_1,x_2\right)_{\alpha_1,\alpha_2} S_F\left(x_2,x_1\right)_{\alpha_2,\alpha_1}. && (160)
\end{aligned}
$$

Returning to the matrix element we get, removing the disconnected loops (155), (159), (160), (160),

$$
\begin{aligned}
\left\langle S_{\epsilon,1}^{(2b)}\right\rangle_c &= -\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} \\
&\quad \left\langle N_p;0\right| : \bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_2}\left(x_2\right)\bar{\psi}_{\alpha_1}\left(x_1\right)\psi_{\alpha_1}\left(x_1\right): && (161)\\
&\quad -S_F\left(x_1,x_2\right)_{\alpha_1,\alpha_2} : \psi_{\alpha_2}\left(x_2\right)\bar{\psi}_{\alpha_1}\left(x_1\right): && (162)\\
&\quad +S_F\left(x_2,x_1\right)_{\alpha_2,\alpha_1} : \bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_1}\left(x_1\right): \left|N_p;0\right\rangle && (163)
\end{aligned}
$$

We evaluate (161) as $\left\langle S_{\epsilon,1}^{(2b,1)}\right\rangle_c$. It writes

$$
\begin{aligned}
\left\langle S_{\epsilon,1}^{(2b,1)}\right\rangle_c &= -\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} \\
&\quad \left\langle N_p;0\right| : \bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_2}\left(x_2\right)\bar{\psi}_{\alpha_1}\left(x_1\right)\psi_{\alpha_1}\left(x_1\right): \left|N_p;0\right\rangle \\
&= -\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} \\
&\quad \times\sum_{n_2m_2}\sum_{n_1m_1} e^{i\left(E_{n_2}-E_{m_2}\right)t_2} e^{i\left(E_{n_1}-E_{m_1}\right)t_1} \\
&\quad \times\phi_{n_2}^{\dagger}\left(\mathbf{x}_2\right)\gamma^0\phi_{m_2}\left(\mathbf{x}_2\right)\phi_{n_1}^{\dagger}\left(\mathbf{x}_1\right)\gamma^0\phi_{m_1}\left(\mathbf{x}_1\right) \\
&\quad \times\left\langle N_p;0\right| : a_{n_2}^{\dagger}a_{n_1}^{\dagger}a_{m_2}a_{m_1} : \left|N_p;0\right\rangle \\
&= -\frac{(\delta m)^2}{2}\sum_{n_2m_2}\sum_{n_1m_1}\int d\mathbf{x}_2\int d\mathbf{x}_1\left(\frac{2}{\epsilon}\right)^2\delta\left(E_{n_2},E_{m_2}\right)\delta\left(E_{n_1},E_{m_1}\right) \\
&\quad \times\phi_{n_2}^{\dagger}\left(\mathbf{x}_2\right)\gamma^0\phi_{m_2}\left(\mathbf{x}_2\right)\phi_{n_1}^{\dagger}\left(\mathbf{x}_1\right)\gamma^0\phi_{m_1}\left(\mathbf{x}_1\right) \\
&\quad \times\left\langle N_p;0\right| : a_{n_2}^{\dagger}a_{n_1}^{\dagger}a_{m_2}a_{m_1} : \left|N_p;0\right\rangle && (164)
\end{aligned}
$$

We proceed in the same manner as in (148). Lets first consider a one-electron state $|N;0>=a_N^\dagger|0;0>$. We obtain

$$\langle N;0|:a_{n_2}^\dagger a_{n_1}^\dagger a_{m_2}a_{m_1}:|N;0\rangle=\langle 0;0|\,\delta_{N,n_2}a_{n_1}^\dagger a_{m_2}\delta_{N,m_1}\,|0;0\rangle=0 \tag{165}$$

Thus in the one-electron cases, $\left\langle S_{\epsilon,1}^{(2b,1)}\right\rangle_{c,1e^-}=0$. For a two-electron state with $|N_1,N_2;0\rangle=a_{N_1}^\dagger a_{N_2}^\dagger|0,0;0>$, the result is given in Eq. (121). For compactness we will usually leave the results without expanding the average values of the creation and annihilation operators. Here however, because of the integral of wave function products like $\phi_{n_2}^\dagger(\mathbf{x}_2)\gamma^0\phi_{m_2}$, which would give 0 if the angular parts of the wave function do not have the same κ and μ, the $\delta(E_n,E_m)$ are really enforcing $n=m$. We obtain

$$\begin{aligned}\left\langle S_{\epsilon,1}^{(2b,1)}\right\rangle_{c,2e^-}&=\frac{(\delta m)^2}{2}\sum_{n_2m_2}\sum_{n_1m_1}\int d\mathbf{x}_2\int d\mathbf{x}_1\left(\frac{2}{\epsilon}\right)^2\\&\quad\times\phi_{n_2}^\dagger(\mathbf{x}_2)\gamma^0\phi_{n_2}(\mathbf{x}_2)\,\phi_{n_1}^\dagger(\mathbf{x}_1)\gamma^0\phi_{n_1}(\mathbf{x}_1)\\&\quad\times(\delta(n1,N2)\,\delta(n2,N1)-\delta(n1,N1)\,\delta(n2,N2))\\&=\left(\frac{2}{\epsilon}\right)^2\frac{(\delta m)^2}{2}\int d\mathbf{x}_2\int d\mathbf{x}_1\\&\quad\times\Big(\phi_{N_1}^\dagger(\mathbf{x}_2)\gamma^0\phi_{N_1}(\mathbf{x}_2)\,\phi_{N_2}^\dagger(\mathbf{x}_1)\gamma^0\phi_{N_2}(\mathbf{x}_1)\\&\quad-\phi_{N_1}^\dagger(\mathbf{x}_2)\gamma^0\phi_{N_1}(\mathbf{x}_2)\,\phi_{N_2}^\dagger(\mathbf{x}_1)\gamma^0\phi_{N_2}(\mathbf{x}_1)\Big)\\&=0\end{aligned} \tag{166}$$

because we can interchange x_1 and x_2 integration in the second term.

The next term (162) is

$$\begin{aligned}\left\langle S_{\epsilon,1}^{(2b,2)}\right\rangle_c&=-\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1e^{-\epsilon(|t_2|+|t_1|)}\\&\quad\left\langle N_p;0\right|-S_F(x_1,x_2)_{\alpha_1,\alpha_2}:\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}(x_1):\left|N_p;0\right\rangle\\&=+\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1e^{-\epsilon(|t_2|+|t_1|)}\sum_{nm}e^{i(t_1E_m-t_2E_n)}\\&\quad\times S_F(x_1,x_2)_{\alpha_1,\alpha_2}\phi_{n,\alpha_2}(\mathbf{x}_2)\,\gamma^0_{\beta,\alpha_1}\phi_{m,\beta}^\dagger(\mathbf{x}_1)\left\langle N_p;0\right|:a_na_m^\dagger:\left|N_p;0\right\rangle\\&=-\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1e^{-\epsilon(|t_2|+|t_1|)}\sum_{nm}e^{i(t_1E_m-t_2E_n)}\\&\quad\times\phi_m^\dagger(\mathbf{x}_1)\gamma^0S_F(x_1,x_2)\,\phi_n(\mathbf{x}_2)\left\langle N_p;0\right|a_m^\dagger a_n\left|N_p;0\right\rangle.\end{aligned} \tag{167}$$

Similarly we obtain for (163)

$$
\begin{aligned}
\left\langle S_{\epsilon,1}^{(2b,4)}\right\rangle_c &= -\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)}\\
&\quad\times\left\langle N_p;0\right|\, S_F\left(x_2,x_1\right)_{\alpha_2,\alpha_1} : \bar{\psi}_{\alpha_2}\left(x_2\right)\psi_{\alpha_1}\left(x_1\right) :\left|N_p;0\right\rangle\\
&= -\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)}\sum_{lm} e^{i\left(t_2E_m-t_1E_l\right)}\\
&\quad\times S_F\left(x_2,x_1\right)_{\alpha_2,\alpha_1}\phi^{\dagger}_{m,\beta}\left(\mathbf{x}_2\right)\gamma^0_{\beta,\alpha_2}\phi_{l,\alpha_1}\left(\mathbf{x}_1\right)\\
&\quad\left\langle N_p;0\right| : a_m^{\dagger}a_l :\left|N_p;0\right\rangle\\
&= -\frac{(\delta m)^2}{2}\int d^4x_2\int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)}\sum_{lm} e^{i\left(t_2E_m-t_1E_l\right)}\\
&\quad\times\phi^{\dagger}_m\left(\mathbf{x}_2\right)\gamma^0 S_F\left(x_2,x_1\right)\phi_l\left(\mathbf{x}_1\right)\left\langle N_p;0\right|\, a_m^{\dagger}a_l\,\left|N_p;0\right\rangle\\
&= \frac{i}{2\pi}\frac{(\delta m)^2}{2}\sum_{lm}\int_{-\infty}^{+\infty} dz\int d^4x_2\int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)}\\
&\quad\times e^{i\left(t_2E_m-t_1E_l\right)}e^{-iz(t_2-t_1)}\\
&\quad\times\phi^{\dagger}_m\left(\mathbf{x}_2\right)\gamma^0 G\left(\mathbf{x}_2,\mathbf{x}_1,z(1+i\delta)\right)\gamma^0\phi_l\left(\mathbf{x}_1\right)\left\langle N_p;0\right|\, a_m^{\dagger}a_l\,\left|N_p;0\right\rangle\\
&= \frac{i}{2\pi}\frac{(\delta m)^2}{2}\sum_{lm}\int_{-\infty}^{+\infty} dz\int d\mathbf{x}_2\int d\mathbf{x}_1\\
&\quad\times\frac{4\epsilon^2}{\left((E_m-z)^2+\epsilon^2\right)\left((E_l-z)^2+\epsilon^2\right)}\\
&\quad\times\phi^{\dagger}_m\left(\mathbf{x}_2\right)\gamma^0 G\left(\mathbf{x}_2,\mathbf{x}_1,z(1+i\delta)\right)\gamma^0\phi_l\left(\mathbf{x}_1\right)\\
&\quad\times\left\langle N_p;0\right|\, a_m^{\dagger}a_l\,\left|N_p;0\right\rangle. \qquad (168)
\end{aligned}
$$

which is exactly $\left\langle S_{\epsilon,1}^{(2b,4)}\right\rangle_c$ as can be seen by exchanging x_1 and x_2 in the above expression. We thus finally have, specializing on the one-electron case,

$$
\begin{aligned}
\left\langle S_{\epsilon,1}^{(2b)}\right\rangle_c &= \frac{i}{2\pi}(\delta m)^2\sum_{lm}\int_{-\infty}^{+\infty} dz\int d\mathbf{x}_2\int d\mathbf{x}_1\\
&\quad\times\frac{4\epsilon^2}{\left((E_m-z)^2+\epsilon^2\right)\left((E_l-z)^2+\epsilon^2\right)}
\end{aligned}
$$

$$\times \phi_m^\dagger(\mathbf{x}_2)\gamma^0 \sum_{n_1} \frac{\phi_{n_1}(\mathbf{x}_2)\phi_{n_1}^\dagger(\mathbf{x}_1)}{E_{n_1} - z(1+i\delta)}\gamma^0\phi_n(\mathbf{x}_1)\delta(l,n)\delta(m,n)$$

$$= \frac{i}{2\pi}(\delta m)^2 \int_{-\infty}^{+\infty} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1 \frac{4\epsilon^2}{\left((E_n - z)^2 + \epsilon^2\right)^2}$$

$$\times \phi_n^\dagger(\mathbf{x}_2)\gamma^0 \sum_{n_1} \frac{\phi_{n_1}(\mathbf{x}_2)\phi_{n_1}^\dagger(\mathbf{x}_1)}{E_{n_1} - z(1+i\delta)}\gamma^0\phi_n(\mathbf{x}_1) \tag{169}$$

The integration over z is performed by changing the contour that can be rotated to go along the imaginary axis. One can then drop δ in the expression and obtain, in the limit $\epsilon \to 0$

$$\int_{-\infty}^{+\infty} dz \frac{\epsilon^2}{\left((E_m - z)^2 + \epsilon^2\right)^2} \frac{1}{E_{n_1} - z(1+i\delta)} = \frac{i\pi}{\epsilon^2}\delta(E_m, E_{n_1}) \tag{170}$$

from which follows

$$\left\langle S_{\epsilon,1}^{(2b)} \right\rangle_c = \frac{i}{2\pi}\frac{4i\pi}{\epsilon^2}(\delta m)^2 \int d\mathbf{x}_2 \int d\mathbf{x}_1 \phi_n^\dagger(\mathbf{x}_2)\gamma^0\phi_n(\mathbf{x}_2)\phi_n^\dagger(\mathbf{x}_1)\gamma^0\phi_n(\mathbf{x}_1)$$

$$= \frac{-2(\delta m)^2}{\epsilon^2}\left(\int d\mathbf{x}\phi_n^\dagger(\mathbf{x})\gamma^0\phi_n(\mathbf{x})\right)^2 \tag{171}$$

The contribution to the energy shift of $\left\langle S_{\epsilon,1}^{(2b)} \right\rangle_c$ as in Eq. (149), including the contribution of $\left\langle S_{\epsilon,1}^{(1)} \right\rangle_c^2$ that belongs to the same order, is then

$$\Delta E_n^{(2b)} = \lim_{\epsilon\to 0} \frac{1}{2} i\epsilon \left(2\left\langle S_{\epsilon,1}^{(2b)} \right\rangle_c - \left\langle S_{\epsilon,1}^{(1)} \right\rangle_c^2\right)$$

$$= \lim_{\epsilon\to 0} \frac{1}{2} i\epsilon \left(\frac{-2(\delta m)^2}{\epsilon^2}\left(\int d\mathbf{x}\phi_n^\dagger(\mathbf{x})\gamma^0\phi_n(\mathbf{x})\right)^2\right.$$

$$\left. - \left(-i\delta m \frac{2}{\epsilon}\int d\mathbf{x}\phi_n^\dagger(\mathbf{x})\gamma^0\phi_n(\mathbf{x})\right)^2\right) = 0 \tag{172}$$

Such a cancellation is necessary to avoid an infinite contribution since both pieces of the S-matrix behaved as $1/\epsilon^2$.

We now turn to the evaluation of $\left\langle S_{\epsilon,1}^{(2a)}\right\rangle_c$ in Eq. (151). It writes

$$\begin{aligned}\left\langle S_{\epsilon,1}^{(2a)}\right\rangle_c &= -\frac{1}{2}\int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} \\ &\quad \times \left\langle N_p;0\right| T\left[j^\mu(x_2)\, j^\nu(x_1)\, A_\mu(x_2)\, A_\nu(x_1)\right]\left|N_p;0\right\rangle_c \\ &= -\frac{1}{2}\int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} \\ &\quad \times \left\langle N_p;0\right| T\Big[\bar\psi(x_2)\,\gamma^\mu\psi(x_2)\,\bar\psi(x_1)\,\gamma^\nu\psi(x_1) \\ &\quad \times A_\mu(x_2)\, A_\nu(x_1)\Big]\left|N_p;0\right\rangle_c \end{aligned} \tag{173}$$

$$\begin{aligned}&= \frac{1}{(4\pi i)}\int_{-\infty}^{+\infty} dq_0 \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)} e^{-iq_0(t_2-t_1)}\frac{e^{-bx_{21}}}{4\pi x_{21}} \\ &\quad \times \left\langle N_p;0\right| T\left[\bar\psi(x_2)\,\gamma^\mu\psi(x_2)\,\bar\psi(x_1)\,\gamma^\nu\psi(x_1)\right]\left|N_p;0\right\rangle_c \end{aligned} \tag{174}$$

with the expressions from (140) and (141) for the photon field. Wick's theorem can be applied to the preceding expression:

$$\begin{aligned} &T\left[\bar\psi(x_2)\,\gamma^\mu\psi(x_2)\,\bar\psi(x_1)\,\gamma^\nu\psi(x_1)\right] \\ &= \gamma^\mu_{\alpha_2\beta_2}\gamma^\nu_{\alpha_1\beta_1} T\left[\bar\psi_{\alpha_2}(x_2)\,\psi_{\beta_2}(x_2)\,\bar\psi_{\alpha_1}(x_1)\,\psi_{\beta_1}(x_1)\right] \\ &= \gamma^\mu_{\alpha_2\beta_2}\gamma^\nu_{\alpha_1\beta_1}\big(\,:\bar\psi_{\alpha_2}(x_2)\,\psi_{\beta_2}(x_2)\,\bar\psi_{\alpha_1}(x_1)\,\psi_{\beta_1}(x_1):\\ &\quad + \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\beta_2}}(x_2) : \bar\psi_{\alpha_1}(x_1)\,\psi_{\beta_1}(x_1) : \\ &\quad + \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}}(x_1) : \psi_{\beta_2}(x_2)\,\bar\psi_{\beta_1}(x_1) : \\ &\quad - \underbrace{\bar\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}}(x_1) : \psi_{\beta_2}(x_2)\,\psi_{\beta_1} : \\ &\quad - \underbrace{\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}}(x_1) : \bar\psi_{\beta_2}(x_2)\,\bar\psi_{\beta_1}(x_1) : \\ &\quad + \underbrace{\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}}(x_1) : \bar\psi_{\beta_2}(x_2)\,\psi_{\beta_1}(x_1) : \\ &\quad + \underbrace{\bar\psi_{\alpha_1}(x_1)\,\psi_{\beta_1}}(x_1) : \bar\psi_{\alpha_2}(x_2)\,\psi_{\beta_2}(x_2) : \\ &\quad + \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\beta_2}}(x_2)\,\underbrace{\bar\psi_{\alpha_1}(x_1)\,\psi_{\beta_1}}(x_1) \\ &\quad - \underbrace{\bar\psi_{\alpha_2}(x_2)\,\bar\psi_{\alpha_1}}(x_1)\,\underbrace{\psi_{\beta_2}(x_2)\,\psi_{\beta_1}}(x_1) \\ &\quad + \underbrace{\bar\psi_{\alpha_2}(x_2)\,\psi_{\alpha_1}}(x_1)\,\underbrace{\psi_{\beta_2}(x_2)\,\bar\psi_{\beta_1}}(x_1)\,\big) \end{aligned} \tag{175}$$

$$= \gamma^{\mu}_{\alpha_2\beta_2}\gamma^{\nu}_{\alpha_1\beta_1} (: \bar{\psi}_{\alpha_2}(x_2)\,\psi_{\beta_2}(x_2)\,\bar{\psi}_{\alpha_1}(x_1)\,\psi_{\beta_1}(x_1) : \tag{176}$$

$$+S_{F\,\beta_2\alpha_2}(x_2,x_2) : \psi_{\beta_1}(x_1)\,\bar{\psi}_{\alpha_1}(x_1) : \tag{177}$$

$$+S_{F\,\alpha_1\alpha_2}(x_1,x_2) : \bar{\psi}_{\beta_1}(x_1)\,\psi_{\beta_2}(x_2) : \tag{178}$$

$$+S_{F\,\alpha_2\alpha_1}(x_2,x_1) : \bar{\psi}_{\beta_2}(x_2)\,\psi_{\beta_1}(x_1) : \tag{179}$$

$$+S_{F\,\alpha_1\beta_1}(x_1,x_1) : \psi_{\beta_2}(x_2)\,\bar{\psi}_{\alpha_2}(x_2) : \tag{180}$$

$$= \gamma^{\mu}_{\alpha_2\beta_2}\gamma^{\nu}_{\alpha_1\beta_1} : \bar{\psi}_{\alpha_2}(x_2)\,\psi_{\beta_2}(x_2)\,\bar{\psi}_{\alpha_1}(x_1)\,\psi_{\beta_1}(x_1) :$$

$$+2\gamma^{\mu}_{\alpha_2\beta_2}\gamma^{\nu}_{\alpha_1\beta_1}S_{F\,\beta_2\alpha_2}(x_2,x_2) : \psi_{\beta_1}(x_1)\,\bar{\psi}_{\alpha_1}(x_1) :$$

$$+2\gamma^{\mu}_{\alpha_2\beta_2}\gamma^{\nu}_{\alpha_1\beta_1}S_{F\,\alpha_2\alpha_1}(x_2,x_1) : \bar{\psi}_{\beta_2}(x_2)\,\psi_{\beta_1}(x_1) :$$

$$=: \bar{\psi}\gamma^{\mu}(x_2)(x_2)\,\bar{\psi}(x_1)\,\gamma^{\nu}\psi(x_1) : \tag{181}$$

$$-2Tr\left[\gamma^{\mu}S_F(x_2,x_2)\right] : \bar{\psi}(x_1)\,\gamma^{\nu}\psi(x_1) : \tag{182}$$

$$+2\gamma^{\mu}_{\alpha_2\beta_2}\gamma^{\nu}_{\alpha_1\beta_1}S_{F\,\alpha_2\alpha_1}(x_2,x_1) : \bar{\psi}_{\beta_2}(x_2)\,\psi_{\beta_1}(x_1) : \tag{183}$$

removing term with contraction of type (128) and disconnected loops, exchanging $\bar{\psi}$ and ψ to get proper order in propagators, and exchanging 1 and 2 in summed indices and in integration variables. We then replace the field operators by their value from Eq. (97) and go back to the S-matrix evaluation

$$\left\langle S^{(2a)}_{\epsilon,1}\right\rangle_c = \frac{1}{(4\pi i)}\int_{-\infty}^{+\infty} dq_0 \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)}e^{-iq_0(t_2-t_1)}\frac{e^{-bx_{21}}}{4\pi x_{21}}$$

$$\times\Big\{\left\langle N_p;0\right| : \bar{\psi}(x_2)\,\gamma^{\mu}(x_2)\,\bar{\psi}(x_1)\,\gamma^{\nu}\psi(x_1) : \left|N_p;0\right\rangle$$

$$-2Tr\left[\gamma^{\mu}S_F(x_2,x_2)\right]\left\langle N_p;0\right| : \bar{\psi}(x_1)\,\gamma^{\nu}\psi(x_1) : \left|N_p;0\right\rangle$$

$$+2\gamma^{\mu}_{\alpha_2\beta_2}\gamma^{\nu}_{\alpha_1\beta_1}S_{F\,\alpha_2\alpha_1}(x_2,x_1)\left\langle N_p;0\right| : \bar{\psi}_{\beta_2}(x_2)\,\psi_{\beta_1}(x_1) : \left|N_p;0\right\rangle\Big\}. \tag{184}$$

We can make the time dependence of the wave functions apparent, leading to

$$\left\langle S^{(2a)}_{\epsilon,1}\right\rangle_c = \frac{1}{(4\pi i)}\int_{-\infty}^{+\infty} dq_0 \int d^4x_2 \int d^4x_1 e^{-\epsilon(|t_2|+|t_1|)}e^{-iq_0(t_2-t_1)}\frac{e^{-bx_{21}}}{4\pi x_{21}}$$

$$\Big\{\sum_{m_2n_2m_1n_1} e^{i\left(E_{n_2}-E_{m_2}\right)t_2}e^{i\left(E_{n_1}-E_{m_1}\right)t_1}$$

$$\times\phi^{\dagger}_{n_2}(\mathbf{x}_2)\,\gamma^0\gamma^{\mu}\phi_{m_2}(\mathbf{x}_2)\,\phi^{\dagger}_{n_1}(\mathbf{x}_1)\,\gamma^0\gamma^{\nu}\phi_{m_1}(\mathbf{x}_1)$$

$$\times\left\langle N_p;0\right| : a^{\dagger}_{n_2}a_{m_2}a^{\dagger}_{n_1}a_{m_1} : \left|N_p;0\right\rangle \tag{185a}$$

$$
\begin{aligned}
&-2Tr\left[\gamma^{\mu}\frac{-i}{2\pi}\int_{-\infty}^{+\infty}dz G\left(\mathbf{x}_2,\mathbf{x}_2,z(1+i\delta)\right)\gamma^0\right] \\
&\times\sum_{mn}e^{i(E_n-E_m)t_1}\phi_n^{\dagger}\left(\mathbf{x}_1\right)\gamma^0\gamma^{\nu}\phi_m\left(\mathbf{x}_1\right)\left\langle N_p;0\right|a_n^{\dagger}a_m\left|N_p;0\right\rangle \qquad (185b)\\
&+\frac{-i}{\pi}\int_{-\infty}^{+\infty}dz\sum_{nm}e^{i(E_nt_2-E_mt_1-iz(t_2-t_1))} \\
&\times\phi_{n_1}^{\dagger}\left(\mathbf{x}_2\right)\gamma^0\gamma^{\mu}G\left(\mathbf{x}_2,\mathbf{x}_1,z(1+i\delta)\right)\gamma^0\gamma^{\nu}\phi_m\left(\mathbf{x}_1\right) \\
&\times\left\langle N_p;0\right|a_n^{\dagger}a_m\left|N_p;0\right\rangle\Bigg\} \qquad (185c)
\end{aligned}
$$

The Feynman diagrams of order α corresponding to the three contributions above are presented in Figs. 5 and 6. The term (185a) is the electron-electron interaction (Fig. 5), the term (185c) is the self-energy (Fig. 6a), and the term (185b) corresponds to vacuum polarization (Fig. 6b). All three diagrams represent contributions of order α to the energy. No other connected diagram can be found at this order beyond those three. The diagrams with closed loops represent radiative QED corrections.

To perform the time integration, Eq. (185) is expressed as a function of the relative time $t_2 - t_1$. One can then establish for a variable z and $\Delta E_{ij} = E_i - E_j$,

$$
\int dt_2\int dt_1 e^{-\epsilon(|t_2|+|t_1|)}e^{-iz(t_2-t_1)}e^{i(\Delta E_{nm})t_2}e^{i(\Delta E_{lk})t_1}
$$

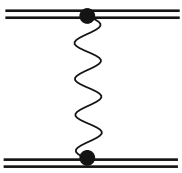


Fig. 5 Electron-electron interaction Feynman diagram

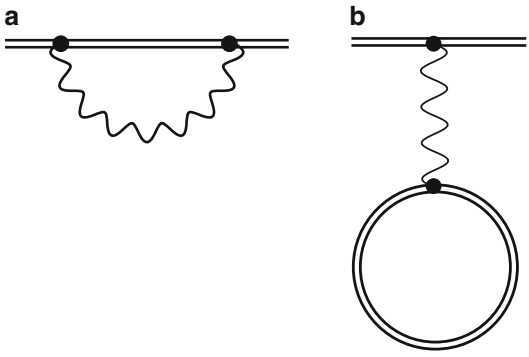


Fig. 6 One-photon, one-electron Feynman diagrams (**a**): self-energy. (**b**): vacuum polarization. The double line represents the Coulomb-bound electron propagator; the wavy line represents the photon propagator

$$= \frac{2\epsilon}{(z+\Delta E_{nm})^2+\epsilon^2}\frac{2\epsilon}{(z-\Delta E_{lk})^2+\epsilon^2}$$

$$= \delta\left(\Delta E_{nm}, -\Delta E_{lk}\right)\frac{2\pi}{\epsilon}\delta\left(z+\Delta E_{nm}\right)+\mathcal{O}(\epsilon) \tag{186}$$

$$= \delta\left(E_n+E_k, E_l+E_n\right)$$
$$\times\frac{1}{\epsilon}\int d\left(t_2-t_1\right)e^{-iz(t_2-t_1)}e^{-i(\Delta E_{nm})t_2}e^{i(\Delta E_{lk})t_1} \tag{187}$$

Equation (187) is then explicit in the relative time.

Contributions of Order α^2

One can then go to the next order of Eq. (113). The third order does not provide any physical contributions, as all odd orders. The next physical contribution is thus provided by $4\left\langle S^{(4)}_{\epsilon,1}\right\rangle - 4\left\langle S^{(1)}_{\epsilon,1}\right\rangle\left\langle S^{(3)}_{\epsilon,1}\right\rangle - 2\left\langle S^{(2)}_{\epsilon,1}\right\rangle^2$. It provides contributions of order α^2 to the energy. In this case there can be two-loop, one-electron diagrams, one-loop two-electron diagrams, and 0-loop three-electron diagrams, which are connected. The one-electron diagrams are presented in Fig. 7. Figure 7a–c represents the two-loop self-energy contribution. They must be evaluated simultaneously to obtain a finite contribution. Figure 7d, e represents mixed vacuum polarization, self-energy (SE-VP diagrams) Fig. 7f mixed vacuum polarization, self-energy (SVPE) diagram); and Fig. 7g, h, i two-loop vacuum polarization diagrams. Figure 7g and h are the all-order version of the Källén-Sabry approximation, in which an expansion of the bound propagator of the loop has been performed. This term also contains two-electron contributions. The first one represented in Fig. 8a, b corresponds to the first electron-electron correlation contributions and its non-radiative QED corrections (the part corresponding to contributions from the negative energies in the ladder diagram Fig. 8a and all of the crossed diagram Fig. 8b). The other contributions Fig. 8c to Fig. 8f are radiative corrections to the electron-electron interaction, often called screened QED contributions. At this order there is also a new possibility, a three-body correction involving three-electrons and two photons, as presented on Fig. 9.

It has required a lot of effort, in the recent years, to finally have an almost complete evaluation of the two-loop diagrams of Fig. 7. Mixed vacuum polarization and self-energy diagrams have been evaluated in several works [65, 66]. The most difficult part has been the two-loop self-energy diagrams evaluation to all orders [67–75]. Both the analytic part (regularization and renormalization) and the numerical part required extensive work, and the low-Z evaluation to all orders is yet to come. For the other diagrams, partial calculations have been performed, often by doing an expansion on the electron-loop bound propagator. Some diagrams have led to potentials in this approximation, like the well-known Källén-Sabry potential [76] for which very accurate numerical evaluations can be performed using, for example [77], approximate expressions for the potential.

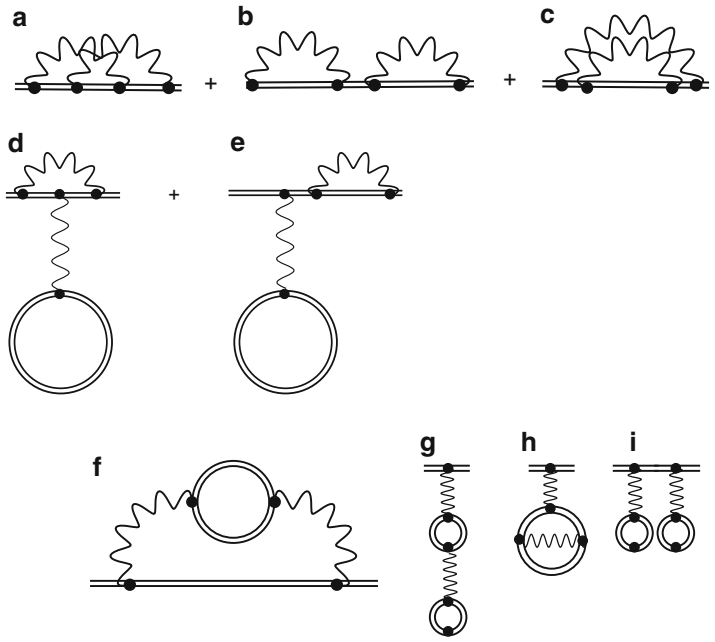


Fig. 7 Two-photon, one-electron Feynman diagrams. (**a**)–(**c**): two-loop self-energy. (**d**), (**e**): mixed vacuum polarization – self-energy (SE-VP diagrams). (**f**): mixed vacuum polarization – self-energy (SVPE diagram). (**g**)–(**i**): two-loop vacuum polarization. (**g**) and (**h**) are the all-order version of the Källén-Sabry contribution

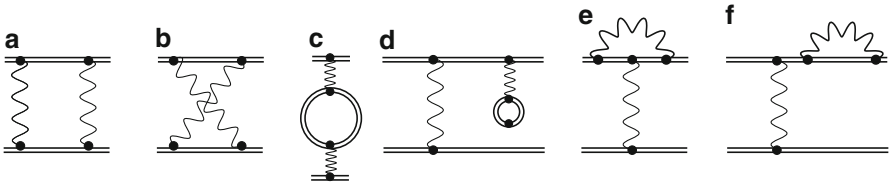


Fig. 8 Two-photon two-electron radiative Feynman diagrams. (**a**), (**b**): two-electron-two-photon interaction (first-order correlation energy and associated non-radiative QED correction). (**c**), (**d**): vacuum polarization correction to the electron-electron interaction (vacuum-polarization screening). (**e**), (**f**): self-energy correction to the electron-electron interaction (self-energy screening)

The screened QED diagrams of Fig. 8c–f have also been the object of many calculations. The vacuum polarization screening has been evaluated in Ref. [78, 79] for two-electron ions and in Ref. [80] for three-electron ions. The self-energy screening has been first evaluated using the method presented in section "Self-Energy Screening" [81]. It has been evaluated for two- and three-electron systems [82].

The two-photon ladder and crossed-ladder diagrams to all orders in $Z\alpha$ have been evaluated by several groups [83–88] using either unperturbed orbitals or orbitals modified by screening. Accurate two-photon exchange has also been

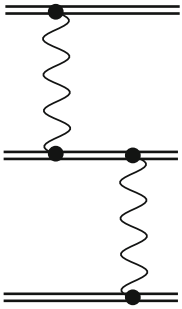


Fig. 9 Two-photon three-electron Feynman diagram

evaluated for two- [5] and three-electron [6, 86, 88, 89] systems. Usually some approximate many-body method is used to include three- and more photon exchange in the ladder approximation as the convergence is only going like $1/Z$ rather than in power of α for the ladder approximation correlation terms.

Reference [89] is based on the S-matrix formalism presented here, which has been extended recently to more complex ions with a valence electron outside a core [90].

Evaluation of the One-Loop Self-Energy

Introduction

There are several methods of practical calculation for radiative corrections. Here we show in detail how to treat in practice the one-electron self-energy diagram in coordinate space [27]. We show in detail the different aspects of the calculation, on how to perform regularization and renormalization.

We choose this example as the largest BSQED effect is in general the lowest-order self-energy correction. Even in the simple case of an electron bound in a pure Coulomb field, the precise theoretical evaluation of this correction has required extensive developments and numerical calculations.

Since the first attempts, a number of methods of evaluating the one-electron self-energy to all orders have been proposed [29, 91–93], and numerical calculations based on these formulations have been done recently [28,94–96]. Specific numerical developments for low-Z have been necessary, as the numerical accuracy needed to have accurate-enough results becomes very high [97,98]. Convergence acceleration techniques as developed in Ref. [99] where the key to obtain the low-Z, high-accuracy results of Refs. [97,98].

A necessary feature of any such work is the subtraction of the infinite mass renormalization in such a way that a numerical evaluation of the remainder is feasible. Here we follow Ref. [27], in which is described a method that has been applied to practical calculations of the one-electron self-energy for $n = 1$ and $n = 2$

states, in [100], to higher-n states in [96], and to the self-energy screening evaluation in [81].

It provides a formulation of the self-energy calculation for an arbitrary spherically symmetric local binding potential in which the infinities associated with mass renormalization are removed by subtractions made in coordinate space. It uses for that a term-by-term subtraction method, which has the advantage that it does not require a mix of coordinate-space and momentum-space calculations as do earlier methods, but works entirely within coordinate space.

The self-energy in Eq. (185c) is rewritten as

$$\Delta E_n(\Lambda) = \frac{\alpha}{2\pi i}\int_{C_F} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1\, \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu G(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^\mu\phi_n(\mathbf{x}_1) \times\left(\frac{e^{-bx_{21}}}{x_{21}}-\frac{e^{-b'x_{21}}}{x_{21}}\right)-\delta m\int d\mathbf{x}\,\phi_n^\dagger(\mathbf{x})\beta\phi_n(\mathbf{x}), \tag{188}$$

where b as explained in section "Photon Field Operators" is $b=-i\sqrt{(E_n-z)^2+i\epsilon}$, with $\mathrm{Re}(b) > 0$ and $b' = -i\sqrt{(E_n-z)^2-\Lambda^2+i\epsilon}$, with $\mathrm{Re}(b') > 0$. Λ corresponds to a mass added to the photon propagator. The contour C_F has been defined in Fig. 3.

In the original paper [29], the QED expression for the self-energy correction, ΔE_{SE}, is then written as the sum $\Delta E_{SE} = \Delta E_L + \Delta E_H$ of a low-energy part ΔE_L and a high-energy part ΔE_H, in units in which $\hbar = c = m_e = 1$. This is obtained by deformation of the contour in Fig. 3 as presented in Fig. 10. Using the analytic properties of the different parts of Eq. (188), and letting z_1 and z_2 go to 0, one can have a final integration path with C_H along the vertical axis and C_L a line between 0 and the E_N pole.

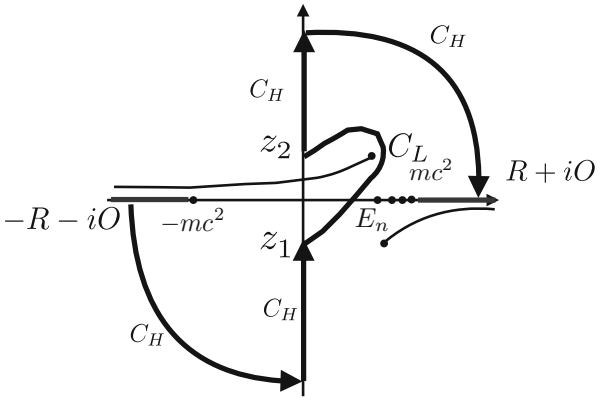


Fig. 10 Integration contour C_L and C_H for the low- and high-energy part of the self-energy

The high- and low-energy parts then read

$$\Delta E_L = \frac{\alpha}{\pi} E_n - \frac{\alpha}{\pi} P \int_0^{E_n} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1 \, \phi_n^\dagger(\mathbf{x}_2) \alpha^l G(\mathbf{x}_2, \mathbf{x}_1, z) \alpha^m \phi_n(\mathbf{x}_1) \times (\delta_{lm} \boldsymbol{\nabla}_2 \cdot \boldsymbol{\nabla}_1 - \nabla_2^l \nabla_1^m) \frac{\sin[(E_n - z)x_{21}]}{(E_n - z)^2 x_{21}} \tag{189}$$

and

$$\Delta E_H = \frac{\alpha}{2\pi i} \int_{C_H} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1 \, \phi_n^\dagger(\mathbf{x}_2) \alpha_\mu G(\mathbf{x}_2, \mathbf{x}_1, z) \alpha^\mu \phi_n(\mathbf{x}_1) \frac{e^{-bx_{21}}}{x_{21}} - \delta m \int d\mathbf{x} \, \phi_n^\dagger(\mathbf{x}) \beta \phi_n(\mathbf{x}), \tag{190}$$

where $\mathbf{x}_{21} = \mathbf{x}_2 - \mathbf{x}_1$ (see definitions and notations in section "Electron Field Propagator"). The indices l and m are summed from 1 to 3, and the index μ is summed from 0 to 3. The contour C_H extends from $-i\infty$ to $0 - i\epsilon$ and from $0 + i\epsilon$ to $+i\infty$, with the appropriate branch of b chosen in each case. Notational details are defined in section "Notations."

The expressions in (189) and (190) contain no assumptions concerning the external electrostatic potential except that the spectrum of the Dirac Hamiltonian resembles the Coulomb spectrum in a way that is evident from the derivation in Ref. [29]. For example, it can correspond to a Coulomb potential with a finite nuclear size correction or to a screened potential. In addition the orders of magnitude of the various Dirac operators for bound states of the external field must be comparable to their counterparts in a Coulomb field so that the identification of lower-order and higher-order terms presented below has validity.

The low-energy part in Eq. (189) is finite, so no subtraction is necessary. Yet, it contains a finite part that contributes to the renormalization, and the physical level shift is smaller by a factor of order $(Z\alpha)^4$ than the apparent order of the contribution. In order to identify the physically relevant part of (189), we explicitly calculate the lower-order terms in section "Low-Energy Part."

The difficult part of the work is to render the high-energy part finite. Here we show how to carry out the isolation of the divergent terms and the mass renormalization in coordinate space. A function is formulated that allows the subtraction of the lower-order terms and the infinities associated with mass renormalization point-wise, prior to carrying out the numerical integrations. Such a subtraction is not necessary to obtain a finite result, but it is found that early removal of the divergent terms improves significantly the convergence of the numerical integrations. This approach enables to renormalize the self-energy diagram whether it is embedded in a higher-order Feynman diagram [81] or for the regular self-energy [96, 100]. It also allows to make calculation without the need to know an explicit form of the Fourier transform of the wave functions. It has recently been extended to vacuum polarization [101].

Low-Energy Part

The low-energy part of the self-energy is finite and can be calculated very accurately even at low-Z, although it is of apparent order αmc^2.

Methods of evaluation of the complete low-energy part are described in Refs. [28–30].

Those calculations can be applied for pure Coulomb field as well as to a broad range of spherically symmetric external fields. The general expression for the real part of the low-energy part is [29]

$$\Delta E_L = \frac{\alpha}{\pi} E_n - \frac{\alpha}{\pi} P \int_0^{E_n} dz \int_0^{\infty} dx_2 x_2^2 \int_0^{\infty} dx_1 x_1^2 \times \sum_{\kappa} \sum_{i,j=1}^{2} f_{\bar{i}}(x_2) G_{\kappa}^{ij}(x_2, x_1, z) f_{\bar{j}}(x_1) A_{\kappa}^{ij}(x_2, x_1), \tag{191}$$

where f_i are the components of the radial wave function, $\bar{i} = 3 - i$, G_{κ}^{ij} are the components of the radial Green functions, and A_{κ}^{ij} are functions that arise from the photon propagator. Detailed definitions of the notation are given in Ref. [29]. The explicit low-order parts is given in Ref. [29] for a Coulomb field, and the corresponding generalization is derived in this section. We identify the lower-order terms not only to isolate the physical part of the correction but also to compare them to the terms that they must cancel in the high-energy part. The identification of the lower-order terms is based on order of magnitudes given in section "Operators Mean Value for the Dirac Equation in a Coulomb Field," which are valid for an electron in a Coulomb field, but the calculations are general for an arbitrary external field.

To calculate the terms of order lower than $(Z\alpha)^4$, we rewrite the low-energy part in terms of operators as done in Ref. [29]

$$\Delta E_L = \frac{\alpha}{\pi} E_n - \frac{\alpha}{4\pi^2} P \int_{k<E_n} d\mathbf{k} \frac{1}{k} \left(\delta_{lm} - \frac{k^l k^m}{k^2} \right) \times \left\langle \alpha^l \frac{1}{\boldsymbol{\alpha}\cdot\mathbf{p} - \boldsymbol{\alpha}\cdot\mathbf{k} + V + \beta - E_n + k - i\delta} \alpha^m \right\rangle \tag{192}$$

and expand the right-hand side of (192), neglecting terms that are higher order than first in V and $1 - E_n$ or higher order than second in $\boldsymbol{\alpha}\cdot\mathbf{p}$

$$\frac{1}{\boldsymbol{\alpha}\cdot\mathbf{p} - \boldsymbol{\alpha}\cdot\mathbf{k} + V + \beta - E_n + k - i\delta} = -\left[\frac{1}{2E_n k} - \frac{1 - E_n^2}{4k^2} \right] (\boldsymbol{\alpha}\cdot\mathbf{k} - \beta - E_n + k)$$

$$-\frac{1}{4k^2}(\boldsymbol{\alpha}\cdot\mathbf{k}-\beta-E_n+k)\,(\boldsymbol{\alpha}\cdot\mathbf{p}+V)(\boldsymbol{\alpha}\cdot\mathbf{k}-\beta-E_n+k)$$
$$-\frac{1}{8k^3}(\boldsymbol{\alpha}\cdot\mathbf{k}-\beta-E_n+k)\,\boldsymbol{\alpha}\cdot\mathbf{p}\,(\boldsymbol{\alpha}\cdot\mathbf{k}-\beta-E_n+k)\boldsymbol{\alpha}\cdot\mathbf{p}$$
$$\times(\boldsymbol{\alpha}\cdot\mathbf{k}-\beta-E_n+k)+\cdots. \tag{193}$$

Neglected terms in this expansion contribute corrections of order $(Z\alpha)^4\ln(Z\alpha)^{-2}$ or higher. By dropping terms that are odd in $\mathbf{k}$, and taking into account the fact that

$$(E_n+\beta)\alpha^l|n\rangle=\alpha^l(E_n-\beta)|n\rangle=\alpha^l(\boldsymbol{\alpha}\cdot\mathbf{p}+V)|n\rangle \tag{194}$$

in estimating orders of magnitude, we find that the right-hand-side of (193) has the same leading terms as

$$\frac{1}{2E_nk}(\beta+E_n-k)+\frac{1-E_n^2}{4k}$$
$$-\frac{1}{4k^2}[2\boldsymbol{\alpha}\cdot\mathbf{k}\,\mathbf{p}\cdot\mathbf{k}-2k\boldsymbol{\alpha}\cdot\mathbf{p}+2k^2V+2(\mathbf{p}\cdot\mathbf{k})^2-kp^2]+\cdots. \tag{195}$$

Substituting (195) into (192), and integrating over $\mathbf{k}$, we obtain

$$\Delta E_L=\frac{\alpha}{\pi}\left[\frac{1}{2}E_n+\langle\beta\rangle-\frac{1}{2}(1-E_n^2)+\frac{1}{6}\langle\boldsymbol{\alpha}\cdot\mathbf{p}\rangle+\frac{1}{2}\langle V\rangle-\frac{1}{3}\langle p^2\rangle\right.$$
$$\left.+\mathcal{O}\big((Z\alpha)^4\ln(Z\alpha)^{-2}\big)\right]. \tag{196}$$

The replacements, based on section "Operators Mean Value for the Dirac Equation in a Coulomb Field" ,

$$1-E_n^2\to\langle\beta\rangle-E_n-\langle V\rangle+\mathcal{O}\big((Z\alpha)^4\big) \tag{197}$$
$$\langle\boldsymbol{\alpha}\cdot\mathbf{p}\,\rangle\to E_n-\langle\beta\rangle-\langle V\rangle \tag{198}$$
$$\langle p^2\rangle\to\langle\boldsymbol{\alpha}\cdot\mathbf{p}\rangle+\mathcal{O}\big((Z\alpha)^4\big) \tag{199}$$

in (196) give

$$\Delta E_L=\frac{\alpha}{\pi}\left[\frac{5}{6}E_n+\frac{2}{3}\langle\beta\rangle+\frac{7}{6}\langle V\rangle+\mathcal{O}\big((Z\alpha)^4\ln(Z\alpha)^{-2}\big)\right]. \tag{200}$$

The physical part of the low-energy part is then contained in the function $f_L(Z\alpha)$ defined by

$$\Delta E_L = \frac{\alpha}{\pi}\left[\frac{5}{6}E_n + \frac{2}{3}\langle\beta\rangle + \frac{7}{6}\langle V\rangle + \frac{(Z\alpha)^4}{n^3} f_L(Z\alpha)\right]. \tag{201}$$

In a pure Coulomb field, $E_n = \langle\beta\rangle$, and this result coincides with Ref. [29].

This direct expansion procedure correctly isolates the lower-order terms because the low-energy part is written here in Coulomb gauge. This fact was recognized early by Kroll and Lamb [35] who employed such an expansion to obtain the lowest-order Lamb shift. This procedure would not work in the Feynman gauge, as it leads to an expansion that does not converge because of the presence of infrared photon contributions. An alternative gauge was identified by Fried and Yennie in 1958 [102], in which spurious low-order terms also vanish.

The low-energy part of the self- energy is calculated by evaluating the complete expression for ΔE_L in (191) and solving numerically Eq. (201) to yield $f_L(Z\alpha)$. In the cases where Z is small, this procedure entails substantial numerical cancellation, but those problems have been solved successfully in Ref. [97, 98].

Regularization

The high-energy part given in Eq. (190) is only a formal expression, as each of the two terms that contribute to it is infinite. Only the difference is finite. To make the terms independently finite, one must employ a scheme such as dimensional regularization (see, e.g., [103, 104] and applications in Refs. [15, 105]) or Pauli-Villars regularization [106], which we employ here. It is implemented by making the replacement

$$\frac{1}{q^2 + i\delta} \to \frac{1}{q^2 + i\delta} - \frac{1}{q^2 - \Lambda^2 + i\delta} \tag{202}$$

in the momentum-space expression for the photon propagator. In coordinate space, this corresponds to the replacement

$$\frac{e^{-bx_{21}}}{x_{21}} \to \frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}}, \tag{203}$$

where

$$b' = -i\left[(E_n - z)^2 - \Lambda^2 + i\delta\right]^{1/2}; \quad \mathrm{Re}(b') > 0. \tag{204}$$

The mass renormalization term δm is obtained by the corresponding term calculated with the regularized photon propagator with the result

$$\delta m \to \delta m(\Lambda) = \frac{\alpha}{\pi}\left(\frac{3}{4}\ln(\Lambda^2) + \frac{3}{8}\right). \tag{205}$$

The full regularized expression is then

$$\Delta E_H = \lim_{\Lambda\to\infty}\left\{\frac{\alpha}{2\pi i}\int_{C_H} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1\, \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu G(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^\mu\phi_n(\mathbf{x}_1) \times\left(\frac{e^{-bx_{21}}}{x_{21}}-\frac{e^{-b'x_{21}}}{x_{21}}\right)-\delta m(\Lambda)\int d\mathbf{x}\,\phi_n^\dagger(\mathbf{x})\beta\phi_n(\mathbf{x})\right\}, \tag{206}$$

which, when combined to the low-energy part ΔE_L, yields the finite self-energy contribution.

Singular Terms: Renormalization in Coordinate Space

Although the integration over z in Eq. (206) is decreasing exponentially at large values of the argument $|z|$, when $x_{21} \neq 0$, the unregulated integral is infinite, since the point $x_{21} = 0$ is included in the coordinate-space integration. To extract the divergent terms, we examine the integrand in the immediate vicinity of $x_{21} = 0$ and employ three expansions around this point. First we employ the expansion of the Dirac Green function in power of $V(x)$ given in (206) and presented in Fig. 4.

Higher-order terms not included in Eq. (135), when substituted into Eq. (206), yield a finite contribution of order $(Z\alpha)^4$. Then we use an expansion of the bound-state wave function in Eq. (206):

$$\phi_n(\mathbf{x}_1) = \phi_n(\mathbf{x}_2) + (\mathbf{x}_1-\mathbf{x}_2)\cdot\boldsymbol{\nabla}_2\,\phi_n(\mathbf{x}_2) + \frac{1}{2}(x_1^l-x_2^l)(x_1^m-x_2^m)\frac{\partial}{\partial x_2^l}\frac{\partial}{\partial x_2^m}\phi_n(\mathbf{x}_2)+\cdots, \tag{207}$$

where the indices l, m are summed from 1 to 3. This expansion is made to take into account the fact that the dominant contribution to the function at large $|z|$ comes from the region where $x_{21} \approx 0$. This series expansion of the wave function, designed to isolate the renormalization terms, has been employed in Ref. [107, 108], to calculate the electromagnetic self-energy of quarks in a cavity.

The series in Eq. (207) corresponds to an asymptotic expansion in $|z|^{-1}$ for the integrand of the integral over z in (206). This will be illustrated by an example in section "Asymptotic Properties." It also corresponds to an expansion in powers of $\mathbf{p}$ in some sense, of power of $Z\alpha$.

We now come to the third expansion,

$$V(\mathbf{x}_3) = V(\mathbf{x}_2)+\cdots \tag{208}$$

in the second term on the right-hand side of Eq. (135). Higher-order terms in Eq. (208) would correspond to commutators of V and F. We have now to regroup the terms corresponding to the first term in Eq. (135) and third term in Eq. (207),

together with the second term in (135), the first term in Eq. (207), and the term in Eq. (208). This will regroup all the divergent contributions and, in the Coulomb case, all the parts of order lower than $(Z\alpha)^4$ in Eq. (206). The combination of the first term of Eq. (135) with the third term of Eq. (207) is finite, but it is of order $p^2 \sim (Z\alpha)^2$. It is removed to improve the numerical accuracy at low-Z.

The expansion of the high-energy part provides the correct power series in $(Z\alpha)$ up to order $(Z\alpha)^4$, even though its expression is written in the Feynman gauge. Yet spurious low-order terms that should appear in the expansion of the complete self-energy in Feynman gauge are absent, because the soft-photon contributions that would lead to these terms appear only in the low-energy part.

Evaluation of the Singular Terms

The singular terms defined above are denoted by $\Delta E_H^{(i,j)}$, where i indicates the order in V and j indicates the order in the power series in Eq. (207).

The leading term in this expansion is

$$\Delta E_H^{(0,0)} = \frac{\alpha}{2\pi i} \int_{C_H} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1 \, \phi_n^\dagger(\mathbf{x}_2) \alpha_\mu F(\mathbf{x}_2, \mathbf{x}_1, z) \alpha^\mu \phi_n(\mathbf{x}_2) \times \left(\frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}} \right), \tag{209}$$

where the free Green function is defined by

$$F(\mathbf{x}_2, \mathbf{x}_1, z) = \lim_{Z\alpha \to 0} G(\mathbf{x}_2, \mathbf{x}_1, z). \tag{210}$$

This function is given explicitly by (see Eq. (133))

$$F(\mathbf{x}_2, \mathbf{x}_1, z) = \left[\left(\frac{c}{x_{21}} + \frac{1}{x_{21}^2} \right) i\boldsymbol{\alpha} \cdot \mathbf{x}_{21} + \beta + z \right] \frac{e^{-cx_{21}}}{4\pi x_{21}}, \tag{211}$$

where $c = \left(1 - z^2\right)^{1/2}$, $\mathrm{Re}(c) > 0$. In Eq. (209), integration over $\mathbf{x}_1$ is elementary

$$\int d\mathbf{x}_1 \, F(\mathbf{x}_2, \mathbf{x}_1, z) \left(\frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}} \right) = (\beta + z) \left(\frac{1}{b + c} - \frac{1}{b' + c} \right). \tag{212}$$

We obtain

$$\Delta E_H^{(0,0)} = \frac{\alpha}{\pi i} \int_{C_H} dz \ \langle 2\beta - z \rangle \left(\frac{1}{b+c} - \frac{1}{b'+c} \right). \tag{213}$$

Integration over z yields as shown in section "Integration of Singular Terms"

$$\begin{aligned} \Delta E_H^{(0,0)} &= \frac{\alpha}{\pi} \Big\{ \langle \beta \rangle \left[\ln(\Lambda^2) - 1 + \frac{1-E_n^2}{E_n^2} \ln\left(1+E_n^2\right) \right] \\ &\quad - E_n \left[\frac{1}{4} \ln(\Lambda^2) + \frac{3E_n^2-2}{8E_n^2} + \frac{1-E_n^4}{4E_n^4} \ln\left(1+E_n^2\right) \right] + \mathcal{O}(\Lambda^{-1}) \Big\}. \end{aligned} \tag{214}$$

The second term is expressed as

$$\begin{aligned} \Delta E_H^{(0,1)} &= \frac{\alpha}{2\pi i} \int_{C_H} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1 \ \phi_n^\dagger(\mathbf{x}_2) \alpha_\mu F(\mathbf{x}_2, \mathbf{x}_1, z) \alpha^\mu \\ &\quad \times (\mathbf{x}_1 - \mathbf{x}_2) \cdot \boldsymbol{\nabla} \ \phi_n(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}_2} \left(\frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}} \right). \end{aligned} \tag{215}$$

Integration over $\mathbf{x}_1$ yields

$$\begin{aligned} &\int d\mathbf{x}_1 \ F(\mathbf{x}_2, \mathbf{x}_1, z)(\mathbf{x}_1 - \mathbf{x}_2) \left(\frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}} \right) = \\ &\quad -\frac{i}{3} \boldsymbol{\alpha} \left[\frac{1}{b+c} - \frac{1}{b'+c} + \frac{c}{(b+c)^2} - \frac{c}{(b'+c)^2} \right]. \end{aligned} \tag{216}$$

This leads to

$$\begin{aligned} \Delta E_H^{(0,1)} &= \frac{\alpha}{\pi i} \int_{C_H} dz \, \frac{1}{3} \langle \boldsymbol{\alpha} \cdot \mathbf{p} \rangle \\ &\quad \times \left[\frac{1}{b+c} - \frac{1}{b'+c} + \frac{c}{(b+c)^2} - \frac{c}{(b'+c)^2} \right], \end{aligned} \tag{217}$$

and to the final result

$$\begin{aligned} \Delta E_H^{(0,1)} &= \frac{\alpha}{\pi} \langle \boldsymbol{\alpha} \cdot \mathbf{p} \rangle \left[\frac{1}{4} \ln(\Lambda^2) - \frac{6 - 3E_n^2 + 7E_n^4}{24E_n^2(1+E_n^2)} \right. \\ &\quad \left. + \frac{1-E_n^4}{4E_n^4} \ln\left(1+E_n^2\right) + \mathcal{O}(\Lambda^{-1}) \right]. \end{aligned} \tag{218}$$

The third term given by

$$\Delta E_H^{(0,2)} = \frac{\alpha}{2\pi i}\int_{C_H} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1\, \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu F(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^\mu \times\frac{1}{2}(x_1^l - x_2^l)(x_1^m - x_2^m)\frac{\partial}{\partial x^l}\frac{\partial}{\partial x^m}\phi_n(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_2} \times\left(\frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}}\right) \tag{219}$$

(using the convention of summation over repeated indices) is evaluated using

$$\int d\mathbf{x}_1\, F(\mathbf{x}_2,\mathbf{x}_1,z)(x_1^l - x_2^l)(x_1^m - x_2^m)\left(\frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}}\right) = \frac{2}{3}\delta_{lm}(\beta + z)\left[\frac{1}{(b+c)^3} - \frac{1}{(b'+c)^3}\right], \tag{220}$$

which leads to

$$\Delta E_H^{(0,2)} = -\frac{\alpha}{\pi i}\int_{C_H} dz\, \frac{1}{3}\left\langle (2\beta - z)p^2\right\rangle\left[\frac{1}{(b+c)^3} - \frac{1}{(b'+c)^3}\right], \tag{221}$$

and finally to

$$\Delta E_H^{(0,2)} = \frac{\alpha}{\pi}\left\{\left\langle \beta p^2\right\rangle\left[-\frac{3 + 6E_n^2 - E_n^4}{3E_n^2(1+E_n^2)^2} + \frac{1}{E_n^4}\ln\left(1+E_n^2\right)\right] + \left\langle E_n p^2\right\rangle\left[\frac{6 + 3E_n^2 + E_n^4}{12E_n^4(1+E_n^2)} - \frac{1}{2E_n^6}\ln\left(1+E_n^2\right)\right] + \mathcal{O}(\Lambda^{-1})\right\} \tag{222}$$

The fourth and last term is the correction linear in V

$$\Delta E_H^{(1,0)} = -\frac{\alpha}{2\pi i}\int_{C_H} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1\, \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu \int d\mathbf{x}_3 F(\mathbf{x}_2,\mathbf{x}_3,z) \times V(\mathbf{x}_2)F(\mathbf{x}_3,\mathbf{x}_1,z)\alpha^\mu\phi_n(\mathbf{x}_2)\left(\frac{e^{-bx_{21}}}{x_{21}} - \frac{e^{-b'x_{21}}}{x_{21}}\right). \tag{223}$$

Using the operator identity

$$\frac{1}{H_0 - z}\,\frac{1}{H_0 - z} = \left[\frac{\partial}{\partial\epsilon}\frac{1}{H_0 - z - \epsilon}\right]_{\epsilon=0}, \tag{224}$$

the integral over the product of Fs is given by

$$\int d\mathbf{x}_3\; F(\mathbf{x}_2,\mathbf{x}_3,z)F(\mathbf{x}_3,\mathbf{x}_1,z) = \left[\frac{\partial}{\partial\epsilon}F(\mathbf{x}_2,\mathbf{x}_1,z+\epsilon)\right]_{\epsilon=0}, \tag{225}$$

leading to

$$\begin{aligned}\Delta E_H^{(1,0)} = &-\frac{\alpha}{2\pi i}\int_{C_H} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1\; \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu V(\mathbf{x}_2) \\ &\times\left[\frac{\partial}{\partial\epsilon}F(\mathbf{x}_2,\mathbf{x}_1,z+\epsilon)\right]_{\epsilon=0}\alpha^\mu\phi_n(\mathbf{x}_2)\left(\frac{e^{-bx_{21}}}{x_{21}}-\frac{e^{-b'x_{21}}}{x_{21}}\right).\end{aligned} \tag{226}$$

Integration over $\mathbf{x}_1$ as expressed in Eq. (212), followed by differentiation with respect to ϵ, provides

$$\begin{aligned}\Delta E_H^{(1,0)} = \frac{\alpha}{\pi i}\int_{C_H} dz \Bigg\{ &\langle V\rangle\left(\frac{1}{b+c}-\frac{1}{b'+c}\right) \\ &-\langle V(2\beta-z)\rangle\,\frac{z}{c}\left[\frac{1}{(b+c)^2}-\frac{1}{(b'+c)^2}\right]\Bigg\},\end{aligned} \tag{227}$$

and

$$\begin{aligned}\Delta E_H^{(1,0)} = \frac{\alpha}{\pi}\Bigg\{ &\langle V\rangle\left[\frac{1}{4}\ln(\Lambda^2)+\frac{6-E_n^2}{8E_n^2}-\frac{3+E_n^4}{4E_n^4}\ln\left(1+E_n^2\right)\right] \\ &-\left\langle\frac{\beta}{E_n}V\right\rangle\left[2-\frac{2}{E_n^2}\ln\left(1+E_n^2\right)\right]+\mathcal{O}(\Lambda^{-1})\Bigg\}.\end{aligned} \tag{228}$$

We combine these four terms to obtain the high-energy analytic part ΔE_{HA}, defined by

$$\Delta E_{HA} = \lim_{\Lambda\to\infty}\left[\Delta E_H^{(0,0)}+\Delta E_H^{(0,1)}+\Delta E_H^{(0,2)}+\Delta E_H^{(1,0)}-\delta m(\Lambda)\,\langle\beta\rangle\right], \tag{229}$$

where the last term is the renormalization term from Eq. (190). We use Dirac equation to replace $\langle\boldsymbol{\alpha}\cdot\mathbf{p}\,\rangle$ by $E_n-\langle\beta\rangle-\langle V\rangle$, to obtain

$$\begin{aligned}\Delta E_{HA} = \frac{\alpha}{\pi}\Bigg\{ &E_n\left[-\frac{2E_n^2}{3(1+E_n^2)}\right] \\ &+\langle\beta\rangle\left[\frac{3-18E_n^2-13E_n^4}{12E_n^2(1+E_n^2)}-\frac{(1-E_n^2)(1-3E_n^2)}{4E_n^4}\ln\left(1+E_n^2\right)\right] \\ &+\langle V\rangle\left[\frac{6+3E_n^2+E_n^4}{6E_n^2(1+E_n^2)}-\frac{1}{E_n^4}\ln\left(1+E_n^2\right)\right]\end{aligned}$$

$$+\left\langle \frac{\beta}{E_n} V \right\rangle \left[-2 + \frac{2}{E_n^2} \ln\left(1 + E_n^2\right) \right]$$
$$+\left\langle E_n p^2 \right\rangle \left[\frac{6 + 3E_n^2 + E_n^4}{12E_n^4(1 + E_n^2)} - \frac{1}{2E_n^6} \ln\left(1 + E_n^2\right) \right]$$
$$+\left\langle \beta p^2 \right\rangle \left[-\frac{3 + 6E_n^2 - E_n^4}{3E_n^2(1 + E_n^2)^2} + \frac{1}{E_n^4} \ln\left(1 + E_n^2\right) \right] \Bigg\}. \tag{230}$$

The divergent terms in Eq. (230) sum up to zero as expected. The terms in p^2 can be eliminated with the aid of the Dirac equation identities given in section "Operators Mean Value for the Dirac Equation in a Coulomb Field":

$$\left\langle E_n p^2 \right\rangle = E_n(E_n^2 - 1) - 2E_n^2 \langle V \rangle + \left\langle E_n V^2 \right\rangle \tag{231}$$
$$\left\langle \beta p^2 \right\rangle = \langle \beta \rangle (E_n^2 - 1) - 2 \langle V \rangle - \left\langle \beta V^2 \right\rangle. \tag{232}$$

We obtain

$$\Delta E_{HA} = \frac{\alpha}{\pi} \Bigg\{ E_n \left[-\frac{6 - 9E_n^2 + 7E_n^4}{12E_n^4} + \frac{1 - E_n^2}{2E_n^6} \ln\left(1 + E_n^2\right) \right]$$
$$+ \langle \beta \rangle \left[\frac{(5 + 9E_n^2)(3 - 6E_n^2 - E_n^4)}{12E_n^2(1 + E_n^2)^2} \right.$$
$$\left. - \frac{(1 - E_n^2)(5 - 3E_n^2)}{4E_n^4} \ln\left(1 + E_n^2\right) \right]$$
$$+ \langle V \rangle \left[2\frac{3 + 6E_n^2 - E_n^4}{3E_n^2(1 + E_n^2)^2} - \frac{2}{E_n^4} \ln\left(1 + E_n^2\right) \right]$$
$$+ \left\langle \frac{\beta}{E_n} V \right\rangle \left[-2 + \frac{2}{E_n^2} \ln\left(1 + E_n^2\right) \right]$$
$$+ \left\langle E_n V^2 \right\rangle \left[\frac{6 + 3E_n^2 + E_n^4}{12E_n^4(1 + E_n^2)} - \frac{1}{2E_n^6} \ln\left(1 + E_n^2\right) \right]$$
$$+ \left\langle \beta V^2 \right\rangle \left[\frac{3 + 6E_n^2 - E_n^4}{3E_n^2(1 + E_n^2)^2} - \frac{1}{E_n^4} \ln\left(1 + E_n^2\right) \right] \Bigg\}. \tag{233}$$

In the general case, i.e., not assuming the Coulomb identity $\langle \beta \rangle = E_n$, we isolate the parts of ΔE_{HA} that are of order lower than $(Z\alpha)^4$ by expanding in powers of $1 - E_n^2$, assuming the potential leads to orders of magnitude identical to the Coulomb potential:

$$\Delta E_{HA} = \frac{\alpha}{\pi} \Bigg\{ E_n \left[-\frac{1}{3} + \frac{2\ln 2 - 1}{4} \left(1 - E_n^2\right) \right]$$

$$+\langle\beta\rangle\left[-\frac{7}{6}+\frac{3-2\ln 2}{4}\left(1-E_n^2\right)\right]$$
$$+\langle V\rangle\left[\frac{4}{3}-2\ln 2\right]+\left\langle\frac{\beta}{E_n}V\right\rangle\left[2\ln 2-2\right]$$
$$+\mathcal{O}\left((Z\alpha)^4\right)\Bigg\}. \tag{234}$$

This expression is simplified by using the expressions given in section "Operators Mean Value for the Dirac Equation in a Coulomb Field":

$$E_n(1-E_n^2) = \langle\beta\rangle - E_n - \langle V\rangle + \mathcal{O}\left((Z\alpha)^4\right) \tag{235}$$
$$\langle\beta\rangle\,(1-E_n^2) = \langle\beta\rangle - E_n - \langle V\rangle + \mathcal{O}\left((Z\alpha)^4\right) \tag{236}$$
$$\left\langle\frac{\beta}{E_n}V\right\rangle = \langle V\rangle + \mathcal{O}\left((Z\alpha)^4\right) \tag{237}$$

leading to

$$\Delta E_{HA} = \frac{\alpha}{\pi}\left[-\frac{5}{6}E_n - \frac{2}{3}\langle\beta\rangle - \frac{7}{6}\langle V\rangle + \mathcal{O}\left((Z\alpha)^4\right)\right]. \tag{238}$$

We do observe here the expected cancellation of the leading terms between the high- and low-energy parts.

Since the leading terms cancel in the total self-energy, it is convenient to express the remainder of the analytic part in terms of a function $f_A(Z\alpha)$ defined by

$$\Delta E_{HA} = \frac{\alpha}{\pi}\left[-\frac{5}{6}E_n - \frac{2}{3}\langle\beta\rangle - \frac{7}{6}\langle V\rangle + \frac{(Z\alpha)^4}{n^3}f_A(Z\alpha)\right]. \tag{239}$$

Equation (233) leads immediately to

$$\frac{(Z\alpha)^4}{n^3}f_A(Z\alpha) = \frac{1}{2}\langle E_n-\beta\rangle + E_n(1-E_n^2)\left[-\frac{2-E_n^2}{4E_n^4}+\frac{1}{2E_n^6}\ln\left(1+E_n^2\right)\right]$$
$$+\langle\beta\rangle\,(1-E_n^2)\left[\frac{15+26E_n^2-5E_n^4}{12E_n^2(1+E_n^2)^2}-\frac{5-3E_n^2}{4E_n^4}\ln\left(1+E_n^2\right)\right]$$
$$+\langle V\rangle\left[\frac{(3+7E_n^2)(4+E_n^2+E_n^4)}{6E_n^2(1+E_n^2)^2}-\frac{2}{E_n^4}\ln\left(1+E_n^2\right)\right]$$
$$+\left\langle\frac{\beta}{E_n}V\right\rangle\left[-2+\frac{2}{E_n^2}\ln\left(1+E_n^2\right)\right]$$
$$+\left\langle E_nV^2\right\rangle\left[\frac{6+3E_n^2+E_n^4}{12E_n^4(1+E_n^2)}-\frac{1}{2E_n^6}\ln\left(1+E_n^2\right)\right]$$

$$+\langle\beta V^2\rangle\left[\frac{3+6E_n^2-E_n^4}{3E_n^2(1+E_n^2)^2}-\frac{1}{E_n^4}\ln\left(1+E_n^2\right)\right]. \tag{240}$$

The first contribution on the right-hand side is isolated to reduce numerical cancellations for small values of $Z\alpha$.

Evaluation of the Subtraction Terms

The singular terms identified in the preceding section will be expressed here as functions of z, x_2, and x_1 to allow for a point-wise numerical subtraction to obtain the best possible accuracy. We calculate the remainder named ΔE_{HB} by point-wise subtraction of the singular terms obtained from the complete high-energy part

$$\Delta E_{HB} = \Delta E_H - \Delta E_{HA}. \tag{241}$$

In this way the numerical integration formula is applied to both terms, so that the dominant integration errors largely cancel. The integral of the difference is finite by construction, so the expression is automatically regularized and no subtraction term is required. The functions $K^{(i,j)}(x_2, x_1, z)$ provide the point by point values of the integrands of the singular terms:

$$\Delta E_H^{(i,j)} = \frac{\alpha}{2\pi i}\int_{C_H} dz \int_0^\infty dx_2\, x_2^2 \int_0^\infty dx_1\, x_1^2\, K^{(i,j)}(x_2, x_1, z). \tag{242}$$

The first term, Eq. (209), corresponds to the integral

$$K^{(0,0)}(x_2, x_1, z) = \int d\Omega_2 \int d\Omega_1\, \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu F(\mathbf{x}_2, \mathbf{x}_1, z)\alpha^\mu \phi_n(\mathbf{x}_2)\frac{e^{-bx_{21}}}{x_{21}}. \tag{243}$$

The integral over Ω_1 is expressed as (see section "Angular Integrations")

$$\int d\Omega_1\, \alpha_\mu\ \ F(\mathbf{x}_2, \mathbf{x}_1, z)\alpha^\mu \frac{e^{-bx_{21}}}{x_{21}} = \int_{-1}^1 d\xi\, \frac{e^{-(b+c)R}}{R^2} \times\left[\left(\frac{c}{R}+\frac{1}{R^2}\right) i\boldsymbol{\alpha}\cdot\hat{x}_2(x_2-\xi x_1)+2\beta-z\right], \tag{244}$$

where

$$R = (x_2^2 - 2x_2x_1\xi + x_1^2)^{1/2}. \tag{245}$$

Using

$$\phi_n^\dagger(\mathbf{x})\boldsymbol{\alpha}\cdot\hat{x}\phi_n(\mathbf{x}) = 0, \tag{246}$$

we have

$$K^{(0,0)}(x_2, x_1, z) = \int d\Omega_2\, \phi_n^\dagger(\mathbf{x}_2) \int_{-1}^{1} d\xi\, \frac{e^{-(b+c)R}}{R^2}(2\beta - z)\phi_n(\mathbf{x}_2), \tag{247}$$

and hence

$$K^{(0,0)}(x_2, x_1, z) = F_n^T(x_2) \int_{-1}^{1} d\xi\, \frac{e^{-(b+c)R}}{R^2}(2\beta - z) F_n(x_2), \tag{248}$$

where the notation is defined in section "Angular Integrations."

The second term is given by

$$\begin{aligned} K^{(0,1)}(x_2, x_1, z) \;=& \int d\Omega_2 \int d\Omega_1 \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu \\ & \times F(\mathbf{x}_2, \mathbf{x}_1, z)\alpha^\mu(\mathbf{x}_1 - \mathbf{x}_2)\cdot\nabla\, \phi_n(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_2} \frac{e^{-bx_{21}}}{x_{21}}, \end{aligned} \tag{249}$$

with the associated angular integral (see section "Angular Integrations").

$$\begin{aligned} & \int d\Omega_1\, \alpha_\mu F(\mathbf{x}_2, \mathbf{x}_1, z)\alpha^\mu (x_1^m - x_2^m)\frac{e^{-bx_{21}}}{x_{21}} \\ &= -\int_{-1}^{1} d\xi\, \frac{e^{-(b+c)R}}{R^2} \\ &\quad \times \left\{ \left(\frac{c}{R} + \frac{1}{R^2}\right) i\alpha^l \left[R^2 \hat{x}_2^l \hat{x}_2^m + \tfrac{1}{2}(1-\xi^2)x_1^2(\delta_{lm} - 3\hat{x}_2^l\hat{x}_2^m)\right] \right. \\ &\qquad \left. + (2\beta - z)(x_2 - \xi x_1)\hat{x}_2^m \right\}. \end{aligned} \tag{250}$$

Using the identities (see section "Notations")

$$\int d\Omega\, \phi_n^\dagger(\mathbf{x}) i(\boldsymbol{\alpha}\cdot\hat{x})(\hat{x}\cdot\nabla)\phi_n(\mathbf{x}) = -F_n^T(x)\alpha\beta\frac{\partial}{\partial x}F_n(x) \tag{251}$$

$$\int d\Omega\, \phi_n^\dagger(\mathbf{x}) i\boldsymbol{\alpha}\cdot\nabla\phi_n(\mathbf{x}) = -F_n^T(x)\alpha\beta\left(\frac{\partial}{\partial x} + \beta\frac{\kappa}{x}\right)F_n(x) \tag{252}$$

$$\int d\Omega\, \phi_n^\dagger(\mathbf{x})(2\beta - z)\hat{x}\cdot\nabla\phi_n(\mathbf{x}) = F_n^T(x)(2\beta - z)\frac{\partial}{\partial x}F_n(x) \tag{253}$$

we obtain the expression

$$\begin{aligned} K^{(0,1)}(x_2, x_1, z) = F_n^T(x_2) \int_{-1}^{1} d\xi\, \frac{e^{-(b+c)R}}{R^2}\Bigg[& (cR+1)\,\alpha\beta\frac{\partial}{\partial x_2} \\ & - \left(\frac{c}{R} + \frac{1}{R^2}\right)(1-\xi^2)x_1^2\alpha\beta\left(\frac{\partial}{\partial x_2} - \beta\frac{\kappa}{2x_2}\right) \end{aligned}$$

$$-(x_2-\xi x_1)(2\beta-z)\frac{\partial}{\partial x_2}\Big]F_n(x_2). \tag{254}$$

We then turn to the term

$$K^{(0,2)}(x_2,x_1,z)=\int d\Omega_2\int d\Omega_1\phi_n^\dagger(\mathbf{x}_2)\alpha_\mu F(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^\mu\frac{1}{2}(x_1^l-x_2^l)$$
$$\times(x_1^m-x_2^m)\frac{\partial}{\partial x^l}\frac{\partial}{\partial x^m}\phi_n(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_2}\frac{e^{-bx_{21}}}{x_{21}}, \tag{255}$$

of which we retain only the diagonal part

$$K_D^{(0,2)}(x_2,x_1,z)=\int d\Omega_2\int d\Omega_1\,\phi_n^\dagger(\mathbf{x}_2)\alpha_\mu F_D(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^\mu\frac{1}{2}(x_1^l-x_2^l)$$
$$\times(x_1^m-x_2^m)\frac{\partial}{\partial x^l}\frac{\partial}{\partial x^m}\phi_n(\mathbf{x})\Big|_{\mathbf{x}=\mathbf{x}_2}\frac{e^{-bx_{21}}}{x_{21}}, \tag{256}$$

which corresponds to the diagonal part of the free Green function

$$F_D(\mathbf{x}_2,\mathbf{x}_1,z)=(\beta+z)\frac{e^{-cx_{21}}}{4\pi x_{21}}. \tag{257}$$

The off-diagonal term is not considered as it gives no contribution to the final result, as can be seen from

$$\int_0^\infty dx_1\,x_1^2\Big[K^{(0,2)}(x_2,x_1,z)-K_D^{(0,2)}(x_2,x_1,z)\Big]=0. \tag{258}$$

We then perform the angle integrals

$$\int d\Omega_1\quad\alpha_\mu F_D(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^\mu(x_1^l-x_2^l)(x_1^m-x_2^m)\,\frac{e^{-bx_{21}}}{x_{21}}$$
$$=\int_{-1}^1 d\xi\,\frac{e^{-(b+c)R}}{R^2}(2\beta-z)$$
$$\times\left[R^2\hat{x}_2^l\hat{x}_2^m+\tfrac{1}{2}(1-\xi^2)x_1^2(\delta_{lm}-3\hat{x}_2^l\hat{x}_2^m)\right], \tag{259}$$

and

$$\int d\Omega\;\phi_n^\dagger(\mathbf{x})(2\beta-z)(\hat{x}\cdot\boldsymbol{\nabla})^2\;\;\phi_n(\mathbf{x})=F_n^T(x)(2\beta-z)\frac{\partial^2}{\partial x^2}F_n(x) \tag{260}$$

$$\int d\Omega\;\phi_n^\dagger(\mathbf{x})(2\beta-z)\nabla^2\phi_n(\mathbf{x})\;\;=F_n^T(x)(2\beta-z)$$
$$\times\left(\frac{1}{x}\frac{\partial^2}{\partial x^2}x-\frac{\kappa(\kappa+\beta)}{x^2}\right)F_n(x) \tag{261}$$

which combined leads to

$$K_D^{(0,2)}(x_2,x_1,z) = F_n^T(x_2)\int_{-1}^{1} d\xi\ \frac{e^{-(b+c)R}}{2R^2}(2\beta - z) \times\left[R^2\frac{\partial^2}{\partial x_2^2} - (1-\xi^2)x_1^2\left(\frac{\partial^2}{\partial x_2^2} - \frac{1}{x_2}\frac{\partial}{\partial x_2} + \frac{\kappa(\kappa+\beta)}{2x_2^2}\right)\right] \times F_n(x_2). \tag{262}$$

The next term, which is proportional to V, is given by

$$\begin{aligned} K^{(1,0)}(x_2,x_1,z) = & -\int d\Omega_2 \int d\Omega_1\ \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu \\ & \times \int d\mathbf{x}_3 F(\mathbf{x}_2,\mathbf{x}_3,z)V(\mathbf{x}_2)F(\mathbf{x}_3,\mathbf{x}_1,z)\alpha^\mu\phi_n(\mathbf{x}_2)\frac{e^{-bx_{21}}}{x_{21}} \\ = & -\int d\Omega_2 \int d\Omega_1\ \phi_n^\dagger(\mathbf{x}_2)\alpha_\mu V(\mathbf{x}_2) \\ & \times\left[\frac{\partial}{\partial\epsilon}F(\mathbf{x}_2,\mathbf{x}_1,z+\epsilon)\right]_{\epsilon=0}\alpha^\mu\phi_n(\mathbf{x}_2)\frac{e^{-bx_{21}}}{x_{21}}. \end{aligned} \tag{263}$$

Using Eqs. (243) to (248), we obtain after differentiating with respect to ϵ and taking into account the spherical symmetry of $V(\mathbf{x})$:

$$K^{(1,0)}(x_2,x_1,z) = F_n^T(x_2)\int_{-1}^{1} d\xi\ \frac{e^{-(b+c)R}}{R^2}V(x_2)\left[1-(2\beta-z)R\frac{z}{c}\right]F_n(x_2), \tag{264}$$

The complete subtraction term $K_A(x_2,x_1,z)$ is then given by the sum of the individual terms obtained above:

$$\begin{aligned} K_A(x_2,x_1,z) = & K^{(0,0)}(x_2,x_1,z) + K^{(0,1)}(x_2,x_1,z) \\ & + K_D^{(0,2)}(x_2,x_1,z) + K^{(1,0)}(x_2,x_1,z) \end{aligned} \tag{265}$$

where

$$K^{(0,0)}(x_2,x_1,z) = F_n^T(x_2)Q_1(x_2,x_1,z)(2\beta-z)F_n(x_2); \tag{266}$$

$$\begin{aligned} K^{(0,1)}(x_2,x_1,z) = & F_n^T(x_2)\left[Q_2(x_2,x_1,z)\alpha\beta\frac{\partial}{\partial x_2} - Q_3(x_2,x_1,z)x_1^2\right. \\ & \times\alpha\beta\left(\frac{\partial}{\partial x_2} - \beta\frac{\kappa}{2x_2}\right) \\ & \left. - Q_4(x_2,x_1,z)(2\beta-z)\frac{\partial}{\partial x_2}\right]F_n(x_2); \end{aligned} \tag{267}$$

$$K_D^{(0,2)}(x_2,x_1,z) = F_n^T(x_2)\left(\beta-\frac{z}{2}\right) \times\Bigg[Q_5(x_2,x_1,z)\frac{\partial^2}{\partial x_2^2} - Q_6(x_2,x_1,z)x_1^2 \times\left(\frac{\partial^2}{\partial x_2^2}-\frac{1}{x_2}\frac{\partial}{\partial x_2}+\frac{\kappa(\kappa+\beta)}{2x_2^2}\right)\Bigg]F_n(x_2); \tag{268}$$

and

$$K^{(1,0)}(x_2,x_1,z) = F_n^T(x_2)V(x_2)\Bigg[Q_1(x_2,x_1,z) -(2\beta-z)\frac{z}{c}Q_7(x_2,x_1,z)\Bigg]F_n(x_2). \tag{269}$$

Finally we eliminate the derivatives of the radial wave function from Eqs. (267)) and (268) using the radial differential Dirac equation giving

$$\frac{\partial}{\partial x}F_n(x) = \left[\alpha(1-\beta E_n)-(1+\beta\kappa)\frac{1}{x}+\alpha\beta V(x)\right]F_n(x) \tag{270}$$

and

$$\frac{\partial^2}{\partial x^2}F_n(x) = \left[\alpha(1-\beta E_n)-(1+\beta\kappa)\frac{1}{x}+\alpha\beta V(x)\right]\frac{\partial}{\partial x}F_n(x) + \left[(1+\beta\kappa)\frac{1}{x^2}+\alpha\beta\frac{\partial V(x)}{\partial x}\right]F_n(x). \tag{271}$$

The seven functions Q_i used above are given by

$$Q_1(x_2,x_1,z) = \int_{-1}^{1} d\xi\,\frac{e^{-(b+c)R}}{R^2} = P_{-2}; \tag{272}$$

$$Q_2(x_2,x_1,z) = \int_{-1}^{1} d\xi\,\frac{e^{-(b+c)R}}{R^2}(cR+1) = cP_{-1}+P_{-2}; \tag{273}$$

$$\begin{aligned} Q_3(x_2,x_1,z) &= \int_{-1}^{1} d\xi\,\frac{e^{-(b+c)R}}{R^4}(cR+1)(1-\xi^2) \\ &= \frac{1}{(2x_2x_1)^2}\big[2(x_2^2+x_1^2)(cP_{-1}+P_{-2})-(cP_1+P_0) \\ &\quad -(x_2^2-x_1^2)^2(cP_{-3}+P_{-4})\big]; \end{aligned} \tag{274}$$

$$Q_4(x_2, x_1, z) = \int_{-1}^{1} d\xi \, \frac{e^{-(b+c)R}}{R^2}(x_2 - \xi x_1) = \frac{1}{2x_2}\left[P_0 + (x_2^2 - x_1^2)P_{-2}\right]; \tag{275}$$

$$Q_5(x_2, x_1, z) = \int_{-1}^{1} d\xi \, e^{-(b+c)R} = P_0; \tag{276}$$

$$\begin{aligned} Q_6(x_2, x_1, z) &= \int_{-1}^{1} d\xi \, \frac{e^{-(b+c)R}}{R^2}(1 - \xi^2) \\ &= \frac{1}{(2x_2x_1)^2}\left[2(x_2^2 + x_1^2)P_0 - P_2 - (x_2^2 - x_1^2)^2 P_{-2}\right]; \end{aligned} \tag{277}$$

$$Q_7(x_2, x_1, z) = \int_{-1}^{1} d\xi \, \frac{e^{-(b+c)R}}{R} = P_{-1}, \tag{278}$$

where the functions Q_i are expressed as functions of the integrals P_i defined by

$$P_i = \int_{-1}^{1} d\xi \, R^i e^{-(b+c)R} = \frac{1}{x_2x_1} \int_{|x_2 - x_1|}^{x_2 + x_1} dR \, R^{(i+1)} e^{-(b+c)R}. \tag{279}$$

In these functions the integration over R is relatively convenient for analytic or numerical evaluation.

We now rewrite the modified high-energy part (265) in terms of the function $K_A(x_2, x_1, z)$

$$\begin{aligned} \Delta E_{HB} = \frac{\alpha}{2\pi i} \int_{C_H} dz \int_0^\infty dx_2 \, x_2^2 \int_0^\infty dx_1 \, x_1^2 \\ \times \Bigg\{ \sum_\kappa \sum_{i,j=1}^{2} \Big[f_i(x_2) G_\kappa^{ij}(x_2, x_1, z) f_j(x_1) A_\kappa(x_2, x_1) \\ - f_{\bar{i}}(x_2) G_\kappa^{ij}(x_2, x_1, z) f_{\bar{j}}(x_1) A_\kappa^{ij}(x_2, x_1) \Big] - K_A(x_2, x_1, z) \Bigg\}, \end{aligned} \tag{280}$$

using the same notations as in Eq. (191) except that the functions A_κ^{ij} are defined differently, as explained in Ref. [29]. We define the corresponding numerical function $f_{HB}(Z\alpha)$ as

$$\Delta E_{HB} = \frac{\alpha}{\pi} \frac{(Z\alpha)^4}{n^3} f_{HB}(Z\alpha). \tag{281}$$

The total high-energy part is thus given by the sum

$$f_H(Z\alpha) = f_{HA}(Z\alpha) + f_{HB}(Z\alpha), \tag{282}$$

with $f_{HA}(Z\alpha)$ given by Eq. (240) and $f_{HB}(Z\alpha)$ by Eq. (280) combined with Eq. (281).

Results and Discussion

The total self-energy is given by

$$\Delta E_n = \frac{\alpha}{\pi}\frac{(Z\alpha)^4}{n^3}F(Z\alpha)m_e c^2 \tag{283}$$

where

$$F(Z\alpha) = f_L(Z\alpha) + f_H(Z\alpha) \tag{284}$$

from (201) and (282).

The foregoing discussion provides a prescription for removing the renormalization terms from the bound-state self-energy diagram in such a way that a complete numerical evaluation is feasible for a broad class of external potentials. The renormalization subtraction is carried out before numerical integration over the two radial coordinates or integration over the intermediate-state energy parameter, so numerical integration errors are not amplified by the subtraction process. The fact that the subtraction term is relatively simple is particularly beneficial in the case where the self-energy diagram is embedded in a more complex Feynman diagram. On the other hand, the subtraction term does not cancel term by term in the summation over intermediate angular momentum in (280), so the slow convergence of that sum is not improved by the subtraction. This factor can be compensated in the numerical evaluation by the introduction of a high-order asymptotic expansion in the sum to carry out an analytic summation of the leading orders [100].

Vacuum Polarization

In this section we follow [101] to perform the renormalization of vacuum polarization, in the Uehling approximation. We start here from Eq. (185b), which correspond to the all-order vacuum polarization. This term was first studied in detail by Whichman and Kroll in 1956 [109]. We integrate over the time difference and obtain

$$E_{\rm VP}^{(2)} = -\alpha \int d\mathbf{x}_2 \int d\mathbf{x}_1 \frac{1}{|\mathbf{x}_2 - \mathbf{x}_1|} \times \mathrm{Tr}\left[\gamma_\mu S_{\rm F}(x_2, x_2)\right] \overline{\phi}_n(\mathbf{x}_1)\gamma^\mu \phi_n(\mathbf{x}_1)\,. \tag{285}$$

where $e\mathrm{Tr}\left[\gamma_0 S_{\rm F}(x_2, x_2)\right]$ plays the role of a charge density. The electron propagation function is given in Eq. (130). Replacing in Eq. (285), one obtains

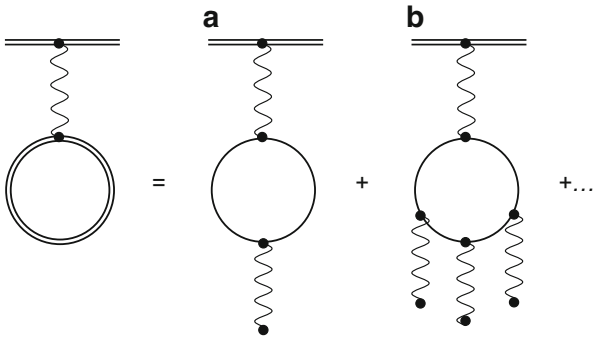


Fig. 11 Expansion of the one-loop vacuum polarization in terms of the free electron propagator (*single line*). The dashed line corresponds to the Coulomb interaction with the nucleus. (**a**): Uehling potential. (**b**): Wichmann and Kroll correction. The two-interaction term, not represented here, vanishes because of Furry's theorem

$$E_{\mathrm{VP}}^{(2)} = \frac{\mathrm{i}\alpha}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}z\int\mathrm{d}\mathbf{x}_2\int\mathrm{d}\mathbf{x}_1\,\frac{1}{|\mathbf{x}_2-\mathbf{x}_1|} \times\Big\{\mathrm{Tr}\left[G\left(\mathbf{x}_2,\mathbf{x}_2,z(1+\mathrm{i}\delta)\right)\right]\phi_n^\dagger(\mathbf{x}_1)\,\phi_n(\mathbf{x}_1) -\mathrm{Tr}\left[\boldsymbol{\alpha}\,G\left(\mathbf{x}_2,\mathbf{x}_2,z(1+\mathrm{i}\delta)\right)\right]\cdot\phi_n^\dagger(\mathbf{x}_1)\,\boldsymbol{\alpha}\,\phi_n(\mathbf{x}_1)\Big\}. \tag{286}$$

The second term vanishes for a spherically symmetric potential, provided infinities are regularized.

The bound electron propagator can be expanded in a series of the free propagator and potential $V(\mathbf{x})$ as in represented in Fig. 4 and expressed in Eq. (135). Contrary to self-energy, these corrections can be expressed as potentials. The two first contributions, corresponding to the Uehling [110] and Wichmann and Kroll potentials [109], are presented in Fig. 11. Each potential can be evaluated numerically to a high level of accuracy [77]. Only the diagrams with an even number of photon insertions are non-zero, due to Furry's theorem [55].

Regularization of the Vacuum Polarization

Zero Potential Contribution

The first term in the expansion when replacing the full Green function depends on the free Green function trace:

$$\mathrm{Tr}\left[F\left(\mathbf{x}_2,\mathbf{x}_1,z\right)\right] = \frac{z\,\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{x}_1|}}{\pi\,|\mathbf{x}_2-\mathbf{x}_1|}\,. \tag{287}$$

This term should vanish, except for the divergent behavior at $\mathbf{x}_2 \approx \mathbf{x}_1$. This divergence is removed by using Pauli-Villars regularization. The free Green function

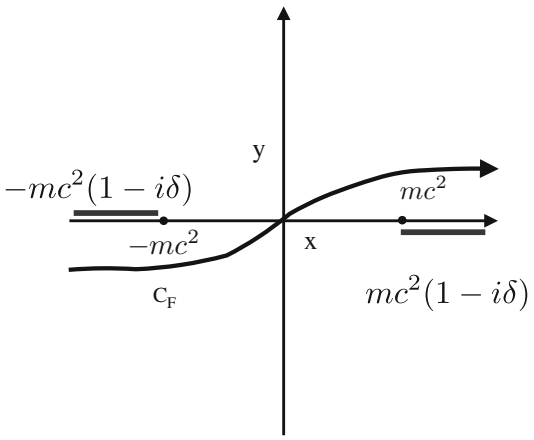


Fig. 12 Integration path for the free Green function. The function is analytic in z and has two branch points located at $z = \pm mc^2(1 - \mathrm{i}\delta)$

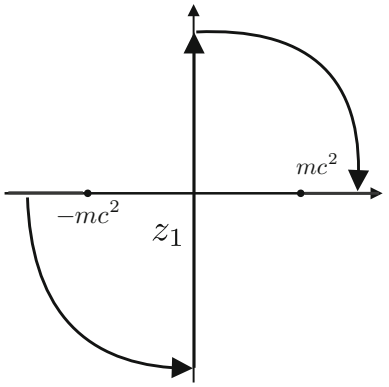


Fig. 13 Integration path for the free Green function to cancel out infinities in the zero-potential term

is analytic in z and it has branch point at $z = \pm m(1 - \mathrm{i}\delta)$. We can modify the integration contour, as presented in Fig. 12 to the one presented in Fig. 13. The contributions from the two quarter circles will vanish when the radius of the circles is enlarged provided the integrand of the regularized function decreases faster than $1/|z|$ for large $|z|$, if $\mathbf{x}_2 = \mathbf{x}_1$.

Taking $z = \mathrm{i}u$ as fully imaginary, we have

$$c_i = \sqrt{m_i^2 - z^2} = \sqrt{u^2} + \frac{m_i^2}{2\sqrt{u^2}} - \frac{m_i^4}{8u^2\sqrt{u^2}} + \dots , \tag{288}$$

so that the leading terms in the expansion in $1/u$ are defined by

$$\frac{\mathrm{e}^{-c_i|\mathbf{x}_2-\mathbf{x}_1|}}{|\mathbf{x}_2-\mathbf{x}_1|} = \frac{\mathrm{e}^{-\sqrt{u^2}|\mathbf{x}_2-\mathbf{x}_1|}}{|\mathbf{x}_2-\mathbf{x}_1|} - \frac{m_i^2\,\mathrm{e}^{-\sqrt{u^2}|\mathbf{x}_2-\mathbf{x}_1|}}{2\sqrt{u^2}}. \tag{289}$$

The two terms on the right-hand side of Eq. (289) give divergent contributions. We eliminate them by introducing two auxiliary propagators with masses m_1 and m_2 with the constraint

$$1 + C_1 + C_2 = 0\,, \tag{290}$$

$$m_0^2 + C_1 m_1^2 + C_2 m_2^2 = 0\,, \tag{291}$$

which is satisfied by

$$C_1 = \frac{m_0^2 - m_2^2}{m_2^2 - m_1^2}\,, \tag{292}$$

$$C_2 = \frac{m_1^2 - m_0^2}{m_2^2 - m_1^2}\,. \tag{293}$$

To regularize the expression, we define

$$F_{\mathrm{R}}(\mathbf{x}_2, \mathbf{x}_1, z) = \sum_{i=0}^{2} C_i\, F_i(\mathbf{x}_2, \mathbf{x}_1, z)\,, \tag{294}$$

with F_i that is defined by Eq. (133) where m is replaced by m_i. One obtains

$$\mathrm{Tr}\left[F_{\mathrm{R}}(\mathbf{x}_2, \mathbf{x}_1, z)\right] = \sum_{i=0}^{2} C_i\, \frac{\mathrm{i}u\, \mathrm{e}^{-c_i |\mathbf{x}_2 - \mathbf{x}_1|}}{\pi\, |\mathbf{x}_2 - \mathbf{x}_1|}\,, \tag{295}$$

and

$$\begin{aligned}
\mathrm{Tr}\left[F_{\mathrm{R}}(\mathbf{x}_2, \mathbf{x}_2, z)\right] &= \lim_{|\mathbf{x}_2 - \mathbf{x}_1| \to 0} \mathrm{Tr}\left[F_{\mathrm{R}}(\mathbf{x}_2, \mathbf{x}_1, z)\right] \\
&= -\frac{\mathrm{i}u}{\pi} \sum_{i=0}^{2} C_i c_i \\
&= \frac{\mathrm{i}}{8\pi u \sqrt{u^2}} \left[\sum_{i=0}^{2} C_i m_i^4 + \mathcal{O}\left(\frac{1}{u^2}\right)\right].
\end{aligned} \tag{296}$$

We see here that the contribution from the quarter circles shown in Fig. 13 vanishes, and since the branches of the square root are specified to give $\sqrt{u^2} = |u|$ for real values of u, the integrand is an odd function of u, and we have

$$\int_{-\infty}^{\infty} \mathrm{d}u\, \mathrm{Tr}\left[F_{\mathrm{R}}(\mathbf{x}_2, \mathbf{x}_2, \mathrm{i}\,u)\right] = 0\,. \tag{297}$$

One Potential Contribution: The Uehling Potential

When substituting the development from Eq. (135) in Eq. (286), and selecting the term with a single potential, one gets

$$G^{(1)}\left(\mathbf{x}_2, \mathbf{x}_1, z\right) = -\int \mathrm{d}\mathbf{w}\, F\left(\mathbf{x}_2, \mathbf{w}, z\right) V(\mathbf{w}) F\left(\mathbf{w}, \mathbf{x}_1, z\right) . \tag{298}$$

First we rewrite $\mathrm{Tr}\left[G^{(1)}\left(\mathbf{x}_2, \mathbf{x}_1, z\right)\right]$ to isolate the singularity in $\mathbf{x}_2 = \mathbf{x}_1$, obtaining a simple term with the aid of section "Details on the Evaluation of the One-Potential Green Function":

$$\begin{aligned}\mathrm{Tr}\left[G^{(1)}\left(\mathbf{x}_2, \mathbf{x}_1, z\right)\right] = &-\frac{Z\alpha}{2\pi}\int \mathrm{d}\mathbf{w}\, \frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\,\rho(\mathbf{w})\,\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|} \\ &-\frac{z^2}{2\pi^2}\int \mathrm{d}\mathbf{w}\, \frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\,V(\mathbf{w})\,\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|} \\ &-\frac{1}{2\pi}\left[V(\mathbf{x}_2)+V(\mathbf{x}_1)\right]\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{x}_1|}}{|\mathbf{x}_2-\mathbf{x}_1|} ,\end{aligned} \tag{299}$$

where Z is the charge of the nucleus and ρ is the nuclear charge density normalized to 1. The limit of $\mathrm{Tr}\left[G^{(1)}\left(\mathbf{x}_2, \mathbf{x}_1, z\right)\right]$ when $\mathbf{x}_2 \to \mathbf{x}_1$ is singular due to the third term, so we introduce a counter term. It is designed to cancel the singularity point by point and goes to zero when integrated over z. It is given by

$$G_{\mathrm{A}}^{(1)}(\mathbf{x}_2, \mathbf{x}_1, z) = -\int \mathrm{d}\mathbf{w}\, F(\mathbf{x}_2, \mathbf{w}, z)\,\frac{V(\mathbf{x}_2)+V(\mathbf{x}_1)}{2}\,F(\mathbf{w}, \mathbf{x}_1, z) , \tag{300}$$

and (see section "Details on the Evaluation of the One-Potential Green Function")

$$\begin{aligned}\mathrm{Tr}\left[G_{\mathrm{A}}^{(1)}(\mathbf{x}_2, \mathbf{x}_1, z)\right] = &-\frac{z^2}{2\pi^2}\int \mathrm{d}\mathbf{w}\, \frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\,\frac{V(\mathbf{x}_2)+V(\mathbf{x}_1)}{2}\,\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|} \\ &-\frac{1}{2\pi}\left[V(\mathbf{x}_2)+V(\mathbf{x}_1)\right]\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{x}_1|}}{|\mathbf{x}_2-\mathbf{x}_1|} .\end{aligned} \tag{301}$$

The difference

$$G_{\mathrm{B}}^{(1)}(\mathbf{x}_2, \mathbf{x}_1, z) = G^{(1)}(\mathbf{x}_2, \mathbf{x}_1, z) - G_{\mathrm{A}}^{(1)}(\mathbf{x}_2, \mathbf{x}_1, z) \tag{302}$$

has the trace

$$\begin{aligned}\mathrm{Tr}\left[G_{\mathrm{B}}^{(1)}(\mathbf{x}_2,\mathbf{x}_1,z)\right] = &-\frac{Z\alpha}{2\pi}\int \mathrm{d}\mathbf{w}\,\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\,\rho(\mathbf{w})\,\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|}\\ &-\frac{z^2}{2\pi^2}\int \mathrm{d}\mathbf{w}\,\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\left[V(\mathbf{w})-\frac{V(\mathbf{x}_2)+V(\mathbf{x}_1)}{2}\right]\\ &\times\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|}\,. \end{aligned} \tag{303}$$

Taking the limit $\mathbf{x}_1 \to \mathbf{x}_2$, we obtain

$$\begin{aligned}\mathrm{Tr}\left[G_{\mathrm{B}}^{(1)}(\mathbf{x}_2,\mathbf{x}_2,z)\right] = &-\frac{Z\alpha}{2\pi}\int \mathrm{d}\mathbf{w}\,\rho(\mathbf{w})\,\frac{\mathrm{e}^{-2c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|^2}\\ &-\frac{z^2}{2\pi^2}\int \mathrm{d}\mathbf{w}\,\left[V(\mathbf{w})-V(\mathbf{x}_2)\right]\\ &\times\frac{\mathrm{e}^{-2c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|^2}\,. \end{aligned} \tag{304}$$

To show that the integral over the counter term vanishes, we consider the expansion in Eq. (136) and Eq. (139) and use the same contour shown in Fig. 13 as before. We find in the regulated trace a nonzero leading term which is

$$\frac{\partial}{\partial z}\,\mathrm{Tr}\left[F_{\mathrm{R}}(\mathbf{x}_2,\mathbf{x}_2,z)\right] = \frac{\partial}{\partial u}\,\frac{1}{8\pi u\sqrt{u^2}}\left[\sum_{i=0}^{2} C_i m_i^4 + \mathcal{O}\left(\frac{1}{u^2}\right)\right]. \tag{305}$$

Finally we integrate, showing that the counter term vanishes:

$$\int_{-\infty}^{\infty}\mathrm{d}z\,\frac{\partial}{\partial z}\,\mathrm{Tr}\left[F_{\mathrm{R}}(\mathbf{x}_2,\mathbf{x}_1,z)\right] = \mathrm{Tr}\left[F_{\mathrm{R}}(\mathbf{x}_2,\mathbf{x}_1,\mathrm{i}u)\right]\Big|_{u=-\infty}^{u=\infty} = 0\,. \tag{306}$$

The subtracted one-potential contribution to the Green function is regularized to be

$$\begin{aligned}G_{\mathrm{BR}}^{(1)}(\mathbf{x}_2,\mathbf{x}_1,z) = &-\int \mathrm{d}\mathbf{w}\left[V(\mathbf{w})-\frac{V(\mathbf{x}_2)+V(\mathbf{x}_1)}{2}\right]\\ &\times\sum_{i=0}^{2} C_i\,F_i(\mathbf{x}_2,\mathbf{w},z)\,F_i(\mathbf{w},\mathbf{x}_1,z)\,, \end{aligned} \tag{307}$$

and the trace of the propagation function at equal coordinates is given by

$$\mathrm{Tr}\left[\gamma^0 S_{\mathrm{FBR}}^{(1)}(x_2,x_2)\right] = \frac{1}{2\pi\mathrm{i}}\int_{-\infty}^{\infty}\mathrm{d}z\,\mathrm{Tr}\left[G_{\mathrm{BR}}^{(1)}\big(\mathbf{x}_2,\mathbf{x}_2,z(1+\mathrm{i}\delta)\big)\right]. \tag{308}$$

Performing a contour rotation to the imaginary axis, and replacing z by iu, and noting that the contribution of the quarter circles as $|z| \to \infty$ vanishes due to the exponential falloff of the integrand, Eq. (308) becomes

$$\begin{aligned}\mathrm{Tr}\left[\gamma^0 S^{(1)}_{\mathrm{FBR}}(x_2,x_2)\right] = \int_0^\infty \mathrm{d}u \sum_{i=0}^{2} C_i \Bigg\{ &- \frac{Z\alpha}{2\pi^2} \int \mathrm{d}\mathbf{w}\, \rho(\mathbf{w})\, \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|^2} \\ &+ \frac{u^2}{2\pi^3} \int \mathrm{d}\mathbf{w}\, [V(\mathbf{w})-V(\mathbf{x}_2)]\, \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|^2} \Bigg\}. \end{aligned} \tag{309}$$

Looking at the singularity $|\mathbf{x}_2 - \mathbf{w}| \approx 0$, we find that

$$\begin{aligned} &\sum_{i=0}^{2} C_i \int \mathrm{d}\mathbf{w}\, \rho(\mathbf{w})\, \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|^2} \\ &\quad \to \rho(\mathbf{x}_2) \sum_{i=0}^{2} C_i \int \mathrm{d}\mathbf{w}\, \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|^2} \\ &\quad = \frac{3\pi}{4}\rho(\mathbf{x}_2) \sum_{i=0}^{2} C_i \frac{m_i^4}{u^5} + \mathcal{O}\left(\frac{1}{u^7}\right) \end{aligned} \tag{310}$$

and

$$\begin{aligned} &\sum_{i=0}^{2} C_i\, u^2 \int \mathrm{d}\mathbf{w}\, [V(\mathbf{w})-V(\mathbf{x}_2)]\, \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|^2} \\ &\quad \to \sum_{i=0}^{2} C_i \frac{u^2}{6} \left[\nabla_2^2 V(\mathbf{x}_2)\right] \int \mathrm{d}\mathbf{w}\, \mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|} \\ &\quad = \frac{5\pi^2 Z\alpha}{4}\, \rho(\mathbf{x}_2) \sum_{i=0}^{2} C_i \frac{m_i^4}{u^5} + \mathcal{O}\left(\frac{1}{u^7}\right). \end{aligned} \tag{311}$$

The integration over u is then convergent and performing the contour rotation was legitimate.

We now integrate by parts in Eq. (309) and obtain

$$\begin{aligned}\mathrm{Tr}\left[\gamma^0 S^{(1)}_{\mathrm{FBR}}(x_2,x_2)\right] = \int_0^\infty \mathrm{d}u \sum_{i=0}^{2} C_i \Bigg\{ &- \frac{Z\alpha u^2}{\pi^2 c_i} \int \mathrm{d}\mathbf{w}\, \rho(\mathbf{w})\, \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|} \\ &+ \frac{u^4}{3\pi^3 c_i} \int \mathrm{d}\mathbf{w}\, [V(\mathbf{w})-V(\mathbf{x}_2)]\, \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|} \Bigg\}. \end{aligned} \tag{312}$$

Here again, an estimation analogous to what was performed in Eqs. (310) and (311) shows that the integration over u converges and that the surface terms at $u = \infty$ from the partial integration vanish even for $|\mathbf{x}_2 - \mathbf{w}| \approx 0$. For the second term in Eq. (312), we have

$$\begin{aligned}\int \mathrm{d}\mathbf{w}\, [V(\mathbf{w}) - V(\mathbf{x}_2)] \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|} &= -Z\alpha \int \mathrm{d}\mathbf{w} \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r}) \\ &\quad \times \left(\frac{1}{|\mathbf{w}-\mathbf{r}|} - \frac{1}{|\mathbf{x}_2-\mathbf{r}|}\right) \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|} \\ &= \frac{\pi Z\alpha}{c_i^2} \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r}) \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{r}|}}{|\mathbf{x}_2-\mathbf{r}|} \end{aligned} \tag{313}$$

leading to

$$\begin{aligned}\mathrm{Tr}\left[\gamma^0 S^{(1)}_{\mathrm{FBR}}(x_2, x_2)\right] = \frac{Z\alpha}{\pi^2} \int_0^\infty \mathrm{d}u \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r}) \sum_{i=0}^{2} C_i \left(-\frac{u^2}{c_i} + \frac{u^4}{3c_i^3}\right) \\ \times \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{r}|}}{|\mathbf{x}_2-\mathbf{r}|}\,. \end{aligned} \tag{314}$$

We then obtain the one-potential level energy shift as

$$\begin{aligned}E^{(2,1)}_{\mathrm{VP}} = -\alpha \int \mathrm{d}\mathbf{x}_2 \int \mathrm{d}\mathbf{x}_1 \frac{1}{|\mathbf{x}_2-\mathbf{x}_1|} \mathrm{Tr}\left[\gamma^0 S^{(1)}_{\mathrm{FBR}}(x_2, x_2)\right] \\ \times \phi_n^\dagger(\mathbf{x}_1)\, \phi_n(\mathbf{x}_1)\,, \end{aligned} \tag{315}$$

which is the expectation value of a vacuum polarization potential given by

$$\begin{aligned}V^{(2,1)}_{\mathrm{VP}}(\mathbf{x}_1) = \frac{Z\alpha^2}{\pi^2} \int \mathrm{d}\mathbf{x}_2 \frac{1}{|\mathbf{x}_2-\mathbf{x}_1|} \int_0^\infty \mathrm{d}u \int \mathrm{d}\mathbf{r}\, \rho(\mathbf{r}) \sum_{i=0}^{2} C_i \left(\frac{u^2}{c_i} - \frac{u^4}{3c_i^3}\right) \\ \times \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{r}|}}{|\mathbf{x}_2-\mathbf{r}|}\,. \end{aligned} \tag{316}$$

Regrouping and isolating the terms depending on $\mathbf{x}_2$, we obtain

$$\int \mathrm{d}\mathbf{x}_2 \frac{1}{|\mathbf{x}_2-\mathbf{x}_1|} \frac{\mathrm{e}^{-2c_i|\mathbf{x}_2-\mathbf{r}|}}{|\mathbf{x}_2-\mathbf{r}|} = \frac{\pi}{c_i^2\,|\mathbf{x}_1-\mathbf{r}|} \left(1 - \mathrm{e}^{-2c_i|\mathbf{x}_1-\mathbf{r}|}\right) \tag{317}$$

The one-interaction potential is then

$$V_{\mathrm{VP}}^{(2,1)}(\mathbf{x}_1) = \frac{Z\alpha^2}{\pi}\int_0^\infty \mathrm{d}u \int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r}) \sum_{i=0}^{2} C_i \left(\frac{u^2}{c_i^3} - \frac{u^4}{3c_i^5}\right) \times \frac{1-\mathrm{e}^{-2c_i|\mathbf{x}_1-\mathbf{r}|}}{|\mathbf{x}_1-\mathbf{r}|}\,. \tag{318}$$

Performing the integral over u for the first term, we get

$$\int_0^\infty \mathrm{d}u \sum_{i=0}^{2} C_i \left(\frac{u^2}{c_i^3} - \frac{u^4}{3c_i^5}\right) = -\frac{1}{3}\sum_{i=0}^{2} C_i \ln m_i^2\,, \tag{319}$$

which produces a potential corresponding to a mass-dependent charge proportional to the charge distribution ρ.

We then proceed to eliminate this mass-dependent part by charge renormalization. For the $\exp(-2c_i\,|\mathbf{x}_1-\mathbf{r}|)$ term in Eq. (318), the contribution from each i is separately finite because of the exponential factor. We now evaluate the nonrelativistic estimate of the energy shift:

$$\begin{aligned} &-\frac{Z\alpha^2}{\pi}\,|\phi_n(0)|^2 \int \mathrm{d}\mathbf{x}_1 \int_0^\infty \mathrm{d}u \int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r})\, C_i \left(\frac{u^2}{c_i^3} - \frac{u^4}{3c_i^5}\right) \frac{\mathrm{e}^{-2c_i|\mathbf{x}_1-\mathbf{r}|}}{|\mathbf{x}_1-\mathbf{r}|} \\ &\quad = -Z\alpha^2\,|\phi_n(0)|^2 \int_0^\infty \mathrm{d}u\, C_i \left(\frac{u^2}{c_i^5} - \frac{u^4}{3c_i^7}\right) \\ &\quad = -\frac{4Z\alpha^2}{15}\,|\phi_n(0)|^2\,\frac{C_i}{m_i^2}\,. \end{aligned} \tag{320}$$

The terms of Eq. (318) with $i > 0$ thus tend to 0 in the limit of large masses, while the $i = 0$ part is just the Uehling potential $V_{\mathrm{U}}(\mathbf{x}_1)$ [110, 111].

$$V_{\mathrm{VP}}^{(2,1)}(\mathbf{x}_1) = -\frac{Z\alpha^2}{3\pi}\int \mathrm{d}\mathbf{r}\,\frac{\rho(\mathbf{r})}{|\mathbf{x}_1-\mathbf{r}|}\sum_{i=0}^{2} C_i\,\ln m_i^2 + V_{\mathrm{U}}(\mathbf{x}_1)\,, \tag{321}$$

where

$$\begin{aligned} V_{\mathrm{U}}(\mathbf{x}_1) &= -\frac{Z\alpha^2}{\pi}\int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r}) \int_0^\infty \mathrm{d}u \left(\frac{u^2}{c_0^3} - \frac{u^4}{3c_0^5}\right) \frac{\mathrm{e}^{-2c_0|\mathbf{x}_1-\mathbf{r}|}}{|\mathbf{x}_1-\mathbf{r}|} \\ &= -\frac{Z\alpha^2}{3\pi}\int \mathrm{d}\mathbf{r}\,\rho(\mathbf{r}) \int_1^\infty \mathrm{d}t\,\sqrt{t^2-1}\left(\frac{2}{t^2}+\frac{1}{t^4}\right) \frac{\mathrm{e}^{-2tm_0|\mathbf{x}_1-\mathbf{r}|}}{|\mathbf{x}_1-\mathbf{r}|}\,. \end{aligned} \tag{322}$$

Specializing to the case of a point charge, we recover the well-known expression:

$$V_{\mathrm{U}}(\mathbf{x}_1) = -\frac{Z\alpha^2}{3\pi}\int_1^\infty \mathrm{d}t\,\sqrt{t^2-1}\left(\frac{2}{t^2}+\frac{1}{t^4}\right)\frac{\mathrm{e}^{-2tm_0x_1}}{x_1}\,, \tag{323}$$

where $x_1 = |\mathbf{x}_1|$.

To summarize, the Green function in the expression for the vacuum polarization is not defined for equal coordinates as it appears formally, so a counter term that removes this singularity is subtracted, the Pauli-Villars regularization sum is made for unequal coordinates, and the regulated expression is taken to be the limit as the coordinates become equal. Then, integration over the energy parameter in the Green function is carried out; the charge is renormalized, after which the auxiliary masses are taken to the infinite limit. The result is just the Uehling potential.

Many approximation and numerical methods have been developed to calculate the Uehling potential. One can cite the approximate method by Fullerton and Rinker [77] that provides both the Uehling and and Källén-Sabry vacuum-polarization potentials. Accurate methods to numerically evaluate the Uehling potential have also been provided by Huang [112] and Klarsfeld [113]. Origin expansion is also available from Refs. [114] and [77]. Finite nuclear size is also taken into account in the previous papers, as well as in Ref. [115].

Details on the Evaluation of the One-Potential Green Function

We derive here the equations used in the derivation of Eqs. (299) and (301). Using Eqs. (133) and (298), we can write

$$\begin{aligned}\mathrm{Tr}\left[G^{(1)}(\mathbf{x}_2,\mathbf{x}_1,z)\right] = &-\frac{1}{4\pi^2}\left[\boldsymbol{\nabla}_2\cdot\boldsymbol{\nabla}_1 + m^2 + z^2\right]\\ &\times\int \mathrm{d}\mathbf{w}\,\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\,V(\mathbf{w})\,\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|}\,.\end{aligned} \tag{324}$$

Since $2\,\boldsymbol{\nabla}_2\cdot\boldsymbol{\nabla}_1 = (\boldsymbol{\nabla}_2+\boldsymbol{\nabla}_1)^2 - \boldsymbol{\nabla}_2^2 - \boldsymbol{\nabla}_1^2$, Eq. (299) follows from

$$\begin{aligned}(\boldsymbol{\nabla}_2+\boldsymbol{\nabla}_1)^2\int \mathrm{d}\mathbf{w}\quad &\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\,V(\mathbf{w})\,\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|}\\ &= \int \mathrm{d}\mathbf{w}\,\frac{\mathrm{e}^{-c|\mathbf{x}_2-\mathbf{w}|}}{|\mathbf{x}_2-\mathbf{w}|}\left[\boldsymbol{\nabla}_w^2\,V(\mathbf{w})\right]\frac{\mathrm{e}^{-c|\mathbf{w}-\mathbf{x}_1|}}{|\mathbf{w}-\mathbf{x}_1|}\end{aligned} \tag{325}$$

together with the Poisson equation

$$\boldsymbol{\nabla}_w^2\,V(\mathbf{w}) = 4\pi\,Z\alpha\,\rho(\mathbf{w}) \tag{326}$$

and

$$\left(\nabla_i^2 - c^2\right) \frac{\mathrm{e}^{-c|\mathbf{x}_i - \mathbf{w}|}}{|\mathbf{x}_i - \mathbf{w}|} = -4\pi\delta(\mathbf{x}_i - \mathbf{w})\,. \tag{327}$$

For the counter term in Eq. (300), we have

$$\begin{aligned}\mathrm{Tr}\left[G_{\mathrm{A}}^{(1)}(\mathbf{x}_2, \mathbf{x}_1, z)\right] &= -\frac{1}{4\pi^2}\Bigg\{\left[\nabla_2 \cdot \nabla_1 + m^2 + z^2\right] \\ &\quad \times \int \mathrm{d}\mathbf{w}\, \frac{\mathrm{e}^{-c|\mathbf{x}_2 - \mathbf{w}|}}{|\mathbf{x}_2 - \mathbf{w}|} \frac{\mathrm{e}^{-c|\mathbf{w} - \mathbf{x}_1|}}{|\mathbf{w} - \mathbf{x}_1|}\Bigg\} \frac{V(\mathbf{x}_2) + V(\mathbf{x}_1)}{2}\end{aligned} \tag{328}$$

and Eq. (301) follows from

$$\begin{aligned}\nabla_2 \cdot \nabla_1 \int \mathrm{d}\mathbf{w}\, \frac{\mathrm{e}^{-c|\mathbf{x}_2 - \mathbf{w}|}}{|\mathbf{x}_2 - \mathbf{w}|} \frac{\mathrm{e}^{-c|\mathbf{w} - \mathbf{x}_1|}}{|\mathbf{w} - \mathbf{x}_1|} &= \frac{2\pi}{c}\, \nabla_2 \cdot \nabla_1\, \mathrm{e}^{-c|\mathbf{x}_2 - \mathbf{x}_1|} \\ &= \frac{2\pi}{c}\left(\frac{2c}{|\mathbf{x}_2 - \mathbf{x}_1|} - c^2\right) \mathrm{e}^{-c|\mathbf{x}_2 - \mathbf{x}_1|} \\ &= \frac{4\pi}{|\mathbf{x}_2 - \mathbf{x}_1|}\, \mathrm{e}^{-c|\mathbf{x}_2 - \mathbf{x}_1|} \\ &\quad + \left(z^2 - m^2\right) \int \mathrm{d}\mathbf{w}\, \frac{\mathrm{e}^{-c|\mathbf{x}_2 - \mathbf{w}|}}{|\mathbf{x}_2 - \mathbf{w}|} \\ &\quad \times \frac{\mathrm{e}^{-c|\mathbf{w} - \mathbf{x}_1|}}{|\mathbf{w} - \mathbf{x}_1|}\,.\end{aligned} \tag{329}$$

Higher-Order Contributions

Furry's theorem implies that the two-potential term in the expansion of the all-order vacuum polarization vanishes. It can be shown by direct calculation using the technique above, as is done in Ref. [101] Sec. C. The next order, the three-potential term in Fig. 11b, corresponds to the Wichmann and Kroll potential [109]. The three-potential Green function is given by

$$\begin{aligned}G^{(3)}(\mathbf{x}_2, \mathbf{x}_1, z) = -\int \mathrm{d}\mathbf{s} \int \mathrm{d}\mathbf{v} \int \mathrm{d}\mathbf{w}\, F(\mathbf{x}_2, \mathbf{s}, z)\, V(\mathbf{s}) F(\mathbf{s}, \mathbf{v}, z)\, V(\mathbf{v}) \\ \times F(\mathbf{v}, \mathbf{w}, z)\, V(\mathbf{w}) F(\mathbf{w}, \mathbf{x}_1, z)\,.\end{aligned} \tag{330}$$

which provides the leading term in powers of $Z\alpha$ in the all-order, point-nucleus calculation of Wichmann and Kroll. All-order calculations with finite nuclei have been performed in Ref. [116, 117].

Equation (330) can be approximated by

$$\begin{aligned} G_{\mathrm{A}}^{(3)}(\mathbf{x}_2, \mathbf{x}_1, z) &= -\int \mathrm{d}\mathbf{s} \int \mathrm{d}\mathbf{v} \int \mathrm{d}\mathbf{w}\, F(\mathbf{x}_2, \mathbf{s}, z)\, V(\mathbf{x}_2) F(\mathbf{s}, \mathbf{v}, z)\, V(\mathbf{x}_2) \\ &\quad \times F(\mathbf{v}, \mathbf{w}, z)\, V(\mathbf{x}_2) F(\mathbf{w}, \mathbf{x}_1, z) \\ &= -\frac{[V(\mathbf{x}_2)]^3}{6} \frac{\partial^3}{\partial z^3} F(\mathbf{x}_2, \mathbf{x}_1, z) \ . \end{aligned} \tag{331}$$

We apply it to the mass-regularized functions F_i, as in the one-potential term above. The zero-component trace of the term of mass m_i is given by

$$\begin{aligned} \frac{\partial^3}{\partial z^3} \mathrm{Tr}\left[F_i(\mathbf{x}_2, \mathbf{x}_1, z)\right] &= \frac{\partial^3}{\partial z^3} \frac{z\, \mathrm{e}^{-c_i |\mathbf{x}_2 - \mathbf{x}_1|}}{\pi\, |\mathbf{x}_2 - \mathbf{x}_1|} \\ &= -\frac{\partial^3}{\partial z^3} \frac{z\, c_i}{\pi} + \mathcal{O}\left(|\mathbf{x}_2 - \mathbf{x}_1|\right), \end{aligned} \tag{332}$$

which is finite for $\mathbf{x}_2 \approx \mathbf{x}_1$, and

$$\begin{aligned} \frac{\partial^3}{\partial z^3} \mathrm{Tr}\left[F_i(\mathbf{x}_2, \mathbf{x}_2, z)\right] &= \frac{\partial^3}{\partial u^3} \frac{u}{\pi} \sqrt{m_i^2 + u^2} \\ &= \frac{\partial}{\partial u} \frac{\sqrt{u^2}}{\pi} \left(\frac{2}{u} - \frac{3 m_i^4}{4 u^5} + \ldots\right) . \end{aligned} \tag{333}$$

Integration over u in Eq. 333 is finite and nonzero, as the branch of the square root is positive at $u = \pm\infty$. This is a well-known property of the "light-by-light" scattering Feynman diagram, which yields a spurious finite gauge-noninvariant part [118]. Yet the integral of the regulated approximate expression vanishes with no ambiguity from the quarter circles.

The higher-order terms in the potential expansion of the vacuum polarization beyond those shown on Fig. 11 are all finite and unambiguous.

Self-Energy Screening

The two-electron diagrams in Fig. 8e, f can be evaluated in coordinate space, following the technique described in section "Evaluation of the One-Loop Self-Energy," as shown in Ref. [81]. There is a new contribution here that is shown in Fig. 14. One must treat separately in the diagram Fig. 8e the case in which the energy in the electron propagator between the photon exchanged with the other electron and the self-energy loop is the same as the bound- state energy on the external leg. This gives

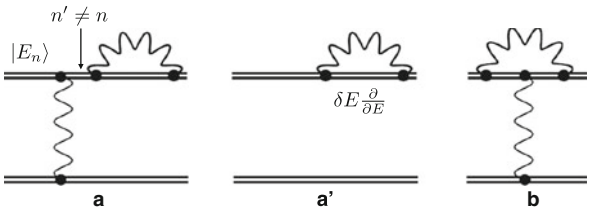


Fig. 14 Feynman diagrams for the self-energy perturbed by an external potential. In the case of the spherically averaged screening potential used in Ref. [81], the diagram labeled A in the left originates in the irreducible part of diagram (e) in Fig. 8 and represents the wave function correction, while the diagram labeled A' in the center comes from the reducible part of diagram (**a**) using the (symbolic) relation $\frac{\partial}{\partial E} G(E) = G(E) \cdot G(E)$, where $G(E)$ is the bound electron propagator. The diagram on the right comes from the vertex correction (f)

rise to an extra contribution, called the reducible contribution, presented in Fig. 14a'. Here we describe a framework for the calculation, in which the formalism is derived in using perturbation theory, replacing the Coulomb potential V by $V + \delta V$ in the equations derived in section "Evaluation of the One-Loop Self-Energy".

In order to establish the expression for the self-energy for a general potential $\mathcal{V}(\mathbf{x})$, we take Eqs. (189) and (190) and make the replacements $V \to \mathcal{V}$ and $E_n \to \mathcal{E}_n$. The potential $\mathcal{V}(\mathbf{x})$ is supposed to be close to a pure Coulomb potential, perturbed by a small potential $\delta V(\mathbf{x})$, which does not have to be spherically symmetric. Some applications of this method have been made with non-spherically symmetric perturbations (see, e.g., Ref. [105]).

We rewrite the high-energy $\mathcal{E}_{\rm H}$ and low-energy $\mathcal{E}_{\rm L}$ parts accordingly, leading to $\mathcal{E}_{\rm SE} = \mathcal{E}_{\rm L} + \mathcal{E}_{\rm H}$, to

$$\mathcal{E}_{\rm L} = \frac{\alpha}{\pi}\mathcal{E}_n - \frac{\alpha}{\pi}{\rm P}\int_0^{\mathcal{E}_n} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1\, \varphi_n^\dagger(\mathbf{x}_2)\alpha^l \mathcal{G}(\mathbf{x}_2, \mathbf{x}_1, z)\alpha^m \varphi_n(\mathbf{x}_1)$$
$$\times (\delta_{lm}\boldsymbol{\nabla}_2 \cdot \boldsymbol{\nabla}_1 - \nabla_2^l \nabla_1^m)\frac{\sin[(\mathcal{E}_n - z)x_{21}]}{(\mathcal{E}_n - z)^2 x_{21}} \tag{334}$$

and to

$$\mathcal{E}_{\rm H} = \frac{\alpha}{2\pi {\rm i}}\int_{C_{\rm H}} dz \int d\mathbf{x}_2 \int d\mathbf{x}_1\, \varphi_n^\dagger(\mathbf{x}_2)\alpha_\mu \mathcal{G}(\mathbf{x}_2, \mathbf{x}_1, z)\alpha^\mu \varphi_n(\mathbf{x}_1)\frac{e^{-bx_{21}}}{x_{21}}$$
$$-\delta m \int d\mathbf{x}\, \varphi_n^\dagger(\mathbf{x})\beta\varphi_n(\mathbf{x})\,. \tag{335}$$

where the indices l and m are summed from 1 to 3 and μ is summed from 0 to 3. The notation is easily derived from the ones in section "Evaluation of the One-Loop Self-Energy," i.e., $b = -i\left[(\mathcal{E}_n - z)^2 + i\delta\right]^{1/2}$, ${\rm Re}(b) > 0$. Here φ_n are the Dirac wave functions in the new potential, and $\mathcal{G}$ is the Green function for the Dirac

equation corresponding to the operator $\mathcal{G} = (\mathcal{H} - z)^{-1}$ The contour C_H is the one defined in Fig. 10.

The screening correction to the self-energy is obtained by applying the transformations

$$\mathcal{V}(\mathbf{x}) \to V(x) + \delta V(\mathbf{x}) \,, \tag{336}$$

$$\varphi_n(\mathbf{x}) \to \phi_n(\mathbf{x}) + \delta\phi_n(\mathbf{x}) \,, \tag{337}$$

$$\mathcal{G}(\mathbf{x}_2, \mathbf{x}_1, z) \to G(\mathbf{x}_2, \mathbf{x}_1, z) + \delta G(\mathbf{x}_2, \mathbf{x}_1, z) \tag{338}$$

$$\mathcal{E}_n \to E_n + \delta E_n \tag{339}$$

in Eqs. (334) and (335) and retaining only the first-order correction terms. As explained above we use in this section the convention that $\mathcal{G}$, $\mathcal{E}_n$, φ_n correspond to the *exact* quantities in the potential $\mathcal{V}(\mathbf{x})$, while G, E_n, ϕ_n correspond to the same objects in a pure Coulomb potential $V(x)$, as defined and used in all preceding sections.

The operators on the unperturbed self-energy that enable to obtain the corrections to the self-energy are defined as δ_ϕ, δ_G, and δ_E, respectively, noted as

$$\delta_\phi = \delta\phi_n \frac{\partial}{\partial\varphi_n} \tag{340}$$

$$\delta_G = \delta G \frac{\partial}{\partial\mathcal{G}} \tag{341}$$

$$\delta_E = \delta E_n \frac{\partial}{\partial\mathcal{E}_n} \,. \tag{342}$$

The total screening correction is then given by

$$\delta\phi_n \frac{\partial}{\partial\varphi_n} + \delta G \frac{\partial}{\partial\mathcal{G}} + \delta E_n \frac{\partial}{\partial\mathcal{E}_n} \tag{343}$$

where the ∂ symbol represents the (formal) differentiation with respect to the indicated variable, while the result is evaluated with the unperturbed functions.

Perturbative Derivation of the Self-Energy Screening Correction

The perturbation expansion is obtained by performing the replacement $\mathcal{V}(\mathbf{x}) \to V(x) + \delta V(x)$, with a spherically symmetric perturbation $\delta V(x)$. The wave functions, the energy, and the Green function appearing in Eqs. (334) and (335) are modified as follows. Perturbation theory gives a first-order energy correction obeying

$$\delta E_n = \int d\mathbf{x}\, \phi_n^\dagger(\mathbf{x})\, \delta V(x) \phi_n(\mathbf{x})$$
$$= \int_0^\infty dx\, x^2 \sum_{i=1}^{2} f_{n,i}(x)\, \delta V(x)\, f_{n,i}(x)\,, \tag{344}$$

where the wave function parts are defined in sections "Spinors" and "Solutions of the Dirac Equation in a Coulomb Potential." The first-order correction to the wave function is obtained by using the reduced Green function $G^R(\mathbf{x}_2, \mathbf{x}_1, E_n)$, defined by (see, e.g., [119]):

$$G^R(\mathbf{x}_2, \mathbf{x}_1, E_n) = \sum_{\substack{m \\ E_m \neq E_n}} \frac{\phi_m(\mathbf{x}_2)\phi_m^\dagger(\mathbf{x}_1)}{E_m - E_n}$$
$$= \lim_{z \to E_n} \left[G(\mathbf{x}_2, \mathbf{x}_1, z) - \frac{\phi_n(\mathbf{x}_2)\phi_n^\dagger(\mathbf{x}_1)}{E_n - z} \right], \tag{345}$$

This expression shows clearly how the separation between reducible and irreducible parts in Fig. 14 is obtained. Using the reduced Green function, the correction to the wave function is obtained as

$$\delta\phi_n(\mathbf{x}_2) = -\int d\mathbf{x}_1 G^R(\mathbf{x}_2, \mathbf{x}_1, E_n)\, \delta V(x_1) \phi_n(\mathbf{x}_1)\,, \tag{346}$$

which transforms to

$$\delta f_{n,i}(x_2) = -\int_0^\infty dx_1 x_1^2 \sum_{j=1}^{2} G_\kappa^{R,ij}(x_2, x_1, E_n)\, \delta V(x_1)\, f_{n,j}(x_1)\,. \tag{347}$$

for a spherically symmetric potential. We define here the components of the radial reduced Green function $G_\kappa^{R,ij}(x_2, x_1, E_n)$ in analogy with the components of the full Green function $G_\kappa^{ij}(x_2, x_1, z)$ following Ref. [29], Eq. (A.14).

To evaluate the first-order correction to the Green function, which would correspond to a diagram expansion similar to the one shown in Fig. 4, we use the operator expansion

$$\mathcal{G}(z) = \frac{1}{H + \delta V - z}$$
$$= \frac{1}{H - z} - \frac{1}{H - z}\, \delta V \frac{1}{H - z}$$
$$+ \frac{1}{H - z}\, \delta V \frac{1}{H - z}\, \delta V \frac{1}{H - z} + \cdots\,. \tag{348}$$

We extract the term of first order in δV as

$$\delta G(z) = -\frac{1}{H-z}\,\delta V\,\frac{1}{H-z}\,. \tag{349}$$

This term has second-order poles in E_n. The first-order correction has thus the expression

$$\delta G(\mathbf{x}_2, \mathbf{x}_1, z) = -\int d\mathbf{x}_3\, G(\mathbf{x}_2, \mathbf{x}_3, z)\delta V(x_3) G(\mathbf{x}_3, \mathbf{x}_1, z) \tag{350}$$

in coordinate space. For a spherically symmetric perturbation δV, we get

$$\delta G_\kappa^{ij}(x_2, x_1, z) = -\int_0^\infty dx_3\, x_3^2 \sum_{k=1}^{2} G_\kappa^{ik}(x_2, x_3, z)\delta V(x_3) G_\kappa^{kj}(x_3, x_1, z). \tag{351}$$

Self-Energy Screening Low-Energy Part

We now turn to the low-energy part defined in Eq. (334). We integrate over the spherical angles of the two vectors $\mathbf{x}_2$ and $\mathbf{x}_1$, for an arbitrary external spherically symmetric potential, which yield

$$\mathcal{E}_\mathrm{L} = \frac{\alpha}{\pi}\mathcal{E}_n + \frac{\alpha}{\pi} P \int_0^{\mathcal{E}_n} dz\, \mathcal{U}(z) \tag{352}$$

where

$$\begin{aligned}\mathcal{U}(z) = &-\int_0^\infty dx_2 x_2^2 \int_0^\infty dx_1 x_1^2 \\ &\times \sum_{\substack{\kappa=+\infty \\ \kappa\neq 0}}^{\kappa=-\infty} \sum_{i,j=1}^{2} \mathcal{F}_{n,\bar{\imath}}(x_2)\mathcal{G}_\kappa^{ij}(x_2, x_1, z)\mathcal{F}_{n,\bar{\jmath}}(x_1)\mathcal{A}_\kappa^{ij}(x_2, x_1, z).\end{aligned} \tag{353}$$

where we have defined $\bar{\imath} = 3 - i$ and $\bar{\jmath} = 3 - j$.

We apply the variations defined in Eqs. (336), (337), (338), (339) to Eqs. (352) and (353), through the radial components of the wave function $\mathcal{F}_{n,i}(x)$, the energy eigenvalue $\mathcal{E}_n$, and the full Green function $\mathcal{G}_\kappa^{ij}(x_2, x_1, z)$, leading to three corrections denoted as $\delta_\phi\,\mathcal{E}_\mathrm{L}$, $\delta_\mathrm{E}\,\mathcal{E}_\mathrm{L}$, and $\delta_\mathrm{G}\,\mathcal{E}_\mathrm{L}$, respectively, summing up to

$$\delta E_\mathrm{L} = \delta_\phi\,\mathcal{E}_\mathrm{L} + \delta_\mathrm{E}\,\mathcal{E}_\mathrm{L} + \delta_\mathrm{G}\,\mathcal{E}_\mathrm{L}\,. \tag{354}$$

Lower-Order Terms

As for the calculation of the regular self-energy described in section "Low-Energy Part," the expression (334) contains spurious parts of lower order in $Z\alpha$ than the complete result. We isolate the physically significant part in a function $\mathcal{F}_{\rm L}(Z\alpha)$ defined as in Eq. (201) for the unperturbed self-energy

$$\mathcal{E}_{\rm L} = \frac{\alpha}{\pi}\left[\frac{5}{6}\mathcal{E}_n + \frac{2}{3}\,\langle\varphi_n|\beta|\varphi_n\rangle + \frac{7}{6}\,\langle\varphi_n|\mathcal{V}|\varphi_n\rangle + \frac{(Z\alpha)^4}{n^3}\mathcal{F}_{\rm L}(Z\alpha)\right] \tag{355}$$

Performing the perturbation expansion in δV up to first order leads to

$$\delta E_{\rm L}(Z\alpha) = \frac{\alpha}{\pi}\left[2\,\delta E_n + \frac{4}{3}\,\langle\phi_n|\beta|\delta\phi_n\rangle + \frac{7}{3}\,\langle\phi_n|V|\delta\phi_n\rangle + \frac{\alpha(Z\alpha)^3}{n^3}\delta F_{\rm L}(Z\alpha)\right] \tag{356}$$

where we have used the indentity

$$\begin{aligned}\delta\,\langle\phi_n|V|\phi_n\rangle &= \langle\phi_n|\delta V|\phi_n\rangle + 2\,\langle\phi_n|V|\delta\phi_n\rangle \\ &= \delta E_n + 2\,\langle\phi_n|V|\delta\phi_n\rangle\,.\end{aligned} \tag{357}$$

Additional transformations must be done to obtain a good numerical accuracy. They are presented in Ref. [81] and will not be dealt with here.

Low-Energy Energy-Level and Wave Function Correction

We turn now to the correction of the low-energy part of the self-energy (334) due to the perturbation to the energy level

$$\delta_{\rm E}\mathcal{E}_{\rm L} = \delta E_n \frac{\partial\mathcal{E}_{\rm L}}{\partial\mathcal{E}_n}\,, \tag{358}$$

where

$$\begin{aligned}\frac{\partial\mathcal{E}_{\rm L}}{\partial\mathcal{E}_n} = \frac{\alpha}{\pi} - \frac{\alpha}{\pi}\Bigg[&\int d\mathbf{x}_2 \int d\mathbf{x}_1\,\phi_n^\dagger(\mathbf{x}_2)\alpha^l G(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^m\phi_n(\mathbf{x}_1) \\ &\times(\delta_{lm}\boldsymbol{\nabla}_2\cdot\boldsymbol{\nabla}_1 - \nabla_2^l\nabla_1^m)\frac{\sin[(E_n-z)x_{21}]}{(E_n-z)^2x_{21}}\Bigg]_{z=E_n} \\ &-\frac{\alpha}{\pi}{\rm P}\int_0^{E_n} dz\int d\mathbf{x}_2\int d\mathbf{x}_1\,\phi_n^\dagger(\mathbf{x}_2)\alpha^l G(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^m\phi_n(\mathbf{x}_1) \\ &\times(\delta_{lm}\boldsymbol{\nabla}_2\cdot\boldsymbol{\nabla}_1 - \nabla_2^l\nabla_1^m)\frac{\partial}{\partial E_n}\frac{\sin[(E_n-z)x_{21}]}{(E_n-z)^2x_{21}}\,.\end{aligned} \tag{359}$$

There is no contribution from the second term on the right-hand side of (359) as seen from

$$(\delta_{lm}\boldsymbol{\nabla}_2\cdot\boldsymbol{\nabla}_1-\nabla_2^l\nabla_1^m)\frac{\sin[(E_n-z)x_{21}]}{(E_n-z)^2x_{21}}=\frac{2}{3}\delta_{lm}(E_n-z)+\mathcal{O}(E_n-z)^3, \quad (360)$$

and

$$\begin{aligned}\int d\mathbf{x}_2\int d\mathbf{x}_1\,\phi_n^\dagger(\mathbf{x}_2)\alpha^l G(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^m\phi_n(\mathbf{x}_1)&\\ =\int d\mathbf{x}_2\int d\mathbf{x}_1\phi_n^\dagger(\mathbf{x}_2)\alpha^l\sum_{E_j=E_n}\frac{\phi_j(\mathbf{x}_2)\phi_j^\dagger(\mathbf{x}_1)}{E_n-z}\alpha^m\phi_n(\mathbf{x}_1)+\mathcal{O}(1)&\\ =\mathcal{O}(1)\,.&\end{aligned} \quad (361)$$

The first term on the right-hand side of Eq. (361) is also zero since

$$\begin{aligned}\int d\mathbf{x}\,\phi_n^\dagger(\mathbf{x})\alpha^l\phi_j(\mathbf{x})&=\mathrm{i}\int d\mathbf{x}\,\phi_n^\dagger(\mathbf{x})[H,x^l]\phi_j(\mathbf{x})\\ &=\mathrm{i}\,(E_n-E_j)\int d\mathbf{x}\,\phi_n^\dagger(\mathbf{x})x^l\phi_j(\mathbf{x})\,,\end{aligned} \quad (362)$$

and the eigenfunctions are orthogonal for $n\neq j$. The final correction due to the perturbation of the bound-state eigenvalue is then

$$\delta_{\mathrm{E}}\mathcal{E}_{\mathrm{L}}=\frac{\alpha}{\pi}\delta E_n+\frac{\alpha}{\pi}P\int_0^{E_n}dz\,\delta_{\mathrm{E}}\mathcal{U}(z)\,, \quad (363)$$

in which

$$\begin{aligned}\delta_{\mathrm{E}}\,\mathcal{U}(z)=-\delta E_n\int_0^\infty dx_2\,x_2^2\int_0^\infty dx_1\,x_1^2\sum_\kappa\sum_{i,j=1}^2&\\ f_i(x_2)G_\kappa^{ij}(x_2,x_1,z)\,f_j(x_1)\frac{\partial}{\partial E_n}A_\kappa^{ij}(x_2,x_1,z)\,.&\end{aligned} \quad (364)$$

The dependence on the bound-state wave function in Eq. (334) is explicit, so the contribution is just

$$\begin{aligned}\delta_\phi\mathcal{E}_{\mathrm{L}}=-\frac{2\alpha}{\pi}\mathrm{P}\int_0^{E_n}dz\int d\mathbf{x}_2\int d\mathbf{x}_1\,\phi_n^\dagger(\mathbf{x}_2)\alpha^l G(\mathbf{x}_2,\mathbf{x}_1,z)\alpha^m\delta\phi_n(\mathbf{x}_1)&\\ \times(\delta_{lm}\boldsymbol{\nabla}_2\cdot\boldsymbol{\nabla}_1-\nabla_2^l\nabla_1^m)\frac{\sin[(E_n-z)x_{21}]}{(E_n-z)^2x_{21}}\,,&\end{aligned} \quad (365)$$

Low-Energy Green Function Correction

The final term to be examined for the low-energy part is due to the variation of the Green function in Eq. (352). It is given by

$$\delta_{\mathrm{G}}\mathcal{E}_{\mathrm{L}} = \frac{\alpha}{\pi}\,\mathrm{Re}\int_{\mathrm{C}_+} dz\,\delta_{\mathrm{G}}\mathcal{U}(z). \tag{366}$$

The first-order change $\delta_{\mathrm{G}}\mathcal{U}(z)$ due to the variation of $\mathcal{G}$ in Eq. (353) is given by

$$\begin{aligned}\delta_{\mathrm{G}}\mathcal{U}(z) = &-\int_0^\infty dx_2\,x_2^2\int_0^\infty dx_1\,x_1^2\\ &\times\sum_{\kappa}\sum_{i,j=1}^{2} f_{n,\bar{\imath}}(x_2)\,\delta G_\kappa^{ij}(x_2,x_1,z)\,f_{n,\bar{\jmath}}(x_1)A_\kappa^{ij}(x_2,x_1,z).\end{aligned} \tag{367}$$

We will define the integration contour C_+ after studying the properties of the integrand. The Green function $\mathcal{G}_\kappa^{ij}(x_2,x_1,z)$ in Eq. (353) has poles along the real axis (see section "Electron Field Propagator" and Fig. 3), which are covered in the integration over z. In the perturbation expansion of $\mathcal{G}_\kappa^{ij}(x_2,x_1,z)$, in the potential δV and the Green function $G_\kappa^{ij}(x_2,x_1,z)$, new, higher-order poles are produced at each order (see section "Perturbative Derivation of the Self-Energy Screening Correction").

We are going to isolate those poles in the relevant orders and evaluate their contribution. We write the Green function in its spectral representation to make explicit the poles:

$$G_\kappa^{ij}(x_2,x_1,z) = \sum_m \frac{f_{m,i}(x_2)\,f_{m,j}(x_1)}{E_m - z}\,. \tag{368}$$

This allows to rewrite Eq. (351) as

$$\begin{aligned}\delta G_\kappa^{ij}(x_2,x_1,z) = &-\int_0^\infty dx_3\,x_3^2\sum_{k=1}^{2}\sum_{m_2,m_1}\\ &\frac{f_{m_2,i}(x_2)\,f_{m_2,k}(x_3)}{E_{m_2}-z}\,\delta V(x_3)\,\frac{f_{m_1,k}(x_3)\,f_{m_1,j}(x_1)}{E_{m_1}-z}\,,\end{aligned} \tag{369}$$

which is explicit for the first- and second-order poles. Only states m_2, m_1 with spin-angular momentum quantum κ can contribute in Eq. (369). We then proceed by extracting the principal parts of $\delta G_\kappa^{ij}(x_2,x_1,z)$. We expand the functions in Eq. (367) in Laurent series about each pole at $z \approx E_m$:

$$G_\kappa^{ij}(x_2,x_1,z) = \frac{f_{m,i}(x_2)\,f_{m,j}(x_1)}{E_m - z} + G_\kappa^{R,ij}(x_2,x_1,E_m) + \mathcal{O}(z-E_m), \tag{370}$$

where $G_\kappa^{R,ij}(x_2, x_1, E_m)$ are the radial components of the reduced Green function given in Eq. (345). We then evaluate the constant term

$$G_\kappa^{R,ij}(x_2, x_1, E_m) = \lim_{z \to E_m} \left[G_\kappa^{ij}(x_2, x_1, z) - \frac{f_{m,i}(x_2) f_{m,j}(x_1)}{E_m - z} \right] = \sum_{\substack{l \\ E_l \neq E_m}} \frac{f_{l,i}(x_2) f_{l,j}(x_1)}{E_l - E_m} . \tag{371}$$

The expression with the first- and second-order poles then reads

$$\begin{aligned} \delta G_\kappa^{ij}(x_2, x_1, z) = & - \int_0^\infty dx_3 \, x_3^2 \\ & \times \sum_{k=1}^{2} \left[f_{m,i}(x_2) f_{m,k}(x_3) \delta V(x_3) f_{m,k}(x_3) f_{m,j}(x_1) \frac{1}{(E_m - z)^2} \right. \\ & + G_\kappa^{R,ik}(x_2, x_3, E_m) \delta V(x_3) f_{m,k}(x_3) f_{m,j}(x_1) \frac{1}{E_m - z} \\ & \left. + f_{m,i}(x_2) f_{m,k}(x_3) \delta V(x_3) G_\kappa^{R,kj}(x_3, x_1, E_m) \frac{1}{E_m - z} \right] \\ & + \mathcal{O}(1) . \end{aligned} \tag{372}$$

Using Eqs. (344)) and (347)), we obtain the final expression

$$\begin{aligned} \delta G_\kappa^{ij}(x_2, x_1, z) = & - f_{m,i}(x_2) f_{m,j}(x_1) \frac{\delta E_m}{(E_m - z)^2} \\ & + \left[\delta f_{m,i}(x_2) \, f_{m,j}(x_1) + f_{m,i}(x_2) \, \delta f_{m,j}(x_1) \right] \frac{1}{E_m - z} \\ & + \mathcal{O}(1) . \end{aligned} \tag{373}$$

Alternatively this result can be obtained by expanding in powers of $\delta V(x)$ the pole contribution to the full Green function $\mathcal{G}_\kappa(x_2, x_1, z)$,

$$\mathcal{G}_\kappa(x_2, x_1, z) = \frac{\mathcal{F}_{m,i}(x_2) \mathcal{F}_{m,j}(x_1)}{\mathcal{E}_m - z} + \mathcal{O}(1) \tag{374}$$

using

$$\begin{aligned} \mathcal{E}_m &= E_m + \delta E_m + \cdots \\ \mathcal{F}_{m,i}(x) &= f_{m,i}(x) + \delta f_{m,i}(x) + \cdots , \end{aligned} \tag{375}$$

and limiting the expansion to the first-order corrections

$$-f_{m,i}(x_2)\,f_{m,j}(x_1)\frac{\delta E_m}{(E_m-z)^2} + \left[\delta f_{m,i}(x_2)\,f_{m,j}(x_1) + f_{m,i}(x_2)\,\delta f_{m,j}(x_1)\right]\frac{1}{E_m-z}\,.$$

In Eq. (335), the level energy also appears in the photon propagator. We thus have to take into account the energy dependence of the $A_\kappa^{ij}(x_2,x_1,z)$ coefficients, issued from the angular expansion of the photon propagator term, as defined in Ref. [29], which appears, e.g., in Eq. (280). The expansion reads

$$\begin{aligned} A_\kappa^{ij}(x_2,x_1,z) = {} & A_\kappa^{ij}(x_2,x_1,E_m) \\ & +(z-E_m)\,B_\kappa^{ij}(x_2,x_1,E_m) \\ & +\mathcal{O}\left((z-E_m)^2\right)\,, \end{aligned} \tag{376}$$

where

$$B_\kappa^{ij}(x_2,x_1,E_m) = \frac{d}{dz}\left[A_\kappa^{ij}(x_2,x_1,z)\right]_{z=E_m}\,. \tag{377}$$

The complete expansion for $z \approx E_m$ is thus given by

$$\begin{aligned} &\int_0^\infty dx_3\,x_3^2\sum_{k=1}^{2} G_\kappa^{ik}(x_2,x_3,z)\delta V(x_3)G_\kappa^{kj}(x_3,x_1,z)\,A_\kappa^{ij}(x_2,x_1,z) \\ &\quad = f_{m,i}(x_2)\,f_{m,j}(x_1)\,\delta E_m \\ &\qquad \times\left[A_\kappa^{ij}(x_2,x_1,E_m)\frac{1}{(E_m-z)^2} - B_\kappa^{ij}(x_2,x_1,E_m)\frac{1}{E_m-z}\right] \\ &\qquad -\left[\delta f_{m,i}(x_2)\,f_{m,j}(x_1) + f_{m,i}(x_2)\,\delta f_{m,j}(x_1)\right] \\ &\qquad \times A_\kappa^{ij}(x_2,x_1,E_m)\frac{1}{E_m-z} + \mathcal{O}(1)\,. \end{aligned} \tag{378}$$

We now calculate the contribution of the low-energy part by subtracting the pole terms from the integrand and evaluating the integral over z numerically. The pole contribution is then added back after analytic evaluation. In the one-electron self-energy calculation, the singularities along the real axis in the interval $(0,1)$ give poles, which leads to the principal value integral in Eq. (352). Here we have double poles as well, so a new analysis is necessary to calculate this new terms. In Ref. [29] it is shown that the integration over z can be written as

$$\mathrm{Re}\int_{C_+} dz\,\mathcal{U}(z), \tag{379}$$

where C_+, extending from $z = 0$ to $z = E_n$ above the real axis in the complex z plane, is shown in Fig. 10. We use here an analytic method as described in Ref. [28]. We note $\mathcal{U}_\mathrm{p}$ the pole contribution in $\mathcal{U}$. We thus write

$$\begin{aligned}\mathrm{Re}\int_{C_+} dz\,\mathcal{U}(z) &= \mathrm{Re}\int_{C_+} dz\,\left[\mathcal{U}(z) - \mathcal{U}_\mathrm{p}(z) + \mathcal{U}_\mathrm{p}(z)\right], \\ &= \int_0^{E_n} dz\left[\mathcal{U}(z) - \mathcal{U}_p(z) + \frac{1}{E_n} Re\int_{C_+} dz'\,\mathcal{U}_\mathrm{p}(z')\right]. \quad (380)\end{aligned}$$

The pole at the endpoint of the z integration, $z = E_n$, can be calculated easily as shown in section "Low-Energy Energy-Level and Wave Function Correction" [Eqs. (359), (360), (361), (362)].

The analytic evaluation of the one- and two-pole terms is thus

$$\mathrm{Re}\int_{C_+} dz\,\frac{1}{(E_m - z)^2} = -\frac{E_n}{E_m(E_n - E_m)}\,, \quad (381)$$

$$\mathrm{Re}\int_{C_+} dz\,\frac{1}{E_m - z} = \ln\left[\frac{E_m}{E_n - E_m}\right], \quad (382)$$

where

$$0 < E_m < E_n\,. \quad (383)$$

Using these two results, we can evaluate Eq. (366) to get

$$\begin{aligned}\delta_\mathrm{G}\mathcal{E}_\mathrm{L} &= \frac{\alpha}{\pi}\,\mathrm{Re}\int_{C_+} dz\,\delta_\mathrm{G}\mathcal{U}(z) \\ &= \frac{\alpha}{\pi}\int_0^{E_n}\left\{\delta_\mathrm{G}\mathcal{U}(z) - \sum_{\substack{m \\ E_m<E_n}}\left[\frac{R^{(2)}_{n,m}}{(E_m - z)^2} + \frac{R^{(1)}_{n,m}}{(E_m - z)}\right]\right\} \\ &\quad + \frac{\alpha}{\pi}\sum_{\substack{m \\ E_m<E_n}}\left[-\frac{R^{(2)}_{n,m}\,E_n}{E_m(E_n - E_m)} + R^{(1)}_{n,m}\ln\left(\frac{E_m}{E_n - E_m}\right)\right] \quad (384)\end{aligned}$$

where $R^{(2)}_{n,m}$ is given by

$$\begin{aligned}R^{(2)}_{n,m} &= \delta E_m\int_0^\infty dx_2\,x_2^2\int_0^\infty dx_1\,x_1^2 \\ &\quad\times\sum_{i,j=1}^{2} f_{n,\bar{i}}(x_2)\,f_{m,i}(x_2)\,f_{m,j}(x_1)\,f_{n,\bar{j}}(x_1)\,A^{ij}_\kappa(x_2, x_1, E_m) \quad (385)\end{aligned}$$

and $R^{(1)}_{n,m}$ by

$$
\begin{aligned}
R^{(1)}_{n,m} = &-\delta E_m \int_0^\infty dx_2\, x_2^2 \int_0^\infty dx_1\, x_1^2 \\
&\times \sum_{i,j=1}^{2} f_{n,\bar{\imath}}(x_2)\, f_{m,i}(x_2)\, f_{m,j}(x_1)\, f_{n,\bar{\jmath}}(x_1) B^{ij}_\kappa(x_2, x_1, E_m) \\
&- \int_0^\infty dx_2\, x_2^2 \int_0^\infty dx_1\, x_1^2 \\
&\times \Bigg\{ \sum_{i,j=1}^{2} f_{n,\bar{\imath}}(x_2) \Big[\delta f_{m,i}(x_2)\, f_{m,j}(x_1) + f_{m,i}(x_2)\, \delta f_{m,j}(x_1) \Big] f_{n,\bar{\jmath}}(x_1) \\
&\qquad \times A^{ij}_\kappa(x_2, x_1, E_m) \Bigg\}
\end{aligned}
\tag{386}
$$

Self-Energy Screening High-Energy Part

We now turn to the high-energy part in Eq. (335). This term is infinite an must be regularized. We follow the same technique as for the normal self-energy (section "Introduction" and Refs. [27,100]) to remove the infinities. Again, the high-energy part $\mathcal{E}_{\mathrm{H}}$, as given in Eq. (335), is separated into two contributions

$$
\delta E_{\mathrm{H}} = \delta E_{\mathrm{HA}} + \delta E_{\mathrm{HB}} \,. \tag{387}
$$

where $\mathcal{E}_{\mathrm{HA}}$ contains all divergences and can be calculated analytically. The other quantity $\mathcal{E}_{\mathrm{HB}}$ can be calculated numerically. In this section we describe the method used to evaluate δE_{HB}. The method to compute analytically the divergent part δE_{HA} is explained in section "Self-Energy Screening: Analytic Terms and Verification of Global Renormalization".

The high-energy remainder with term-by-term subtraction, from Eq. (280) (see also Eq. (32) in Ref. [100]), is written as

$$
\begin{aligned}
\mathcal{E}_{\mathrm{HB}} &= \mathcal{E}_{\mathrm{H}} - \mathcal{E}_{\mathrm{HA}} \\
&= \frac{\alpha}{2\pi \mathrm{i}} \int_{\mathrm{C_H}} dz \int_0^\infty dx_2\, x_2^2 \int_0^\infty dx_1\, x_1^2 \\
&\quad \times \Bigg\{ \sum_{|\kappa|=1}^{\infty} \Big[\mathcal{K}_\kappa(x_2, x_1, z) - \mathcal{K}^{(0,0)}_\kappa(x_2, x_1, z) - \mathcal{K}^{(0,1)}_\kappa(x_2, x_1, z) \\
&\qquad - \mathcal{K}^{(1,0)}_\kappa(x_2, x_1, z) \Big] - \mathcal{K}^{(0,2)}_{\mathrm{D}}(x_2, x_1, z) \Bigg\} \,,
\end{aligned}
\tag{388}
$$

where the different $\mathcal{K}$ functions are defined as

$$\mathcal{K}_\kappa(x_2,x_1,z) = \sum_{i,j=1}^{2} \left[\mathcal{F}_{n,i}(x_2)\mathcal{G}_\kappa^{ij}(x_2,x_1,z)\mathcal{F}_{n,j}(x_1)A_\kappa(x_2,x_1)\right.$$
$$\left.-\mathcal{F}_{n,\bar{i}}(x_2)\mathcal{G}_\kappa^{ij}(x_2,x_1,z)\mathcal{F}_{n,\bar{j}}(x_1)A_\kappa^{ij}(x_2,x_1)\right], \tag{389}$$

$$\mathcal{K}_\kappa^{(0,0)}(x_2,x_1,z) = A_\kappa\left\{F_\kappa^{11}(x_2,x_1,z)\left[\mathcal{F}_{n,1}^2(x_2)-3\mathcal{F}_{n,2}^2(x_2)\right]\right.$$
$$\left.+F_{-\kappa}^{22}(x_2,x_1,z)\left[\mathcal{F}_{n,2}^2(x_2)-3\mathcal{F}_{n,1}^2(x_2)\right]\right\}, \tag{390}$$

$$\mathcal{K}_\kappa^{(0,1)}(x_2,x_1,z)$$
$$= \left\{F_\kappa^{11}(x_2,x_1,z)\left[\mathcal{F}_{n,1}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1+\kappa_n}{x_2}\right)\mathcal{F}_{n,1}(x_2)\right.\right.$$
$$\left.-\mathcal{F}_{n,2}(x_2)\left(3B_\kappa\frac{\partial}{\partial x_2}-C_\kappa\frac{1-\kappa_n}{x_2}\right)\mathcal{F}_{n,2}(x_2)\right]$$
$$+F_{-\kappa}^{22}(x_2,x_1,z)\left[\mathcal{F}_{n,2}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1-\kappa_n}{x_2}\right)\mathcal{F}_{n,2}(x_2)\right.$$
$$\left.-\mathcal{F}_{n,1}(x_2)\left(3B_\kappa\frac{\partial}{\partial x_2}-C_\kappa\frac{1+\kappa_n}{x_2}\right)\mathcal{F}_{n,1}(x_2)\right]$$
$$+2\,F_{-\kappa}^{12}(x_2,x_1,z)\mathcal{F}_{n,1}(x_2)$$
$$\times\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1-\kappa_n}{x_2}\right)\mathcal{F}_{n,2}(x_2)$$
$$+2\,F_\kappa^{21}(x_2,x_1,z)\mathcal{F}_{n,2}(x_2)$$
$$\left.\times\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1+\kappa_n}{x_2}\right)\mathcal{F}_{n,1}(x_2)\right\}, \tag{391}$$

and

$$\mathcal{K}_\kappa^{(1,0)}(x_2,x_1,z) = -\mathcal{V}(x_2)A_\kappa\left\{\frac{d}{dz}F_\kappa^{11}(x_2,x_1,z)\left[\mathcal{F}_{n,1}^2(x_2)-3\mathcal{F}_{n,2}^2(x_2)\right]\right.$$
$$\left.+\frac{d}{dz}F_{-\kappa}^{22}(x_2,x_1,z)\left[\mathcal{F}_{n,2}^2(x_2)-3\mathcal{F}_{n,1}^2(x_2)\right]\right\}. \tag{392}$$

In Eqs. (389), (390), (391), and (392) A_κ^{ij}, A_κ, B_κ, and C_κ are integrals over angular coordinates (see Refs. [28–30, 100]), $\mathcal{F}_{n,i}(x)$ are radial components of the bound-state wave function defined in the previous sections, and κ_n is the spin-angular momentum quantum number of the bound-state n. The expressions for the free Green function radial components $F_\kappa^{ij}(x_2,x_1,z)$ and their derivatives can be found in [100] and those of the Coulomb Green function can be found in [29].

The high-energy remainder for the self-energy screening is obtained from Eq. (388) as described in section "Perturbative Derivation of the Self-Energy Screening Correction" as the sum $\delta_\phi \mathcal{E}_{\mathrm{HB}} + \delta_{\mathrm{G}} \mathcal{E}_{\mathrm{HB}} + \delta_{\mathrm{E}} \mathcal{E}_{\mathrm{HB}}$. We now derive the expressions for $\delta_\phi \mathcal{E}_{\mathrm{HB}}$, $\delta_{\mathrm{G}} \mathcal{E}_{\mathrm{HB}}$, and $\delta_{\mathrm{E}} \mathcal{E}_{\mathrm{HB}}$ from Eqs. (389), (390), (391), and (392) in sections "Wave Function Correction for the Screened Self-Energy," "Green Function Correction," and "Energy Correction."

Wave Function Correction for the Screened Self-Energy

The expression for the high-energy remainder for the wave function correction is obtained using the functional derivatives of Eqs. (389), (390), (391), and (392) with respect to the radial wave functions $\mathcal{F}_{n,i}(x)$. For the full expression (389), we have

$$\begin{aligned} K_{\phi,\kappa}(x_2, x_1, z) = \sum_{i,j=1}^{2} \Big\{ \Big[& f_{n,i}(x_2) G_\kappa^{ij}(x_2, x_1, z)\, \delta f_{n,j}(x_1) \\ & + \delta f_{n,i}(x_2) G_\kappa^{ij}(x_2, x_1, z)\, f_{n,j}(x_1) \Big] A_\kappa(x_2, x_1) \\ & - \Big[f_{n,\bar{i}}(x_2) G_\kappa^{ij}(x_2, x_1, z)\, \delta f_{n,\bar{j}}(x_1) \\ & + \delta f_{n,\bar{i}}(x_2) G_\kappa^{ij}(x_2, x_1, z)\, f_{n,\bar{j}}(x_1) \Big] A_\kappa^{ij}(x_2, x_1) \Big\} . \end{aligned} \tag{393}$$

The subtraction terms are evaluated as

$$\begin{aligned} & K_{\phi,\kappa}^{(0,0)}(x_2, x_1, z) \\ & \quad = 2A_\kappa \Big\{ F_\kappa^{11}(x_2, x_1, z) \Big[f_{n,1}(x_2)\, \delta f_{n,1}(x_2) - 3 f_{n,2}(x_2)\, \delta f_{n,2}(x_2) \Big] \\ & \qquad + F_{-\kappa}^{22}(x_2, x_1, z) \Big[f_{n,2}(x_2)\, \delta f_{n,2}(x_2) - 3 f_{n,1}(x_2)\, \delta f_{n,1}(x_2) \Big] \Big\}, \end{aligned} \tag{394}$$

$$\begin{aligned} & K_{\phi,\kappa}^{(0,1)}(x_2, x_1, z) \\ & \quad = \Big\{ F_\kappa^{11}(x_2, x_1, z) \Big[f_{n,1}(x_2) \left(B_\kappa \frac{\partial}{\partial x_2} + C_\kappa \frac{1+\kappa_n}{x_2} \right) \delta f_{n,1}(x_2) \\ & \qquad - f_{n,2}(x_2) \left(3B_\kappa \frac{\partial}{\partial x_2} - C_\kappa \frac{1-\kappa_n}{x_2} \right) \delta f_{n,2}(x_2) \\ & \qquad + \delta f_{n,1}(x_2) \left(B_\kappa \frac{\partial}{\partial x_2} + C_\kappa \frac{1+\kappa_n}{x_2} \right) f_{n,1}(x_2) \\ & \qquad - \delta f_{n,2}(x_2) \left(3B_\kappa \frac{\partial}{\partial x_2} - C_\kappa \frac{1-\kappa_n}{x_2} \right) f_{n,2}(x_2) \Big] \end{aligned}$$

$$
\begin{aligned}
&+F^{22}_{-\kappa}(x_2,x_1,z)\Bigg[f_{n,2}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1-\kappa_n}{x_2}\right)\delta f_{n,2}(x_2)\\
&-f_{n,1}(x_2)\left(3B_\kappa\frac{\partial}{\partial x_2}-C_\kappa\frac{1+\kappa_n}{x_2}\right)\delta f_{n,1}(x_2)\\
&+\delta f_{n,2}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1-\kappa_n}{x_2}\right)f_{n,2}(x_2)\\
&-\delta f_{n,1}(x_2)\left(3B_\kappa\frac{\partial}{\partial x_2}-C_\kappa\frac{1+\kappa_n}{x_2}\right)f_{n,1}(x_2)\Bigg]\\
&+2\,F^{12}_{-\kappa}(x_2,x_1,z)\Bigg[f_{n,1}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1-\kappa_n}{x_2}\right)\delta f_{n,2}(x_2)\\
&+\delta f_{n,1}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1-\kappa_n}{x_2}\right)f_{n,2}(x_2)\Bigg]\\
&+2\,F^{21}_{\kappa}(x_2,x_1,z)\Bigg[f_{n,2}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1+\kappa_n}{x_2}\right)\delta f_{n,1}(x_2)\\
&+\delta f_{n,2}(x_2)\left(B_\kappa\frac{\partial}{\partial x_2}+C_\kappa\frac{1+\kappa_n}{x_2}\right)f_{n,1}(x_2)\Bigg]\Bigg\},
\end{aligned}
\tag{395}
$$

and

$$
\begin{aligned}
&K^{(1,0)}_{\phi,\kappa}(x_2,x_1,z)=-2V(x_2)A_\kappa\\
&\times\left\{\frac{d}{dz}F^{11}_{\kappa}(x_2,x_1,z)\left[f_{n,1}(x_2)\,\delta f_{n,1}(x_2)-3f_{n,2}(x_2)\,\delta f_{n,2}(x_2)\right]\right.\\
&\left.+\frac{d}{dz}F^{22}_{-\kappa}(x_2,x_1,z)\left[f_{n,2}(x_2)\,\delta f_{n,2}(x_2)-3f_{n,1}(x_2)\,\delta f_{n,1}(x_2)\right]\right\}.
\end{aligned}
\tag{396}
$$

Using Eqs. (393), (394), (395), and (396), the first-order wave function correction to $\mathcal{E}_{\mathrm{HB}}$ is given by

$$
\begin{aligned}
\delta_\phi\mathcal{E}_{\mathrm{HB}}=\frac{\alpha}{2\pi\mathrm{i}}\int_{C_{\mathrm{H}}}dz\int_0^\infty dx_2\,x_2^2\int_0^\infty dx_1\,x_1^2\\
\times\sum_{|\kappa|=1}^{\infty}\Bigg[K_{\phi,\kappa}(x_2,x_1,z)-K^{(0,0)}_{\phi,\kappa}(x_2,x_1,z)\\
-K^{(0,1)}_{\phi,\kappa}(x_2,x_1,z)-K^{(1,0)}_{\phi,\kappa}(x_2,x_1,z)\Bigg].
\end{aligned}
\tag{397}
$$

In order to evaluate the expression in Eq. (395), we need the derivative of the bound-state Dirac wave function and the derivative of its first-order correction in the potential δV. The first part is found in section "Bound States of the Dirac Equation,"

Eq. (55). Using this equation, we obtain the expressions for the perturbation of the wave-function components in the potential $V(x)+\delta V(x)$. Retaining only first-order terms in $\delta V(x)$,it reads

$$\begin{aligned}\frac{d}{dx}\delta f_{n,1}(x) &= -\frac{1+\kappa_n}{x}\delta f_{n,1}(x) + [1+E_n-V(x)]\,\delta f_{n,2}(x)\\ &\quad +[\delta E_n - \delta V(x)]\,f_{n,2}(x)\\ \frac{d}{dx}\delta f_{n,2}(x) &= [1-E_n+V(x)]\,\delta f_{n,1}(x) - \frac{1-\kappa_n}{x}\delta f_{n,2}(x)\\ &\quad +[\delta V(x) - \delta E_n]\,f_{n,1}(x)\,. \end{aligned} \tag{398}$$

Green Function Correction

The effect of the variation of the Coulomb Green function under a change of the potential on the high-energy term is straightforward. The Coulomb Green function has no poles on the high-energy integration contour so we use directly Eq. (351) and get

$$\begin{aligned}K_{\mathrm{G},\kappa}(x_2,x_1,z) &= -\sum_{i,j,k=1}^{2}\int_0^\infty dx_3\, x_3^2\,\delta V(x_3)\\ &\times\Big[f_{n,i}(x_2)G_\kappa^{ik}(x_2,x_3,z)G_\kappa^{kj}(x_3,x_1,z)\,f_{n,j}(x_1)A_\kappa(x_2,x_1)\\ &\quad -f_{n,\bar{i}}(x_2)G_\kappa^{ik}(x_2,x_3,z)G_\kappa^{kj}(x_3,x_1,z)\,f_{n,\bar{j}}(x_1)A_\kappa^{ij}(x_2,x_1)\Big]\,. \end{aligned} \tag{399}$$

Here only the $K^{(1,0)}$ term contributes to the subtraction term, and Eq. (392) leads to

$$\begin{aligned}&K_{\mathrm{G},\kappa}^{(1,0)}(x_2,x_1,z)\\ &\quad = -\delta V(x_2)A_\kappa\Big\{\frac{d}{dz}F_\kappa^{11}(x_2,x_1,z)\left[f_{n,1}^2(x_2)-3f_{n,2}^2(x_2)\right]\\ &\qquad +\frac{d}{dz}F_{-\kappa}^{22}(x_2,x_1,z)\left[f_{n,2}^2(x_2)-3f_{n,1}^2(x_2)\right]\Big\}\,. \end{aligned} \tag{400}$$

The Green function correction to $\mathcal{E}_{\mathrm{HB}}$ is thus

$$\begin{aligned}\delta_{\mathrm{G}}\mathcal{E}_{\mathrm{HB}} &= \frac{\alpha}{2\pi\mathrm{i}}\int_{\mathrm{C_H}}dz\int_0^\infty dx_2\,x_2^2\int_0^\infty dx_1\,x_1^2\\ &\quad\times\sum_{|\kappa|=1}^{\infty}\Big[K_{\mathrm{G},\kappa}(x_2,x_1,z)-K_{\mathrm{G},\kappa}^{(1,0)}(x_2,x_1,z)\Big] \end{aligned} \tag{401}$$

Energy Correction

We finally turn to the evaluation of the perturbation effect on $\delta_{\rm E}\mathcal{E}_{\rm HB}$, given by

$$\delta_{\rm E}\mathcal{E}_{\rm HB} = \delta E_n \frac{\partial}{\partial \mathcal{E}_n}\mathcal{E}_{\rm HB} \,, \tag{402}$$

Although this calculation can be performed by direct numerical differentiation of $\mathcal{E}_{\rm HB}$,it has been found preferable in Ref. [81] to differentiate directly the running term in the sum over angular momenta in Eq. (388). The sum of derivatives obtained by term-by-term differentiation is given by

$$\begin{aligned}\frac{\partial}{\partial E_n} S_A(r,y,u) = \sum_{|\kappa|=1}^{\infty} \frac{\partial}{\partial E_n}\Big[& K_{\phi,\kappa}(ry,y,iu) \\ & -K^{(0,0)}_{\phi,\kappa}(ry,y,iu) - K^{(0,1)}_{\phi,\kappa}(ry,y,iu) \\ & -K^{(1,0)}_{\phi,\kappa}(ry,y,iu)\Big]\end{aligned} \tag{403}$$

where $u = (1/2)(1/t - t)$ and $K^{(i,j)}_{\phi,\kappa}$ are defined in Eqs. (393), (394), (395), and (396).

Self-Energy Screening: Analytic Terms and Verification of Global Renormalization

We finally turn to the analytic evaluation of contributions of the renormalization terms from the three diagrams presented in Fig. 14 to $\delta E_{\rm HA}$. These terms were subtracted in the numerical calculation described in section "Self-Energy Screening High-Energy Part" to obtain finite contribution. The calculations reported in the previous sections and in Ref. [27, 81, 100] have been done to keep explicit the dependence in V and E_n to be able to obtain these analytic contributions by direct differentiation.

Wave Function Correction

The analytic portion of the wave-function correction follows from the previous section (see Refs. [27, 100]). In each of the terms $E^{(i,j)}_{\rm H}$, we calculate the variation with respect to a change in the wave function based on the explicit dependence on the wave function in V and E_n and obtain

$$\delta_\phi \mathcal{E}^{(0,0)}_{\rm H} = \frac{2\alpha}{\pi}\langle\phi_n|\beta|\delta\phi_n\rangle\left[\ln(\Lambda^2) - 1 + \frac{1-E_n^2}{E_n^2}\ln\left(1+E_n^2\right) + \mathcal{O}(\Lambda^{-1})\right] \tag{404}$$

$$\delta_\phi \mathcal{E}_{\mathrm{H}}^{(0,1)} = \frac{2\alpha}{\pi} \langle \phi_n | \boldsymbol{\alpha} \cdot \mathbf{p} \, | \delta\phi_n \rangle \left[\frac{1}{4} \ln(\Lambda^2) - \frac{6 - 3E_n^2 + 7E_n^4}{24E_n^2(1 + E_n^2)} + \frac{1 - E_n^4}{4E_n^4} \ln\left(1 + E_n^2\right) + \mathcal{O}(\Lambda^{-1}) \right], \tag{405}$$

and

$$\delta_\phi \mathcal{E}_{\mathrm{H}}^{(1,0)} = \frac{2\alpha}{\pi} \left\{ \langle \phi_n | V | \delta\phi_n \rangle \left[\frac{1}{4} \ln(\Lambda^2) + \frac{6 - E_n^2}{8E_n^2} - \frac{3 + E_n^4}{4E_n^4} \ln\left(1 + E_n^2\right) \right] - \langle \phi_n | \beta V | \delta\phi_n \rangle \left[\frac{2}{E_n} - \frac{2}{E_n^3} \ln\left(1 + E_n^2\right) \right] + \mathcal{O}(\Lambda^{-1}) \right\}. \tag{406}$$

In analogy with what is described in section "Evaluation of the One-Loop Self-Energy," we define

$$\delta_\phi \mathcal{E}_{\mathrm{HA}} = \lim_{\Lambda \to \infty} \left[\delta_\phi \mathcal{E}_{\mathrm{H}}^{(0,0)} + \delta_\phi \mathcal{E}_{\mathrm{H}}^{(0,1)} + \delta_\phi \mathcal{E}_{\mathrm{H}}^{(1,0)} - 2\,\delta m(\Lambda) \, \langle \phi_n | \beta | \delta\phi_n \rangle \right], \tag{407}$$

where the last term is the renormalization term with

$$\delta m(\Lambda) = \frac{\alpha}{\pi} \left[\frac{3}{4} \ln(\Lambda^2) + \frac{3}{8} \right], \tag{408}$$

which reabsorb the infinite contributions into the physical mass of the particle.

We now regroup the contributions to $\ln(\Lambda^2)$ in Eq. (407), leading to

$$\frac{\alpha}{2\pi} \langle \phi_n | \beta + \boldsymbol{\alpha} \cdot \mathbf{p} + V | \delta\phi_n \rangle = \frac{\alpha}{2\pi} E_n \, \langle \phi_n | \delta\phi_n \rangle = 0 \tag{409}$$

using the differential equations from Eqs. (395) and (398) for ϕ_n and the fact that $\langle \phi_n | \delta\phi_n \rangle = 0$. Equation (409) show that the wave function correction is finite. We can then deduce more generally that the contribution of diagram of Fig. 14b is independently finite.

Energy Correction

Following the same route as in the previous subsection, we differentiate all terms with respect to $\mathcal{E}_n$. We obtain three terms

$$\delta_{\mathrm{E}} \mathcal{E}_{\mathrm{H}}^{(0,0)} = \frac{\alpha}{\pi} \delta E_n \left[\langle \phi_n | \beta | \phi_n \rangle \left(\frac{2}{E_n} \frac{1 - E_n^2}{1 + E_n^2} - \frac{2}{E_n^3} \ln\left(1 + E_n^2\right) \right) - \frac{1}{4} \ln(\Lambda^2) - \frac{6 - E_n^2}{8E_n^2} + \frac{3 + E_n^4}{4E_n^4} \ln\left(1 + E_n^2\right) + \mathcal{O}(\Lambda^{-1}) \right], \tag{410}$$

$$\delta_{\rm E}\mathcal{E}_{\rm H}^{(0,1)} = \frac{\alpha}{\pi}\delta E_n \,\langle\phi_n|\boldsymbol{\alpha}\cdot\mathbf{p}|\phi_n\rangle \left[\frac{6+9E_n^2-8E_n^4-3E_n^6}{6E_n^3\left(1+E_n^2\right)^2} - \frac{1}{E_n^5}\log(1+E_n^2) + \mathcal{O}(\Lambda^{-1})\right], \tag{411}$$

and

$$\delta_{\rm E}\mathcal{E}_{\rm H}^{(1,0)} = \frac{\alpha}{\pi}\delta E_n \Bigg\{ \langle\phi_n|V|\phi_n\rangle \left[-\frac{6+3E_n^2+E_n^4}{2E_n^3(1+E_n^2)} + \frac{3}{E_n^5}\log(1+E_n^2)\right] + \langle\phi_n|\beta V|\phi_n\rangle \left[\frac{6+2E_n^2}{E_n^2(1+E_n^2)} - \frac{6}{E_n^4}\log(1+E_n^2)\right] + \mathcal{O}(\Lambda^{-1})\Bigg\} . \tag{412}$$

Summing up all three contributions, we obtain

$$\delta_{\rm E}\mathcal{E}_{\rm HA} = \delta_{\rm E}\mathcal{E}_{\rm H}^{(0,0)} + \delta_{\rm E}\mathcal{E}_{\rm H}^{(0,1)} + \delta_{\rm E}\mathcal{E}_{\rm H}^{(1,0)} , \tag{413}$$

which contains the divergent term

$$\delta_{\rm E}\mathcal{E}_{\rm HA} = -\frac{\alpha\,\delta E_n}{4\pi}\ln(\Lambda^2) + \mathcal{O}(1) . \tag{414}$$

It shows that the contribution of the diagram of Fig. 14a' is not finite.

Green Function Correction

Here the whole contribution results from $\mathcal{E}_{\rm H}^{(1,0)}$, which is linear in the potential in the Green function. Taking into account the definition of the perturbed energy correction $\langle\phi_n|\delta V|\phi_n\rangle = \delta E_n$, we obtain

$$\delta_{\rm G}\mathcal{E}_{\rm HA} = \delta_{\rm G}\mathcal{E}_{\rm H}^{(1,0)} = \frac{\alpha}{\pi}\Bigg\{\delta E_n\left[\frac{1}{4}\ln(\Lambda^2) + \frac{6-E_n^2}{8E_n^2} - \frac{3+E_n^4}{4E_n^4}\ln\left(1+E_n^2\right)\right] - \langle\phi_n|\beta\delta V|\phi_n\rangle\left[\frac{2}{E_n} - \frac{2}{E_n^3}\ln\left(1+E_n^2\right)\right] + \mathcal{O}(\Lambda^{-1})\Bigg\}\Bigg] . \tag{415}$$

We find again a divergent contribution

$$\delta_{\rm G}\mathcal{E}_{\rm HA} = \frac{\alpha\,\delta E_n}{4\pi}\ln(\Lambda^2) + \mathcal{O}(1) . \tag{416}$$

Comparing with Eq. (414), we see that the infinities cancel in the sum of the two terms. This shows that the sum of the reducible and irreducible contributions

represented by the two diagrams Fig. 14a and a' must be taken into account simultaneously to obtain a finite contribution. This completes the derivation of the total corrections

$$\delta E_{\mathrm{HA}} = \delta_{\phi}\mathcal{E}_{\mathrm{HA}} + \delta_{\mathrm{E}}\mathcal{E}_{\mathrm{HA}} + \delta_{\mathrm{G}}\mathcal{E}_{\mathrm{HA}}, \tag{417}$$

which is finite as expected.

Two-Photon, Two-Electron Diagrams

Here we give some indications on the evaluation of the so-called ladder and cross-ladder diagrams (Fig. 8a, b). These diagrams are finite, since there are no closed loops. Yet there are specific new issues that have to be dealt with, which were not present in any of the diagram treated in the preceding sections. In Fig. 15, we present the electron-electron interaction and non-radiative QED corrections. Non-radiative means that they are no photon or electron loops. Many-body calculations (see, e.g., Ref. [120]) in atoms were performed very early in the framework of standard perturbation theory, or the one of the (multiconfiguration) Hartree-Fock or many-body approximation [121], and more recently in the framework of the multiconfiguration Dirac-Fock (MCDF) [122–125] or of the relativistic many-body perturbation theory, the relativistic configuration interaction (RCI), or coupled cluster techniques [126]. All those calculations were meant to calculate the *electron correlation* energy. In a diagram language, these calculations represent a sum of the ladder-approximation diagrams (Fig. 15, bottom part). It was shown that in the case of the relativistic many-body techniques [127], the electron-electron interaction, because it couples positive- and negative-energy states, must include projection operators on the negative-energy states; otherwise the contribution is divergent starting at second order in perturbation [46–48, 50].

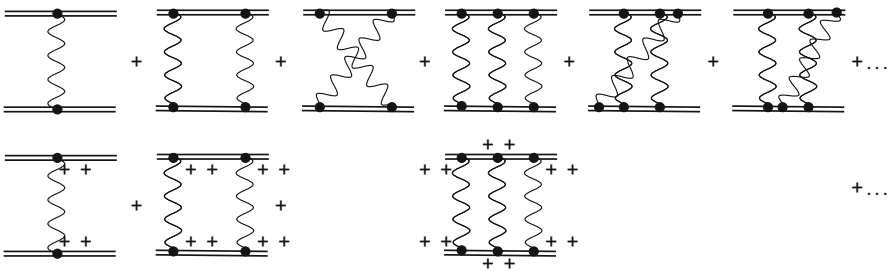


Fig. 15 Non-radiative two-electron Feynman diagrams (*top*) and the ladder approximation (*bottom*). The ++ symbols mean that the electron-electron interaction is between projectors on the positive- energy eigenvalues. This series of diagrams corresponds to the correlation energy as evaluated by many-body techniques like RMBPT, MCDF, RCI, etc.

Here we discuss the QED evaluation of the first- and second-order diagrams, following Ref. [51, 83]. The QED expression for the upper leftmost diagram in Fig. 15 between two states with energy E_a and E_b is given by

$$\Delta E^{(2)} = \alpha \int d\mathbf{x}_2 \int d\mathbf{x}_1 \left\{ \phi_a^\dagger(\mathbf{x}_2)\, \phi_b^\dagger(\mathbf{x}_1)\, \frac{1}{x_{21}} \alpha_\mu^{(2)} \alpha^{\mu\,(1)} \phi_a(\mathbf{x}_2)\, \phi_b(\mathbf{x}_1) \right.$$
$$\left. - \phi_a^\dagger(\mathbf{x}_2)\, \phi_b^\dagger(\mathbf{x}_1)\, \frac{\cos\left[|E_b - E_a|\, x_{21}\right]}{x_{21}} \alpha_\mu^{(2)} \alpha^{\mu\,(1)} \phi_a(\mathbf{x}_1)\, \phi_b(\mathbf{x}_2) \right\}, \quad (418)$$

using unperturbed Dirac wave functions. The cosine term corresponds to the retardation part of the interaction (taking into account that the photon speed is finite) and the $\alpha_\mu^{(2)} \alpha^{\mu\,(1)}$ correspond to the magnetic part of the interaction. One recognizes the usual perturbation theory result.

The second-order ladder (L) and cross-ladder (X) diagrams come from the same fourth-order term in Eq. (113) as the screened self-energy,

$$\Delta E^{(4)} = \lim_{\epsilon \to 0} \frac{i\epsilon}{2} \left(4 \left\langle S_\epsilon^{(4)} \right\rangle_C - 2 \left\langle S_\epsilon^{(2)} \right\rangle_C^2 - 4 \left\langle S_\epsilon^{(1)} \right\rangle_C \left\langle S_\epsilon^{(3)} \right\rangle_C \right.$$
$$\left. + 4 \left\langle S_\epsilon^{(1)} \right\rangle_C^2 \left\langle S_\epsilon^{(2)} \right\rangle_C - \left\langle S_\epsilon^{(1)} \right\rangle_C^4 \right). \quad (419)$$

Only the two first terms contribute. Using the expression in Eq. (115), the energy shift for the ladder diagram writes

$$\left\langle S_\epsilon^{(4)} \right\rangle_L = \frac{e^2}{2} \int d^4\mathbf{x}_4 \int d^4\mathbf{x}_3 \int d^4\mathbf{x}_2 \int d^4\mathbf{x}_1 e^{(|t_4|+|t_3|+|t_2|+|t_1|)}$$
$$\times \left\{ \sum_{n_4,n_3,n_2,n_1} \left[\left\langle a_{n_4}^\dagger a_{n_2}^\dagger a_{n_1} a_{n_3} \right\rangle \right.\right.$$
$$\times \bar{\phi}_{n_4}(\mathbf{x}_4)\, \gamma_\mu S_F(\mathbf{x}_4, \mathbf{x}_3)\, \gamma_\nu \phi_{n_3}(\mathbf{x}_3)$$
$$\times \bar{\phi}_{n_2}(\mathbf{x}_2)\, \gamma^\mu S_F(\mathbf{x}_2, \mathbf{x}_1)\, \gamma^\nu \phi_{n_1}(\mathbf{x}_1)$$
$$\left.\left. \times D_F(\mathbf{x}_4 - \mathbf{x}_2)\, D_F(\mathbf{x}_3 - \mathbf{x}_1) \right] \right\}. \quad (420)$$

Similarly the energy shift for the cross-ladder diagram writes

$$\left\langle S_\epsilon^{(4)} \right\rangle_X = \frac{e^2}{2} \int d^4\mathbf{x}_4 \int d^4\mathbf{x}_3 \int d^4\mathbf{x}_2 \int d^4\mathbf{x}_1 e^{(|t_4|+|t_3|+|t_2|+|t_1|)}$$
$$\times \left\{ \sum_{n_4,n_3,n_2,n_1} \left[\left\langle a_{n_4}^\dagger a_{n_2}^\dagger a_{n_1} a_{n_3} \right\rangle \right.\right.$$

$$\times \bar{\phi}_{n_4}(\mathbf{x}_4)\,\gamma_\mu S_F(\mathbf{x}_4,\mathbf{x}_3)\,\gamma_\nu \phi_{n_3}(\mathbf{x}_3)$$
$$\times \bar{\phi}_{n_2}(\mathbf{x}_2)\,\gamma^\mu S_F(\mathbf{x}_2,\mathbf{x}_1)\,\gamma^\nu \phi_{n_1}(\mathbf{x}_1)$$
$$\times D_F(\mathbf{x}_4-\mathbf{x}_1)\,D_F(\mathbf{x}_3-\mathbf{x}_2)\Big]\Big\}. \tag{421}$$

Performing the time integration in Eqs. (420) and (421) replacing the propagators by the relevant expressions as defined in sections "Electron Field Propagator" and "Photon Field Operators," one gets

$$\begin{aligned}\left\langle S_\epsilon^{(4)}\right\rangle_L = \frac{e^2}{2}\int d^4\mathbf{x}_4\int d^4\mathbf{x}_3\int d^4\mathbf{x}_2\int d^4\mathbf{x}_1\int_{-\infty}^{+\infty} dq_2\int_{-\infty}^{+\infty} dq_1\\ \times\int_{-\infty}^{+\infty} dz_2\int_{-\infty}^{+\infty} dz_1\Bigg\{\sum_{n_4,n_3,n_2,n_1}\Big[\left\langle a^\dagger_{n_4}a^\dagger_{n_2}a_{n_1}a_{n_3}\right\rangle\\ \times\frac{\epsilon}{\epsilon^2+(E_{n_4}-z_2-q_2)^2}\frac{\epsilon}{\epsilon^2+(E_{n_3}-z_2+q_1)^2}\\ \times\frac{\epsilon}{\epsilon^2+(E_{n_2}-z_1+q_2)^2}\frac{\epsilon}{\epsilon^2+(E_{n_1}-z_1-q_1)^2}\\ \times\phi^\dagger_{n_4}(\mathbf{x}_4)\,\alpha_\mu G(\mathbf{x}_4,\mathbf{x}_3,z_2(1+i\delta))\,\alpha_\nu\phi_{n_3}(\mathbf{x}_3)\,H(\mathbf{x}_4-\mathbf{x}_2,q_2,)\\ \times\phi^\dagger_{n_2}(\mathbf{x}_2)\,\alpha^\mu G(\mathbf{x}_2,\mathbf{x}_1,z_1(1+i\delta))\,\alpha^\nu\phi_{n_1}(\mathbf{x}_1)\,H(\mathbf{x}_3-\mathbf{x}_1,q_1)\Big]\Bigg\},\end{aligned} \tag{422}$$

and

$$\begin{aligned}\left\langle S_\epsilon^{(4)}\right\rangle_X = \frac{e^2}{2}\int d^4\mathbf{x}_4\int d^4\mathbf{x}_3\int d^4\mathbf{x}_2\int d^4\mathbf{x}_1\int_{-\infty}^{+\infty} dq_2\int_{-\infty}^{+\infty} dq_1\\ \times\int_{-\infty}^{+\infty} dz_2\int_{-\infty}^{+\infty} dz_1\Bigg\{\sum_{n_4,n_3,n_2,n_1}\Big[\left\langle a^\dagger_{n_4}a^\dagger_{n_2}a_{n_1}a_{n_3}\right\rangle\\ \times\frac{\epsilon}{\epsilon^2+(E_{n_4}-z_2-q_2)^2}\frac{\epsilon}{\epsilon^2+(E_{n_3}-z_2+q_1)^2}\\ \times\frac{\epsilon}{\epsilon^2+(E_{n_2}-z_1+q_1)^2}\frac{\epsilon}{\epsilon^2+(E_{n_1}-z_1-q_2)^2}\\ \times\phi^\dagger_{n_4}(\mathbf{x}_4)\,\alpha_\mu G(\mathbf{x}_4,\mathbf{x}_3,z_2(1+i\delta))\,\alpha_\nu\phi_{n_3}(\mathbf{x}_3)\,H(\mathbf{x}_4-\mathbf{x}_1,q_2,)\\ \times\phi^\dagger_{n_2}(\mathbf{x}_2)\,\alpha^\mu G(\mathbf{x}_2,\mathbf{x}_1,z_1(1+i\delta))\,\alpha^\nu\phi_{n_1}(\mathbf{x}_1)\,H(\mathbf{x}_3-\mathbf{x}_2,q_1)\Big]\Bigg\}.\end{aligned} \tag{423}$$

We define

$$g_{a,b,c,d}(q) = -4\pi\alpha \int d^4\mathbf{x}_2 \int d^4\mathbf{x}_1 H\left(\mathbf{x}_2 - \mathbf{x}_1, q\right) \phi_a^\dagger\left(\mathbf{x}_2\right) \alpha_\mu \phi_c\left(\mathbf{x}_2\right) \times \phi_b^\dagger\left(\mathbf{x}_1\right) \alpha^\mu \phi_d\left(\mathbf{x}_1\right), \tag{424}$$

and replace $G\left(\mathbf{x}_a, \mathbf{x}_b, z(1+i\delta)\right)$ by its spectral decomposition (131) in Eqs. (422) and (423). As an example, we treat the ground state $\left|1s^2\,{}^1S_0\right\rangle = a_b^\dagger a_a^\dagger |0\rangle$. In this case, $E_{n_4} = E_{n_3} = E_{n_2} = E_{n_1} = E_0$. We can then use

$$\sum_{n_4,n_3,n_2,n_1} \left\langle a_{n_4}^\dagger a_{n_2}^\dagger a_{n_1} a_{n_3} \right\rangle f\left(n_4, n_3, n_2, n_1\right) = 2\left[f\left(a,a,b,b\right) - f\left(a,b,b,a\right)\right] \tag{425}$$

to rewrite Eq. (422) as

$$\begin{aligned}\left\langle S_\epsilon^{(4)}\right\rangle_L = \frac{1}{\pi^4} \int_{-\infty}^{+\infty} dq_2 \int_{-\infty}^{+\infty} dq_1 \Bigg\{ \sum_{n,m} \Big[& g_{a,b,n,m}(q_2) g_{n,m,a,b}(q_1) \\ & - g_{a,b,n,m}(q_2) g_{n,m,b,a}(q_1) \Big] \Bigg\} \\ \times \int_{-\infty}^{+\infty} dz_2 \int_{-\infty}^{+\infty} dz_1 & \frac{1}{E_n - z_2(1+i\delta)} \frac{1}{E_m - z_1(1+i\delta)} \\ \times & \frac{\epsilon}{\epsilon^2 + (E_0 - z_2 - q_2)^2} \frac{\epsilon}{\epsilon^2 + (E_0 - z_2 + q_1)^2} \\ \times & \frac{\epsilon}{\epsilon^2 + (E_0 - z_1 + q_2)^2} \frac{\epsilon}{\epsilon^2 + (E_0 - z_1 - q_1)^2},\end{aligned} \tag{426}$$

and Eq. (423) as

$$\begin{aligned}\left\langle S_\epsilon^{(4)}\right\rangle_X = \frac{1}{\pi^4} \int_{-\infty}^{+\infty} dq_2 \int_{-\infty}^{+\infty} dq_1 \Bigg\{ \sum_{n,m} \Big[& g_{a,n,m,b}(q_2) g_{b,n,m,a}(q_1) \\ & - g_{a,m,n,a}(q_2) g_{b,n,m,b}(q_1) \Big] \Bigg\} \\ \times \int_{-\infty}^{+\infty} dz_2 \int_{-\infty}^{+\infty} dz_1 & \frac{1}{E_n - z_2(1+i\delta)} \frac{1}{E_m - z_1(1+i\delta)} \\ \times & \frac{\epsilon}{\epsilon^2 + (E_0 - z_2 - q_2)^2} \frac{\epsilon}{\epsilon^2 + (E_0 - z_2 + q_1)^2} \\ \times & \frac{\epsilon}{\epsilon^2 + (E_0 - z_1 + q_1)^2} \frac{\epsilon}{\epsilon^2 + (E_0 - z_1 - q_2)^2}.\end{aligned} \tag{427}$$

Finally, we have for the second-order term, which will be needed for having the right order in $1/\epsilon$:

$$\begin{aligned}\left\langle S_\epsilon^{(2)}\right\rangle_E &= -\frac{e^2}{2}\int d^4\mathbf{x}_2 \int d^4\mathbf{x}_1 e^{(|t_2|+|t_1|)}\Bigg\{\sum_{n_4,n_3,n_2,n_1}\Bigg[\left\langle a_{n_4}^\dagger a_{n_2}^\dagger a_{n_1} a_{n_3}\right\rangle \\ &\quad\times \bar\phi_{n_4}(\mathbf{x}_2)\,\gamma_\mu\phi_{n_3}(\mathbf{x}_2)\,\bar\phi_{n_2}(\mathbf{x}_1)\,\gamma^\mu\phi_{n_1}(\mathbf{x}_1)\,D_F(\mathbf{x}_2-\mathbf{x}_1)\Bigg]\Bigg\} \\ &= -\frac{2i}{\pi}\int_{-\infty}^{+\infty} dq\,[g_{a,b,a,b}(q)-g_{a,b,b,a}(q)]\,\frac{\epsilon^2}{(q^2+\epsilon^2)^2}. \end{aligned} \tag{428}$$

Here we will only describe how to deal with the limit $\epsilon \to 0$, which requires distinguishing as in the case of the self-energy screening in section "Self-Energy Screening," between reducible and irreducible states. The reducible states correspond to $E_n = E_m = E_0$. In this case the term $\left\langle S_\epsilon^{(4)}\right\rangle_L$ has a singularity of order $1/\epsilon^2$ that cancels out only when combining with $\left\langle S_\epsilon^{(2)}\right\rangle_E^2$.

Equation (426), when specializing to the ground state of He-like ions, can be transformed into

$$\begin{aligned}\left\langle S_\epsilon^{(4)}\right\rangle_{L\mathrm{red}} &= \frac{1}{\pi^4}\int_{-\infty}^{+\infty} dq_2\int_{-\infty}^{+\infty} dq_1\Bigg\{\sum_{n,m\in 1s_{\pm 1/2}}\Bigg[g_{a,n,m,b}(q_2)g_{b,n,m,a}(q_1) \\ &\quad -g_{a,m,n,a}(q_2)g_{b,n,m,b}(q_1)\Bigg]\Bigg\} f_\epsilon(q_2,q_1)\,f_\epsilon(q_1,q_2)\,, \end{aligned} \tag{429}$$

where

$$\begin{aligned} f_\epsilon(q_2,q_1) &= \int_{-\infty}^{+\infty} dz \frac{1}{E_0 - z(1+i\delta)}\frac{\epsilon}{\epsilon^2+(E_0-z-q_2)^2}\frac{\epsilon}{\epsilon^2+(E_0-z+q_1)^2} \\ &= \frac{-\pi\epsilon}{(q_2+q_1)^2+4\epsilon^2}\frac{q_2-q_1-4i\epsilon}{(q_2-i\epsilon)(q_1+i\epsilon)}. \end{aligned} \tag{430}$$

In the same way one obtains

$$\begin{aligned}\left\langle S_\epsilon^{(4)}\right\rangle_{X\mathrm{red}} &= \frac{1}{\pi^4}\int_{-\infty}^{+\infty} dq_2\int_{-\infty}^{+\infty} dq_1\Bigg\{\sum_{n,m\in 1s_{\pm 1/2}}\Bigg[g_{a,n,m,b}(q_2)g_{b,n,m,a}(q_1) \\ &\quad -g_{a,m,n,a}(q_2)g_{b,n,m,b}(q_1)\Bigg]\Bigg\} f_\epsilon(q_2,q_1)^2\,. \end{aligned} \tag{431}$$

The singularity can be isolated by noticing that

$$\frac{1}{\pi^4} f_\epsilon(q_2, q_1)\, f_\epsilon(q_1, q_2) = \frac{[f_\epsilon(q_2, q_1) + f_\epsilon(q_1, q_2)]^2 - f_\epsilon^2(q_2, q_1) - f_\epsilon^2(q_1, q_2)}{2\pi^4}. \quad (432)$$

The first term writes

$$\frac{[f_\epsilon(q_2, q_1) + f_\epsilon(q_1, q_2)]^2}{2\pi^4} = \frac{-2\epsilon^4}{\pi^2 \left(q_2^2 + \epsilon^2\right)^2 \left(q_1^2 + \epsilon^2\right)^2} \quad (433)$$

We can then evaluate

$$\left\langle S_\epsilon^{(2)} \right\rangle_E^2 = -\frac{4}{\pi^2} \int_{-\infty}^{+\infty} dq_2 \int_{-\infty}^{+\infty} dq_1 \left[g_{a,b,a,b}(q_2) - g_{a,b,b,a}(q_2)\right] \times \left[g_{a,b,a,b}(q_1) - g_{a,b,b,a}(q_1)\right] \frac{\epsilon^2}{\left(q_2^2 + \epsilon^2\right)^2} \frac{\epsilon^2}{\left(q_1^2 + \epsilon^2\right)^2}. \quad (434)$$

and use the identity

$$\left[g_{a,b,a,b}(q_2) - g_{a,b,b,a}(q_2)\right] \left[g_{a,b,a,b}(q_1) - g_{a,b,b,a}(q_1)\right] = \sum_{n,m \in 1s_{\pm 1/2}} \left[g_{a,n,m,b}(q_2) g_{b,n,m,a}(q_1) - g_{a,m,n,a}(q_2) g_{b,n,m,b}(q_1)\right] \quad (435)$$

to show that $-2\left\langle S_\epsilon^{(2)} \right\rangle_E^2$ as presented in (434) exactly cancels the $[f_\epsilon(q_2, q_1) + f_\epsilon(q_1, q_2)]^2$ part of the $\left\langle S_\epsilon^{(4)} \right\rangle_L$ term in $4\left\langle S_\epsilon^{(4)} \right\rangle_{L\mathrm{red}}$.

This formalism has been applied to the ground state of He-like ions [83] and to the $1s2s\,{}^1S_0$ and ${}^3S_1$ [87]. It has also been applied to the $1s^2 2p_{1/2} \to 1s^2 2s_{1/2}$ transition [86] and to the $1s^2 2p_{3/2}$ level in Li-like ions [88].

Conclusion

In his chapter, we have reviewed the formalism of QED, starting from the Dirac equation, introducing field-theoretical approches and perturbation theory. We have then developed the formalism of bound-state quantum electrodynamics in space-coordinate approach, giving practical methods to perform explicit calculations. These were illustrated by several examples, from one-electron self-energy and vacuum polarization, two-electron self-energy screening, and second-order electron-electron interaction. In the following chapters, the two-time Green function

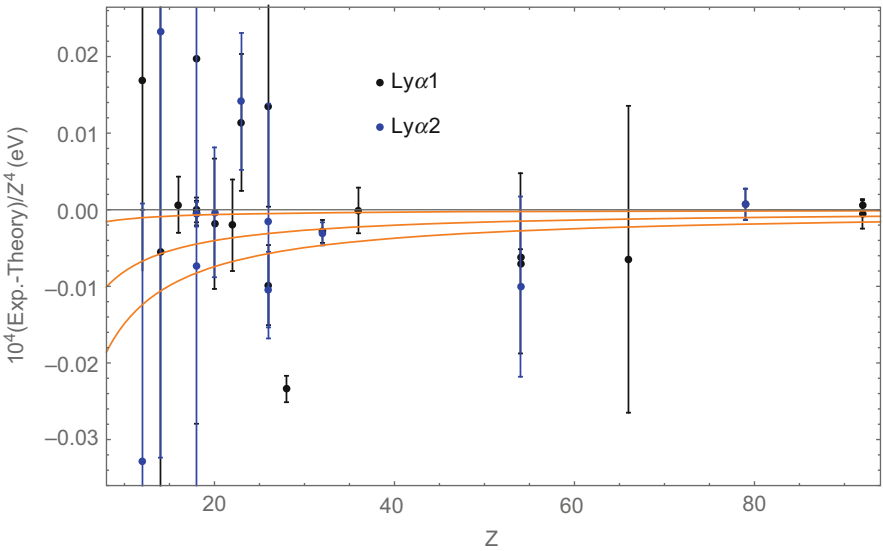


Fig. 16 Comparison between theory and experiment for H-like ions for $16 \leq Z \leq 92$. The red lines correspond to a fit using a aZ^3 functions, with the upper and lower 1σ error bands. Theory: Ref. [4]. Experiment: $Z = 12$: Ref. [138], $Z = 14$: Ref. [139], $Z = 16$: Ref. [140], $Z = 18$: Refs. [140–142], $Z = 20$: Ref. [143], $Z = 22$: Ref. [144], $Z = 23$: Ref. [145], $Z = 26$: Refs. [146, 147], $Z = 28$: Ref. [148], $Z = 32$: Ref. [149], $Z = 36$: Ref. [150], $Z = 54$: Refs. [151, 152], $Z = 66$: Ref. [153], $Z = 79$: Refs. [154, 155], $Z = 92$: Refs. [156, 157]

approach is presented. Then the covariant-evolution-operator method, which aims at unifying BSQED and many-body perturbation theory techniques, is described.

Many experimental tests of BSQED in few-electron ions have been performed, using either accelerators, ion traps, or storage rings.

A compilation of theoretical values for hydrogen-like ions has been published recently [4]. Agreement with hydrogen and highly charged ions is excellent as can be seen from Fig. 16. Yet there is a strong difficulty connected to the 7σ discrepancy between the proton size obtained from muonic and normal hydrogen [128, 129], which is unexplained up to now. This disagreement can be a hint of new physics as well as an incomplete understanding of the proton structure.

Measurements of $1s2\ell \rightarrow 1s^2\,{}^1S_0$ transitions in two-electron ions also show an excellent agreement with BSQED calculations as reported in [5]. The comparison between theory and experiment for all lines that have been measured ($1s2p\,{}^1P_1 \rightarrow 1s^2\,{}^1S_0$, $1s2p\,{}^3P_1 \rightarrow 1s^2\,{}^1S_0$, $1s2p\,{}^3P_2 \rightarrow 1s^2\,{}^1S_0$, and $1s2s\,{}^3S_1 \rightarrow 1s^2\,{}^1S_0$) is shown in Fig. 17. High-Z measurements of the $1s2p\,{}^3P_2 \rightarrow 1s2s\,{}^3S_1$ transition [130, 131] are also in excellent agreement with theory.

It can be seen from the error bands of a fit with a aZ^3 function plotted in Fig. 17 that there is no trend on the theory and experiment difference, although the experimental values tend to be slightly above the Z-axis. Fits with constant, linear, or quadratic functions lead to the same conclusion, which contradicts the

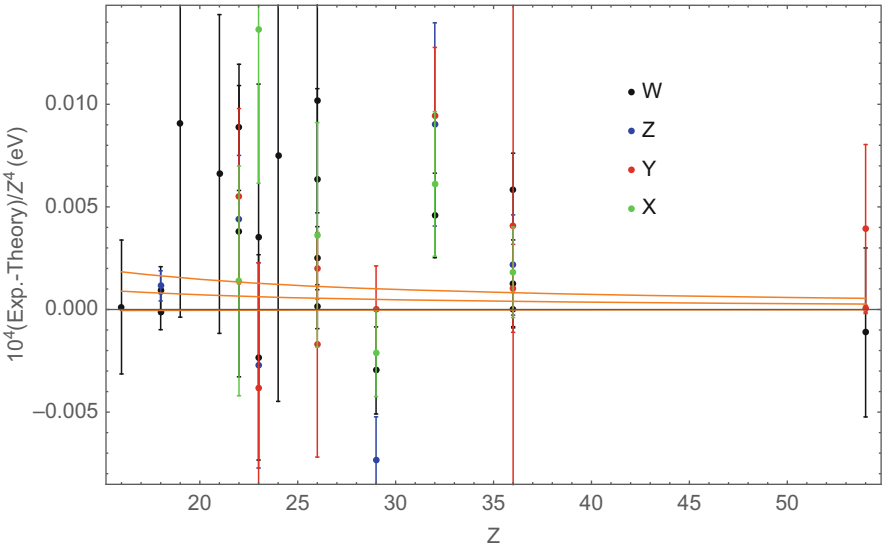


Fig. 17 Comparison between theory and experiment for He-like ions for $16 \leq Z \leq 54$. The orange lines correspond to a fit using a aZ^3 functions, with the upper and lower 1σ error bands. Theory: Ref. [5]. Experiment: $Z = 16$: Ref. [140], Ref. $Z = 18$: Refs. [158, 159], $Z = 19$, $Z = 21$, $Z = 23$: Ref. [160], $Z = 22$: Ref. [132, 160, 161], $Z = 26$: Refs. [160, 162, 163], $Z = 29$: Ref. [164], $Z = 32$: Ref. [165], $Z = 36$: Refs. [166–168], $Z = 54$: Ref. [169]

conclusions of Ref. [132]. It is clear that including all the recent, much more accurate $Z = 18$, $Z = 22$, $Z = 29$, and $Z = 36$ values makes the discrepancy disappear.

Many experiments have also been made to measure the fine structure splitting $1s^2 2p_j \rightarrow 1s^2 2s_{1/2}$ in lithium-like ions. Theoretical calculations have been improved over the years, with better and better account of BSQED corrections. The most recent, in the framework of the S-matrix formalism described here, is presented in Ref. [89]. MCDF calculations are available [14, 133], as well as RMBPT calculations [134], RCI calculations [135], RCI calculations with relaxed QED [136], and BSQED calculations by St Petersburg group [71, 86, 137]. The accuracy for the most recent calculations is in the range of 0.001eV at low-Z up to $\pm$0.15eV at high-Z for the $1s^2 2p_{1/2} \rightarrow 1s^2 2s_{1/2}$ transition and up to $\pm$0.4eV $1s^2 2p_{3/2} \rightarrow 1s^2 2s_{1/2}$ transition. We show in Fig. 18 the comparison between the calculations in Ref. [89] and available experiments with $10 \leq Z \leq 92$. Again the agreement is excellent and no deviation can be observed. All fits with functions of the form aZ^n, $0 \leq n \leq 3$ have 1σ error bands that are slightly above the Z-axis.

Finally, extensive comparisons have been performed for the K, L, and M transitions in atoms with an inner-shell hole [182–185], in the framework of RMBPT, with QED corrections including self-energy screening in the Welton

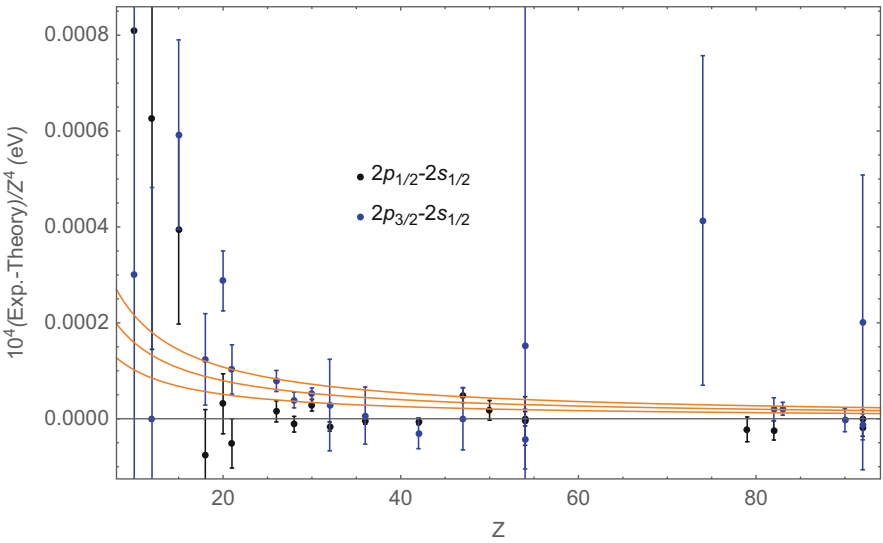


Fig. 18 Comparison between theory and experiment for the $1s^2 2p_j \rightarrow 1s^2 2s_{1/2}$ in Li-like ions with $10 \leq Z \leq 92$. The orange lines correspond to a fit using a Z^3 function, with the upper and lower 1σ error bands. Theory: Ref. [89]. Experiment from NIST database [170] with the following exceptions. For the $1s^2 2p_{1/2} \rightarrow 1s^2 2s_{1/2}$ transition – $Z = 47$: Ref. [171], $Z = 50$: Ref. [172], $Z = 54$: Ref. [173, 174], $Z = 74$: Ref. [175], $Z = 82$: Ref. [176], $Z = 90$, Ref. [177], $Z = 92$: Ref. [178, 179]. For the $1s^2 2p_{3/2} \rightarrow 1s^2 2s_{1/2}$ transition–for $Z = 47$: Ref. [171], $Z = 50$: Ref. [172], $Z = 54$: Ref. [172, 174], $Z = 79$, $Z = 82$, Ref. [180], $Z = 92$: Ref. [180, 181]

approximation. The agreement is excellent for $10 \leq Z \leq 100$, taking into account the complexity of calculations with hole states, which requires taking care simultaneously of correlation and auto-ionization.

It is then clear that all available experiments to date, using few-electron or neutral atoms, for all Z are in excellent agreement with BSQED as presented here, comforting the huge theoretical and experimental effort that has been performed in the last 40 years. The mystery of the so-called proton size puzzle remains, and there is not hint that this large discrepancy has repercussions in the high-Z region of electronic atoms, even though the $1s$ electron distance with the nucleus, for $Z \geq 42$ is smaller than the distance between the muon and the proton in the $2s$ state of muonic hydrogen.

Acknowledgements Laboratoire Kastler-Brossel http://www.lkb.upmc.fr is *Unité Mixte de Recherche* UMR #8552 of Sorbonne Université, UPMC Paris-6; PSL-ENS; Collège de France and CNRS. The author is a member of the ExtreMe Matter Institute (EMMI) https://www.gsi.de/work/wissenschaftliche_netzwerke/helmholtz_allianz_emmi.htm, an Helmholtz Association Alliance (HA216/EMMI). P.I. wishes to thank P.B. for her continuous support and encouragements during the redaction of this chapter.

Appendix: Useful Properties

Properties of Dirac γ Matrices

Basic Relations

To recover the relativistic invariant

$$p^\mu p_\mu = E^2/c^2 - \mathbf{p}^2 = m^2c^2, \tag{436}$$

the γ^μ matrices must obey

$$(\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu) = 2g^{\mu\nu} \tag{437}$$

which can be rewritten as an anticommutator (noted $\{,\}$)

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}, \tag{438}$$

where $g^{\mu\nu} = (1, -1, -1, -1)$ is the metric tensor for special relativity. We first evaluate the determinant of the γ^μ. Taking the determinant of $\gamma^\mu\gamma^\nu$, we must have using (437) for $\mu \neq \nu$ $Det\,[\gamma^\mu\gamma^\nu] = Det\,[-\gamma^\mu\gamma^\nu] = (-1)^N Det\,[\gamma^\mu\gamma^\nu]$, where N is the dimension of the matrix. But since $Det\,[\gamma^\mu\gamma^\nu] = Det\,[\gamma^\mu\gamma^\nu]$, one must have $(-1)^N = 1$, *i.e.,* N even. The smallest value for which this can be realized with 4 linearly independent matrices is $N = 4$. One can express the γ in terms of the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} et\ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{439}$$

There are many representations used for the Dirac matrices. In atomic physics and BSQED, it is more convenient to use them in the form:

$$\gamma^0 = \beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}, \boldsymbol{\gamma} = \beta\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix}, \boldsymbol{\alpha} = \beta\boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}. \tag{440}$$

Among other interesting properties, we can deduce from (437) $\left(\gamma^0\right)^2 = \beta^2 = 1$ and $\left(\gamma^i\right)^2 = -1$.

More Properties of the Dirac Matrices

One can rewrite the tensor $\gamma^\mu\gamma^\nu$ as a sum of a symmetric and antisymmetric tensor as

$$\gamma^\mu\gamma^\nu = g^{\mu\nu} + \sigma^{\mu\nu}. \tag{441}$$

The Dirac γ matrix commutators and anticommutators can be rewritten as

$$[\gamma^\mu, \gamma^\nu] = \gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu = 2\sigma^{\mu\nu}. \tag{442}$$

$$\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu} \tag{443}$$

where with our conventions the antisymmetric tensors $\sigma^{\mu\nu}$ is given by

$$\sigma^{\mu\nu} = (\boldsymbol{\alpha}, i\,\boldsymbol{\Sigma})\,, \tag{444}$$

from which one can deduce the following contractions, using only Eq. (437):

$$\begin{aligned}
\gamma^\mu\gamma_\mu &= 4 \\
\gamma^\mu\gamma^\nu\gamma_\mu &= -2\gamma^\nu \\
\gamma^\mu\gamma^\lambda\gamma^\nu\gamma_\mu &= 4g^{\lambda\nu} \\
\gamma^\mu\gamma^\lambda\gamma^\nu\gamma^\rho\gamma_\mu &= -2\gamma^\rho\gamma^\nu\gamma^\lambda \\
\gamma^\mu\gamma^\lambda\gamma^\nu\gamma^\rho\gamma^\sigma\gamma_\mu &= 2\left(\gamma^\sigma\gamma^\lambda\gamma^\nu\gamma^\rho + \gamma^\rho\gamma^\nu\gamma^\lambda\gamma^\sigma\right).
\end{aligned} \tag{445}$$

The traces are found to be

$$\begin{aligned}
Tr[I] &= 4 \\
Tr[\gamma^\mu] &= 0.
\end{aligned} \tag{446}$$

We will use the *Feynman slash* in QED expressions:

$$\not{p} = \gamma\, p \equiv \gamma^\mu p_\mu = \beta p_0 - \boldsymbol{\gamma}\mathbf{p}. \tag{447}$$

Another important properties of the Dirac matrices are given by the adjoints. One can easily show that

$$(\gamma^\mu)^\dagger = \gamma^0\gamma^\mu\gamma^0. \tag{448}$$

Using the commutation rules, one can see easily that $\left(\gamma^0\right)^\dagger = \gamma^0$ and $\left(\gamma^i\right)^\dagger = -\gamma^i$.

Asymptotic Properties

This section gives an example that indicates that the power series expansion of the wave function discussed in section "Singular Terms: Renormalization in Coordinate Space" leads to an asymptotic expansion for large values of $|z|$ of the integrand in Eq. (190). A simple expression with the essential features of (190) is

$$u(y) = \int d\mathbf{x}_2 \int d\mathbf{x}_1\, f(\mathbf{x}_2)\,[a + y\mathbf{b}\cdot(\mathbf{x}_2 - \mathbf{x}_1)]\,\frac{e^{-y|\mathbf{x}_2-\mathbf{x}_1|}}{4\pi|\mathbf{x}_2-\mathbf{x}_1|^2}\,g(\mathbf{x}_1). \tag{449}$$

If $g(\mathbf{x}_1)$ is expanded about the point $\mathbf{x}_2$, we have

$$g(\mathbf{x}_1) = g(\mathbf{x}_2) + (\mathbf{x}_1-\mathbf{x}_2)\cdot\boldsymbol{\nabla}_2\, g(\mathbf{x}_2) + \frac{1}{2}(x_1^l - x_2^l)(x_1^m - x_2^m)\frac{\partial}{\partial x_2^l}\frac{\partial}{\partial x_2^m}g(\mathbf{x}_2) + \cdots. \tag{450}$$

Term by term integration over $\mathbf{x}_1$ in (449) is elementary and yields

$$u(y) = \frac{a}{y}\int d\mathbf{x}\, f(\mathbf{x})\, g(\mathbf{x}) - \frac{2}{3y^2}\int d\mathbf{x}\, f(\mathbf{x})\,\mathbf{b}\cdot\boldsymbol{\nabla}\, g(\mathbf{x}) + \frac{a}{3y^3}\int d\mathbf{x}\, f(\mathbf{x})\nabla^2 g(\mathbf{x}) + \cdots, \tag{451}$$

where the three terms correspond to the three terms in (450). Since gradients of the wave function are proportional to the momentum, the series depicted above is a power series in $\mathbf{p}$ or $Z\alpha$, as well as an asymptotic series in y^{-1}.

Integration of Singular Terms

In section "Evaluation of the Singular Terms," the result of integration over z is given for a number of singular terms in the high-energy part. Here, we indicate a method of evaluation of the integrals for one example:

$$I = \frac{1}{i}\int_{C_H} dz\, z\left(\frac{1}{b+c} - \frac{1}{b'+c}\right). \tag{452}$$

Changes of variables $y = -iz$ on the positive imaginary axis and $y = iz$ on the negative imaginary axis lead to

$$I = 2\,\mathrm{Im}\int_0^\infty dy\; y\left[\frac{1}{y - iE_n + (1+y^2)^{1/2}} - \frac{1}{(\Lambda^2 + (y - iE_n)^2)^{1/2} + (1+y^2)^{1/2}}\right]. \tag{453}$$

In the second term in (453), we make the replacement $(1+y^2)^{1/2} \to y$ with a resulting change in the integral of order Λ^{-1}. To integrate each of the two terms in (453) separately, we introduce a temporary cutoff

$$I = 2\,\mathrm{Im}\lim_{Y\to\infty}\int_0^Y dy\; y\left[\frac{1}{y - iE_n + (1+y^2)^{1/2}} - \frac{1}{(\Lambda^2 + (y - iE_n)^2)^{1/2} + y}\right] + \mathcal{O}(\Lambda^{-1}) \tag{454}$$

where the limit $Y \to \infty$ is taken before the limit $\Lambda \to \infty$. Each integral can be evaluated analytically, with the result for large Y that

$$2\,\mathrm{Im}\int_0^Y dy\; \frac{y}{y - iE_n + (1+y^2)^{1/2}} = \frac{E_n}{4}\left[2\,\ln(2Y) - \frac{1}{E_n^2} + \frac{1 - E_n^4}{E_n^4}\ln(1+E_n^2)\right] + \mathcal{O}(Y^{-1}) \tag{455}$$

and

$$2\,\mathrm{Im}\int_0^Y dy\; \frac{y}{(\Lambda^2 + (y - iE_n)^2)^{1/2} + y} = \frac{E_n}{4}\left[2\,\ln(2Y) - \ln\Lambda^2 - \frac{\Lambda^2}{E_n^2} - \left(\frac{\Lambda^2}{E_n^2} - 1\right)^2 \ln\left(1 - \frac{E_n^2}{\Lambda^2}\right)\right] + \mathcal{O}(Y^{-1}). \tag{456}$$

Taking the difference for large Λ yields

$$I = \frac{E_n}{4}\left[\ln(\Lambda^2) + \frac{3E_n^2 - 2}{2E_n^2} + \frac{1 - E_n^4}{E_n^4}\ln\left(1 + E_n^2\right)\right] + \mathcal{O}(\Lambda^{-1}). \tag{457}$$

The remaining integrals in Section "Evaluation of the Singular Terms," can be evaluated in this way.

Angular Integrations

Formulas pertaining to integration over $d\Omega_1$ in section "Evaluation of the Subtraction Terms" are given here. The calculation is facilitated by expressing the integral in terms of spherical angles θ and ϕ of $\mathbf{x}_1$ relative to the direction of $\mathbf{x}_2$. In particular, we write

$$\mathbf{x}_1 = x_1 \cos\phi \sin\theta\, \hat{a} + x_1 \sin\phi \sin\theta\, \hat{b} + x_1 \cos\theta\, \hat{x}_2 \tag{458}$$

where $\hat{a}$ and $\hat{b}$ are orthogonal unit vectors in the plane perpendicular to $\hat{x}_2$. For $\xi = \cos\theta$, $R = (x_2^2 - 2x_2 x_1 \xi + x_1^2)^{1/2}$, and f a function of R, we have

$$\begin{aligned}\int d\Omega_1 (\mathbf{x}_1 - \mathbf{x}_2)\, f(R) &= \int_{-1}^{1} d\xi \int_0^{2\pi} d\phi \,\big(x_1 \cos\phi \sqrt{1-\xi^2}\,\hat{a} \\ &\quad + x_1 \sin\phi \sqrt{1-\xi^2}\,\hat{b} + x_1 \xi\, \hat{x}_2 - \mathbf{x}_2\big)\, f(R) \\ &= 2\pi \int_{-1}^{1} d\xi (\xi x_1 - x_2)\hat{x}_2\, f(R)\end{aligned} \tag{459}$$

and

$$\begin{aligned}\int d\Omega_1 (x_1^l - x_2^l)(x_1^m - x_2^m)\, f(R) &= \int_{-1}^{1} d\xi \int_0^{2\pi} d\phi \,\big(x_1 \cos\phi \sqrt{1-\xi^2}\,\hat{a}^l \\ &\quad + x_1 \sin\phi \sqrt{1-\xi^2}\,\hat{b}^l + x_1 \xi\, \hat{x}_2^l - x_2^l\big) \\ &\quad \times \big(x_1 \cos\phi \sqrt{1-\xi^2}\,\hat{a}^m \\ &\quad + x_1 \sin\phi \sqrt{1-\xi^2}\,\hat{b}^m \\ &\quad + x_1 \xi\, \hat{x}_2^m - x_2^m\big)\, f(R) \\ &= 2\pi \int_{-1}^{1} d\xi \big[x_1^2 \tfrac{1}{2}(1-\xi^2)(\hat{a}^l \hat{a}^m + \hat{b}^l \hat{b}^m) \\ &\quad + (x_1 \xi - x_2)^2 \hat{x}_2^l \hat{x}_2^m\big] f(R) \\ &= 2\pi \int_{-1}^{1} d\xi\, f(R) \Big[R^2 \hat{x}_2^l \hat{x}_2^m \\ &\quad + \tfrac{1}{2}(1-\xi^2) x_1^2 (\delta_{lm} - 3\hat{x}_2^l \hat{x}_2^m)\Big].\end{aligned} \tag{460}$$

References

1. Jauch JM, Rohrlich F (1976) The theory of photons and electrons: the relativistic quantum field theory of charged particles with spin one-half. Texts and monographs in physics, 2nd edn. Springer, New York
2. Indelicato P, Shabaev VM, Volotka AV (2004) Interelectronic-interaction effect on the transition probability in high-Z He-like ions. Phys Rev A 69(6):062506. http://link.aps.org/abstract/PRA/v69/e062506
3. Dirac PAM (1928) The quantum theory of the electron. Proc R Soc Lond Ser A 117:610–624

4. Yerokhin VA, Shabaev VM (2015) Nuclear recoil effect in the lamb shift of light hydrogen-like atoms. J Phys Chem Ref. Data 44(3):033103. http://scitation.aip.org/content/aip/journal/jpcrd/44/3/10.1063/1.4927487
5. Artemyev AN, Shabaev VM, Yerokhin VA, Plunien G, Soff G (2005) QED calculation of the n = 1 and n = 2 energy levels in He-like ions. Phys Rev A 71(6):062104. http://link.aps.org/abstract/PRA/v71/e062104
6. Yerokhin VA, Artemyev AN, Shabaev VM (2007) QED treatment of electron correlation in Li-like ions. Phys Rev A 75:062501. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.75.062501
7. Yerokhin VA, Surzhykov A (2012) Relativistic configuration-interaction calculation of energy levels of core-excited states in lithiumlike ions: argon through krypton. Phys Rev A 86(4):042507. http://link.aps.org/doi/10.1103/PhysRevA.86.042507
8. Malyshev AV, Volotka AV, Glazov DA, Tupitsyn II, Shabaev VM, Plunien G (2014) QED calculation of the ground-state energy of berylliumlike ions. Phys Rev A 90(6):062517. http://link.aps.org/doi/10.1103/PhysRevA.90.062517
9. Yerokhin VA, Surzhykov A, Fritzsche S (2014) Relativistic configuration-interaction calculation of Ka transition energies in beryllium-like iron. Phys Rev A 90(2):022509. http://link.aps.org/doi/10.1103/PhysRevA.90.022509
10. Yerokhin VA, Surzhykov A, Fritzsche S (2015) Relativistic configuration-interaction calculation of K ? transition energies in beryllium-like argon. Phys Scr 90(5):054003. http://stacks.iop.org/1402-4896/90/i=5/a=054003
11. Artemyev AN, Shabaev VM, Tupitsyn II, Plunien G, Surzhykov A, Fritzsche S (2013) Ab initio calculations of the 2p3/2-2p1/2 fine-structure splitting in boronlike ions. Phys Rev A 88(3):032518. http://link.aps.org/doi/10.1103/PhysRevA.88.032518
12. Dyall KG, Grant IP, Johnson CT, Parpia FA, Plummer EP (1989) GRASP: a general-purpose relativistic atomic structure program. Comput Phys Commun 55:425
13. Indelicato P, Gorceix O, Desclaux JP (1987) MCDF studies of two electron ions II: radiative corrections and comparison with experiment. J Phys B At Mol Opt Phys 20(4):651. http://dx.doi.org/10.1088/0022-3700/20/4/007
14. Indelicato P, Desclaux JP (1990) Multiconfiguration Dirac-Fock calculations of transition energies with QED corrections in three-electron ions. Phys Rev A 42(9):5139. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.42.5139
15. Cheng KT, Johnson WR, Sapirstein J (1993) Lamb-shift calculations for non-Coulomb potentials. Phys Rev A 47(3):1817. http://link.aps.org/doi/10.1103/PhysRevA.47.1817
16. Cheng KT, Chen H, Johnson WR, Sapirstein J (2008) High-precision relativistic atomic structure calculations and the EBIT: Tests of quantum electrodynamics in highly charged ions. Can J Phys 86(1):33
17. Shabaev VM, Tupitsyn II, Yerokhin VA (2013) Model operator approach to the Lamb shift calculations in relativistic many-electron atom. Phys Rev A 88(1):012513. http://link.aps.org/doi/10.1103/PhysRevA.88.012513
18. Shabaev VM, Tupitsyn II, Yerokhin VA (2015) QEDMOD: Fortran program for calculating the model Lamb-shift operator. Comput Phys Commun 189:175. http://www.sciencedirect.com/science/article/pii/S0010465514004081
19. Pauli W (1941) Relativistic field theories of elementary particles. Rev Mod Phys 13(3):203
20. Foldy LL (1952) The electromagnetic properties of Dirac particles. Phys Rev 87(5):688. http://journals.aps.org/pr/abstract/10.1103/PhysRev.87.688
21. Cheng KT, Childs WJ (1985) Ab initio calculation of $4f^n\ 6s^2$ hyperfine structure in neutral rare-earth atoms. Phys Rev A 31(5):2775
22. Boucard S, Indelicato P (2000) Relativistic many-body and QED effects on the hyperfine structure of lithium-like ions. Eur Phys J D 8(1):59
23. Martin PC, Glauber RJ (1958) Relativistic theory of radiative orbital electron capture. Phys Rev 109(4):1307

24. Rose ME (1961) Relativistic electron theory. Wiley, New York
25. Baym G (1969) Lectures in quantum mechanics. Benjamin, Cumming
26. Johnson MH, Lippmann BA (1950) Relativistic keppler problem. Phys Rev (A) 78:329. http://journals.aps.org/pr/abstract/10.1103/PhysRev.78.313
27. Indelicato P, Mohr PJ (1992) Coordinate space approach to the one-electron self-energy. Phys Rev A 46(1):172
28. Mohr PJ, Kim YK (1992) Self-energy of excited states in a strong Coulomb field. Phys Rev A 45(5):2727. http://link.aps.org/doi/10.1103/PhysRevA.45.2727
29. Mohr PJ (1974) Self-energy radiative corrections in hydrogen-like systems. Ann Phys 88:26
30. Mohr PJ (1982) Self-energy of the n=2 states in a strong Coulomb field. Phys Rev A 26(5):2338
31. Lamb WE, Retherford RC (1947) Fine structure of the hydrogen atom by a microwave method. Phys Rev 72(3):241. http://link.aps.org/doi/10.1103/PhysRev.72.241
32. Lamb WE Jr, Retherford RC (1950) Fine structure of the hydrogen atom. Part I. Phys Rev 79(4):549. http://journals.aps.org/pr/abstract/10.1103/PhysRev.79.549
33. Kusch P, Foley HM (1948) Phys. The magnetic moment of the electron. Rev. 74(3): 250
34. Bethe HA (1947) The electromagnetic shift of energy levels. Phys Rev 72(4):339. http://journals.aps.org/pr/abstract/10.1103/PhysRev.72.339
35. Kroll NM, Lamb JW (1949) On the self-energy a the bound electron. Phys Rev 75(3): 388
36. Feynman RP (1948) A relativistic cut-off for quantum electrodynamics. Phys Rev 74(10):1430. http://journals.aps.org/pr/abstract/10.1103/PhysRev.74.1430
37. Feynman RP (1949) The theory of positrons. Phys Rev 76(6):749. http://journals.aps.org/pr/abstract/10.1103/PhysRev.76.749
38. Feynman RP (1949) Space-time approach to quantum electrodynamics. Phys Rev 76(6):769. http://journals.aps.org/pr/abstract/10.1103/PhysRev.76.769
39. Schwinger J (1948) On quantum-electrodynamics and the magnetic moment of the electron. Phys Rev 73:416. http://journals.aps.org/pr/abstract/10.1103/PhysRev.73.416
40. Schwinger J (1948) Quantum electrodynamics. I. A covariant formulation. Phys Rev 74(10):1439. http://journals.aps.org/pr/abstract/10.1103/PhysRev.74.1439
41. Schwinger J (1949) Quantum electrodynamics. II. Vacuum polarization and self-energy. Phys Rev 75(4):651. http://link.aps.org/doi/10.1103/PhysRev.75.651
42. Schwinger J (1949) Quantum electrodynamics. III. The electromagnetic properties of the electron-radiative corrections to scattering. Phys Rev 76(6):790. http://journals.aps.org/pr/abstract/10.1103/PhysRev.76.790
43. Dyson FJ (1949) The S matrix in quantum electrodynamics. Phys Rev 75(11):1736. http://journals.aps.org/pr/abstract/10.1103/PhysRev.75.1736
44. Dyson FJ (1949) The radiation theories of Tomonaga, Schwinger, and Feynman. Phys Rev 75(3):486. http://journals.aps.org/pr/abstract/10.1103/PhysRev.75.486
45. Brown GE, Ravenhall DE (1951) On the interaction of two electrons. Proc R Soc Lond Ser A 208:552
46. Sucher J (1980) Foundation of the relativistic theory of many-electron atoms. Phys Rev A 22(2):348
47. Mittleman MH (1981) Theory of relativistic effects on atoms: configuration-space hamiltonian. Phys Rev A 24(3):1167
48. Heully JL, Lindgren I, lindroth E, Mårtensson-Pendrill AM (1986) Comment on relativistic wave equation and negative-energy state. Phys Rev A 33(6):4426
49. Grant IP (1987) On the relativistic electron-electron interaction. J Phys B At Mol Phys 20:L735
50. Indelicato P (1995) Projection operators in multiconfiguration Dirac-Fock calculations. Application to the ground state of heliumlike ions. Phys Rev A 51(2):1132

51. Mohr PJ (1989) Quantum electrodynamics of high-Z few-electron atoms. In: Marrus R (ed) Physics of Highly-Ionized Atoms. NATO ASI Series, vol 201. Plenum, New York, pp 111–141
52. Mohr PJ (1989) Relativistic, quantum electrodynamic, and weak interaction effects in atoms : Santa Barbara, CA 1988. In: Johnson WR, Mohr PJ, Sucher J (eds) Proceeding of the Institute for Theoretical Physics (Santa Barbara). AIP Conference Proceedings, vol 189. American Institute of Physics, New York
53. Mohr PJ, Plunien G, Soff G (1998) QED corrections in heavy atoms. Phys Rep 293(5&6), 227
54. Cohen-Tannoudji C, Diu B, Laloë F (1996) Mécanique quantique. Hermann, Paris
55. Furry WH (1951) On bound states and scattering in Positrons theory. Phys Rev 81(1):115. http://journals.aps.org/pr/abstract/10.1103/PhysRev.81.115
56. Mohr PJ (1985) Quantum electrodynamics of high-Z few-electron atoms. Phys Rev A 32(4):1949
57. Gell-Mann M, Low F (1951) Bound states in quantum field theory. Phys Rev 84(2):350
58. Fetter AL, Walecka JD (1971) Quantum theory of many-particle systems. International series in pure and applied physics. Mc Graw-Hill, New-York
59. Sucher J (1957) S-matrix formalism for level-shift calculations. Phys Rev 107:1448
60. Mohr PJ (1996) QED in one electron systems. In: Drake GWF (ed) Handbook of atomic physics. American Institute of Physics, New-York, p 125
61. Dyson FJ (1952) Divergence of perturbation theory in quantum electrodynamics. Phys Rev 85(4):631. http://journals.aps.org/pr/abstract/10.1103/PhysRev.85.631
62. Wick GC (1950) The evaluation of the collision matrix. Phys Rev 80(2):268. http://link.aps.org/doi/10.1103/PhysRev.80.268
63. Mathematica Edition: Version 10 (2015) Wolfram Research, Inc., Champaign. http://www.wolfram.com
64. Helton JW, Miller RL, Stankus M (1996) Ncalgebra: a mathematica package for doing noncommuting algebra. http://math.ucsd.edu/~ncalg
65. Lindgren I, Persson H, Salomonson S, Karasiev V, Labzowsky L, Mitrushenkov A, Tokman M (1993) Second-order QED corrections for few-electron heavy ions: reducible Breit-Coulomb correction and mixed self-energy- vacuum polarization correction. J Phys B: At Mol Opt Phys 26:L503
66. Persson H, Lindgren I, Labzowsky LN, Plunien G, Beier T, Soff G (1996) Second-order self-energy-vacuum-polarization contributions to the Lamb shift in highly charged few-electron ions. Phys Rev A 54(4):2805
67. Yerokhin VA, Indelicato P, Shabaev VM (2003) Two-loop self-energy correction in high-Z hydrogen-like ions. Phys Rev Lett 91(7):073001
68. Yerokhin VA, Indelicato P, Shabaev VM (2003) Evaluation of the two-loop self-energy correction to the ground state energy of H-like ions to all orders in Z? Eur Phys J D 25(3): 203
69. Yerokhin VA, Indelicato P, Shabaev VM (2005) Two-loop self-energy contribution to the Lamb shift in H-like ions. Phys Rev A 71(4):040101(R). http://link.aps.org/abstract/PRA/v71/e040101
70. Yerokhin VA, Indelicato P, Shabaev VM (2005) Two-loop self-energy correction in a strong Coulomb nuclear field. J Exp Theor Phys 101(2):280–293
71. Yerokhin VA, Indelicato P, Shabaev VM (2006) Nonperturbative calculation of the two-loop Lamb shift in Li-like ions. Phys Rev Lett 97(25):253004. http://link.aps.org/abstract/PRL/v97/e253004
72. Yerokhin VA, Indelicato P, Shabaev VM (2007) Two-loop QED corrections in few-electron ions. Can J Phys 85(5):521. http://dx.doi.org/10.1140/epjd/e2006-00064-8
73. Yerokhin VA, Indelicato P, Shabaev VM (2008) Two-loop QED corrections with closed fermion loops. Phys Rev A 77(6):062510 (12)
74. Yerokhin VA (2009) Two-loop self-energy for the ground state of medium-Z hydrogenlike ion. Phys Rev A 80(4):040501. http://link.aps.org/abstract/PRA/v80/e040501

75. Yerokhin VA (2010) The two-loop self-energy: diagrams in the coordinate-momentum representation. Eur Phys J D 58(1):57. http://dx.doi.org/10.1140/epjd/e2010-00089-4
76. Källén G, Sabry A (1955) Fourth order vacuum polarization. Det Kongelige Danske Videnskabernes Selskab Matematisk-Fysiske Meddelelser 29(17):3
77. Fullerton LW, Rinker GA (1976) Accurate and efficient methods for the evaluation of vacuum-polarization potentials of order Za and Za2. Phys Rev A 13(3):1283
78. Artemyev AN, Shabaev VM, Yerokhin VA (1997) Vacuum polarization screening corrections to the ground-state energy of two-electron ions. Phys Rev A 56(5):3529. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.56.3529
79. Artemyev AN, Beier T, Plunien G, Shabaev VM, Soff G, Yerokhin VA (2000) Vacuum-polarization screening corrections to the energy levels of heliumlike ions. Phys Rev A 62(2):022116. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.62.022116
80. Artemyev AN, Beier T, Plunien G, Shabaev VM, Soff G, Yerokhin VA (1999) Vacuum-polarization screening corrections to the energy levels of lithiumlike ions. Phys Rev A 60(1):45. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.60.45
81. Indelicato P, Mohr PJ (2001) Coordinate-space approach to the bound-electron self-energy: Self-Energy screening calculation. Phys Rev A 63(4):052507
82. Yerokhin VA, Artemyev AN, Beier T, Plunien G, Shabaev VM, Soff G (1999) Two-electron self-energy corrections to the 2p1/2-2s transition energy in Li-like ions. Phys Rev A 60(5):3522. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.60.3522
83. Blundell SA, Mohr PJ, Johnson WR, Sapirstein J (1993) Evaluation of two-photon exchange graphs for highly charged heliumlike ions. PRA 48(4):2615. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.48.2615
84. Lindgren I, Persson H, Salomonson S, Labzowsky L (1995) Full QED calculations of two-photon exchange for heliumlike systems : analysis in the coulomb and feynman gauge. Phys Rev A 51(2):1167
85. Mohr PJ, Sapirstein J (2000) Evaluation of two-photon exchange graphs for excited states of highly charged Heliumlike ions. Phys Rev A 62(5):052501
86. Yerokhin VA, Artemyev AN, Shabaev VM, Sysak MM, Zherebtsov OM, Soff G (2001) Evaluation of the two-photon exchange graphs for the $2p_{1/2} - 2s$ transition in Li-like ions. Phys Rev A 64(3):032109. http://link.aps.org/doi/10.1103/PhysRevA.64.032109
87. øAsén B, Salomonson S, Lindgren I (2002) Two-photon-exchange QED effects in the 1s2s 1S and 3S states of heliumlike ions. Phys Rev A 65(3):032516 (16)
88. Artemyev AN, Shabaev VM, Sysak MM, Yerokhin VA, Beier T, Plunien G, Soff G (2003) Evaluation of the two-photon exchange diagrams for the (1s)22p3/2 electron configuration in Li-like ions. Phys Rev A 67(6):062506 (7). http://journals.aps.org/pra/abstract/10.1103/PhysRevA.67.062506
89. Sapirstein J, Cheng KT (2011) S-matrix calculations of energy levels of the lithium isoelectronic sequence. Phys Rev A 83(1):012504. http://link.aps.org/doi/10.1103/PhysRevA.83.012504
90. Sapirstein J, Cheng KT (2015) S-matrix calculations of energy levels of sodiumlike ions. Phys Rev A 91(6):062508. http://link.aps.org/doi/10.1103/PhysRevA.91.062508
91. Brown GE, Langer JS, Schaefer GW (1959) Lamb shift of a tightly bound electron: I. method. Proc R Soc Lond Ser A 251:92
92. Desiderio AM, Johnson WR (1971) Lamb Shift and binding energies of K electron in heavy atoms. Phys Rev A 3(4):1267
93. Snyderman NJ (1991) Electron radiatives self-energy of highly stripped heavy atoms. Ann Phys 211:43
94. Blundell SA, Snyderman NJ (1991) Basis set approach to calculating the radiative self-energy in highly ionized atoms. Phys Rev A 44(3):R1427
95. Cheng KT, Johnson WR, Sapirstein J (1991) Screened Lamb shift calculations for lithiumlike uranium, sodiumlike platinum and copperlike gold. Phys Rev Lett 66(23):2960
96. Le Bigot EO, Indelicato P, Mohr PJ (2001) QED self-energy contribution to highly-excited atomic states. Phys Rev A 64(5):052508 (14)

97. Jentschura UD, Mohr PJ, Soff G (1999) Calculation of the electron self-energy for low nuclear charge. Phys Rev Lett 82(1):53
98. Jentschura UD, Mohr PJ, Soff G (2001) Electron self-energy for the K and L shells at low nuclear charge. Phys Rev A 63(4):042512
99. Jentschura UD, Mohr PJ, Soff G, Weniger EJ (1999) Convergence acceleration via combined nonlinear-condensation transformations. Comput Phys Commun 116(1):28. http://www.sciencedirect.com/science/article/B6TJ5-3W3FP53-3/2/42586ad49205db06e1b5c90230bd7150
100. Indelicato P, Mohr PJ (1998) Coordinate-space approach to the bound-electron self energy: Coulomb field calculation. Phys Rev A 58(1):165
101. Indelicato P, Mohr PJ, Sapirstein J (2014) Coordinate-space approach to vacuum polarization. Phys Rev A 89(4):042121 http://link.aps.org/doi/10.1103/PhysRevA.89.042121
102. Fried HM, Yennie DR (1958) New techniques in the Lamb shift calculation. Phys Rev 112(4):1391. http://link.aps.org/doi/10.1103/PhysRev.112.1391
103. Coquereaux R (1980) Renormalization schemes in QED. Ann Phys 125(2):401. http://www.sciencedirect.com/science/article/pii/0003491680901396
104. Narison S (1982) Techniques of dimensional regularization and the two-point functions of QCD and QED. Phys Rep 84(4):263. http://www.sciencedirect.com/science/article/pii/0370157382900230
105. Blundell SA, Cheng KT, Sapirstein J (1997) Radiative corrections in atomic physics in the presence of perturbing potentials. Phys Rev A 55(3):1857. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.55.1857
106. Pauli W, Villars F (1949) On the invariant regularization in relativistic quantum theory. Rev Mod Phys 21(3):434. http://journals.aps.org/rmp/abstract/10.1103/RevModPhys.21.434
107. Mohr PJ, Sapirstein JR (1985) Quantum electrodynamic energy shifts of quarks bound in a cavity. Phys Rev Lett 54(6):514
108. Mohr PJ (1987) In: Greiner W (ed) Physics of strong fields. Plenum, New York, p 17
109. Wichmann EH, Kroll NM (1956) Vacuum polarization in a strong Coulomb field. Phys Rev 101(2):843
110. Uehling EA (1935) Polarization effects in the positron theory. Phys Rev 48(1):55
111. Serber R (1935) Linear modifications in the Maxwell field equations. Phys Rev 48(1):49. http://link.aps.org/doi/10.1103/PhysRev.48.49
112. Huang KN (1976) Calculation of the vacuum-polarization potential. Phys Rev A 14(4):1311
113. Klarsfeld S (1977) Analytical expression for the evaluation of vacuum-polarization potentials in muonic atoms. Phys Lett 66B(1), 86
114. Blomqvist J (1972) Vacuum polarization in exotic atoms. Nucl Phys B 48:95
115. Hylton DJ (1985) Finite-nuclear-size corrections to the Uehling potential. Phys Rev A 32(3):1303. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.32.1303
116. Soff G, Mohr PJ (1988) Vacuum polarization in a strong external field. Phys Rev A 38(10):5066
117. Persson H, Lindgren I, Salomonson S, Sunnergren P (1993) Accurate vacuum-polarization calculations. Phys Rev A 48(4):2772. http://link.aps.org/doi/10.1103/PhysRevA.48.2772
118. Karplus R, Neuman M (1950) Non-linear interactions between electromagnetic fields. Phys Rev 80(3):380. http://link.aps.org/doi/10.1103/PhysRev.80.380
119. Hylton DJ (1984) The reduced Dirac-Green function for the Coulomb potential. J Math Phys 25(4):1125. http://scitation.aip.org/content/aip/journal/jmp/25/4/10.1063/1.526255
120. Lindgren I, Morrison J (1982) Atomic many-body theory. Atoms and plasmas, 2nd edn. Springer, Berlin
121. Kelly HP (1963) Correlation effects in atoms. Phys Rev 131(2):684. http://journals.aps.org/pr/abstract/10.1103/PhysRev.131.684
122. Grant IP (1965) Relativistic self-consistent fields. Proc Phys Soc Lond 86:523
123. Grant IP (1970) Relativistic calculation of atomic structures. Adv Phys 19(82):747
124. Desclaux JP, Mayers DF, O'Brien F (1971) Relativistic atomic wavefunction. J Phys B: At Mol Optic Phys 4:631

125. Desclaux JP (1975) A multiconfiguration relativistic Dirac-Fock program. Comput Phys Commun 9:31
126. Lindgren I (1978) A coupled-cluster approach to the many-body perturbation theory for open-shell systems. Int J Quantum Chem 14(S12), 33. http://dx.doi.org/10.1002/qua.560140804
127. Johnson WR, Sapirstein J (1986) Computation of second-order many-body corrections in relativistic atomic systems. Phys Rev Lett 57(9):1126. http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.57.1126
128. Pohl R, Antognini A, Nez F, Amaro FD, Biraben F, Cardoso JMR, Covita DS, Dax A, Dhawan S, Fernandes LMP, Giesen A, Graf T, Hänsch TW, Indelicato P, Julien L, Kao CY, Knowles P, Bigot EOL, Liu YW, Lopes JAM, Ludhova L, Monteiro CMB, Mulhauser F, Nebel T, Rabinowitz P, dos Santos JMF, Schaller LA, Schuhmann K, Schwob C, Taqqu D, Veloso JFCA, Kottmann F (2010) The size of the proton. Nature 466(7303):213. http://www.nature.com/doifinder/10.1038/nature09250
129. Antognini A, Nez F, Schuhmann K, Amaro FD, Biraben F, Cardoso JMR, Covita DS, Dax A, Dhawan S, Diepold M, Fernandes LMP, Giesen A, Gouvea AL, Graf T, Hänsch TW, Indelicato P, Julien L, Kao CY, Knowles P, Kottmann F, Le Bigot EO, Liu YW, Lopes JAM, Ludhova L, Monteiro CMB, Mulhauser F, Nebel T, Rabinowitz P, dos Santos JMF, Schaller LA, Schwob C, Taqqu D, Veloso JFCA, Vogelsang J, Pohl R (2013) Proton structure from the measurement of 2S-2P transition frequencies of muonic hydrogen. Science 339(6118):417. http://www.sciencemag.org/content/339/6118/417.abstract
130. Trassinelli M, Kumar A, Beyer HF, Indelicato P, Märtin R, Reuschl R, Kozhedu YS, Brandau C, Bräuning H, Geyer S, Gumberidze A, Hess S, Jagodzinski P, Kozhuharov C, Liesen D, Trotsenko S, Weber G, Winters DFA, Stöhlker T (2009) Observation of the $2p_{3/2} \rightarrow 2s_{1/2}$ intra-shell transition in He-like uranium. Europhys Lett 87(6):63001 (6). http://www.iop.org/EJ/abstract/0295-5075/87/6/63001
131. Trassinelli M, Kumar A, Beyer HF, Indelicato P, Märtin R, Reuschl R, Kozhedub YS, Brandau C, Bräuning H, Geyer S, Gumberidze A, Hess S, Jagodzinski P, Kozhuharov C, Liesen D, Spillmann U, Trotsenko S, Weber G, Winters DFA, Stöhlker T (2011) Differential energy measurement between He- and Li-like uranium intra-shell transitions. Phys Scr T144:014003. http://stacks.iop.org/1402-4896/2011/i=T144/a=014003
132. Chantler CT, Kinnane MN, Gillaspy JD, Hudson LT, Payne AT, Smale LF, Henins A, Pomeroy JM, Tan JN, Kimpton JA, Takacs E, Makonyi K (2012) Testing three-body quantum electrodynamics with trapped Ti^{20+} ions: evidence for a Z-dependent divergence between experiment and calculation. Phys Rev Lett 109(15):153001. http://link.aps.org/doi/10.1103/PhysRevLett.109.153001
133. Kim YK, Baik DH, Indelicato P, Desclaux JP (1991) Resonance transition energies of Li-, Na-, and Cu-like ions. Phys Rev A 44(1):148. http://link.aps.org/doi/10.1103/PhysRevA.44.148
134. Blundell SA (1993) Calculations of the screened self-energy and vacuum polarization in Li-like, Na-like and Cu-like ions. Phys Rev A 47(3):1790. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.47.1790
135. Chen MH, Cheng KT, Johnson WR, Sapirstein J (1995) Relativistic configuration-interaction calculations for the n=2 states of lithiumlike ions. Phys Rev A 52(1):266. http://link.aps.org/doi/10.1103/PhysRevA.52.266
136. Cheng KT, Chen MH, Sapirstein J (2000) Quantum electrodynamic corrections in high-Z Li-like and Be-like ions. Phys Rev A 62(3):054501. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.62.054501
137. Kozhedub YS, Volotka AV, Artemyev AN, Glazov DA, Plunien G, Shabaev VM, Tupitsyn II, Stöhlker T (2010) Relativistic recoil, electron-correlation, and QED effects on the 2pj-2s transition energies in Li-like ions. Phys Rev A 81(4):042513. http://link.aps.org/doi/10.1103/PhysRevA.81.042513
138. Hölzer G, Förster E, Klöpfel D, Beiersdorfer P, Brown GV, Crespo López-Urrutia JR, Widman K (1998) Absolute wavelength measurement of the Lyman-a transition of hydrogenic Mg^{11+}. PRA 57(2):945. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.57.945

139. Tschischgale J, Klöpfel D, Beiersdorfer P, Brown GV, Förster E, Schulte-Schrepping H, Utter SB (2002) Absolute wavelength measurement of the Lyman-? transition of hydrogen-like silicon. CJP 80(8):867. http://rparticle.web-p.cisti.nrc.ca/rparticle/AbstractTemplateServlet?calyLang=eng&journal=cjp&volume=80&year=0&issue=8&msno=p02-011
140. Kubiçek K, Mokler PH, Mäckel V, Ullrich J, López-Urrutia JRC (2014) Transition energy measurements in hydrogenlike and heliumlike ions strongly supporting bound-state QED calculations. Phys Rev A 90(3):032508. http://link.aps.org/doi/10.1103/PhysRevA.90.032508
141. Briand JP, Mossé JP, Indelicato P, Chevallier P, Girard-Vernhet D, Chétioui A, Ramos MT, Desclaux JP (1983) Spectroscopy of hydrogenlike and heliumlike argon. Phys Rev A 28(3):1413. http://link.aps.org/doi/10.1103/PhysRevA.28.1413
142. Beyer HF, Deslattes RD, Folkmann F, LaVilla RE (1985) Determination of the 1s Lamb shift in one-electron argon recoil ions. J Phys B: At Mol Opt Phys 18(2):207. http://dx.doi.org/10.1088/0022-3700/18/2/008
143. Rice JE, Reinke ML, Ashbourn JMA, Gao C, Victora MM, Chilenski MA, Delgado-Aparicio L, Howard NT, Hubbard AE, Hughes JW, Irby JH (2014) X-ray observations of Ca^{19+}, Ca^{18+} and satellites from Alcator C-Mod tokamak plasmas. J Phys B: At Mol Opt Phys 47(7):075701. http://stacks.iop.org/0953-4075/47/i=7/a=075701
144. Tarbutt MR, Silver JD (2002) Measurement of the ground-state Lamb shift of hydrogen-like Ti 21+. J Phys B: At Mol Opt Phys 35(6):1467. http://stacks.iop.org/0953-4075/35/i=6/a=305
145. Gillaspy JD, Chantler CT, Paterson D, Hudson LT, Serpa FG, Takács E (2010) First measurement of Lyman alpha x-ray lines in hydrogen-like vanadium: results and implications for precision wavelength metrology and tests of QED. J Phys B: At Mol Opt Phys 43(7):074021. http://stacks.iop.org/0953-4075/43/i=7/a=074021
146. Briand JP, Tavernier M, Indelicato P, Marrus R, Gould H (1983) High precision spectroscopic studies of Lyman a lines of hydrogenlike iron: a measurement of the 1s Lamb-shift. Phys Rev Lett 50(11):832. http://link.aps.org/doi/10.1103/PhysRevLett.50.832
147. Chantler CT, Laming JM, Dietrich DD, Hallett WA, McDonald R, Silver JD (2007) Hydrogenic Lamb shift in iron Fe25+ and fine-structure Lamb shift. Phys Rev A 76(4):042116. http://link.aps.org/doi/10.1103/PhysRevA.76.042116
148. Beyer HF, Indelicato P, Finlayson KD, Liesen D, Deslattes RD (1991) Measurement of the 1s lamb-shift in hydrogenlike nickel. Phys Rev A 43(1):223. http://link.aps.org/doi/10.1103/PhysRevA.43.223
149. Chantler CT, Laming JM, Silver JD, Dietrich DD, Mokler PH, Finch EC, Rosner SD (2009) Hydrogenic Lamb shift in Ge31+ and the fine-structure Lamb shift. Phys Rev A 80(2):022508. http://link.aps.org/doi/10.1103/PhysRevA.80.022508
150. Tavernier M, Briand JP, Indelicato P, Liesen D, Richard P (1985) Measurement of the 1s Lamb shift of hydrogenlike krypton. J Phys B: At Mol Opt Phys 18(11):L327. http://dx.doi.org/10.1088/0022-3700/18/11/004
151. Briand JP, Indelicato P, Simionovici A, San Vicente V, Liesen D, Dietrich D (1989) Spectroscopic study of hydrogenlike and heliumlike xenon ions. Europhys Lett 9(3):225. http://iopscience.iop.org/0295-5075/9/3/007
152. Widmann K, Beiersdorfer P, Brown GV, Löpez Urrutia JRC, Osterheld AL, Reed KJ, Scofield JH, Utter SB (2000) High-resolution measurements of the K-shell spectral lines of hydrogenlike and heliumlike xenon. In: X-Ray and inner-shell processes, vol 506, Chicago, pp 444–66. http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.1302773
153. Beyer HF, Finlayson KD, Liesen D, Indelicato P, Chantler CT, Deslattes RD, Schweppe J, Bosch F, Jung M, Klepper O, König W, Moshammer R, Beckert K, Heuckhoff H, Franzke B, Gruber A, Nolden F, Spädtke P, Steck M (1993) X-ray transitions associated with electron capture into bare dysprosium. J Phys B: At Mol Opt Phys 26:1557. http://iopscience.iop.org/article/10.1088/0953-4075/26/9/004/meta
154. Beyer HF, Liesen D, Bosch F, Finlayson KD, Jung M, Klepper O, Moshammer R, Beckert K, Eickhoff H, Franzke B, Nolden F, Spädtke P, Steck M, Menzel G, Deslattes RD (1994) X rays from radiative electron capture of free cooling electrons for precise Lamb-shift measurements at high Z: Au78+. Phys Lett A 184:435. http://www.sciencedirect.com/

science?_ob=Article&_udi=B6TVM-46SPK2H-8S&_user=870092&_coverDate=01%2F24%2F1994&_alid=1427368629&_ rdoc=1&_fmt=high&_orig=search&_cdi=5538&_sort=r&_docanchor=&view=c&_ct=1&_acct=C000046800 &_version=1&_urlVersion=0&_userid=870092&md5=95fca9083bd953fcc051dc24ee02498b

155. Beyer HF, Menzel G, Liesen D, Gallus A, Bosch F, Deslattes RD, Indelicato P, Stöhlker T, Klepper O, Moshammer R, Nolden F, Eickhoff H, Franzke B, Steck M (1995) Measurement of the ground-state Lambshift of hydrogenlike uranium at the electron cooler of the ESR. Zeitschrift für Physik D 35:169. http://www.springerlink.com/content/h686711823m02g75/?p=83d9bb408bf8475db4f9330d57005a48&pi=1
156. Stöhlker T, Mokler PH, Bosch F, Dunford RW, Franzke F, Klepper O, Kozhuharov C, Ludziejewski T, Nolden F, Reich H, Rymuza P, Stachura Z, Steck M, Swiat P, Warczak A (2000) 1*s* Lamb shift in hydrogenlike uranium measured on cooled, decelerated ion beams. Phys Rev A 85(15):3109
157. Gumberidze A, Stöhlker T, Banaś D, Beckert K, Beller P, Beyer HF, Bosch F, Hagmann S, Kozhuharov C, Liesen D, Nolden F, Ma X, Mokler PH, Steck M, Sierpowski D, Tashenov S (2005) Quantum electrodynamics at strong electric fields: the ground state Lamb shift in hydrogenlike uranium. Phys Rev Lett 94(22):223001. http://link.aps.org/doi/10.1103/PhysRevLett.94.223001
158. Amaro P, Schlesser S, Guerra M, Le Bigot EO, Isac JM, Travers P, Santos JP, Szabo CI, Gumberidze A, Indelicato P (2012) Absolute measurement of the "relativistic M1" transition energy in heliumlike argon. Phys Rev Lett 109(4):043005. http://link.aps.org/doi/10.1103/PhysRevLett.109.043005
159. Szabo CI, Amaro P, Guerra M, Schlesser S, Gumberidze A, Santos JP, Indelicato P (2013) Reference free, high-precision measurements of transition energies in few electron argon ions. In: CAARI. AIP Conference Proceedings, vol 1525. AIP, Fort Worth, pp 68–72
160. Beiersdorfer P, Bitter M, von Goeler S, Will KW (1989) Experimental study of the x-ray transitions in the heliumlike isoelectronic sequence. Phys Rev A 40(1):150. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.40.150
161. Payne AT, Chantler CT, Kinnane MN, Gillaspy JD, Hudson LT, Smale LF, Henins A, Kimpton JA, Takacs E (2014) Helium-like titanium x-ray spectrum as a probe of QED computation. J Phys B: At Mol Opt Phys 47(18):185001. http://stacks.iop.org/0953-4075/47/i=18/a=185001
162. Briand JP, Tavernier M, Marrus R, Desclaux JP (1984) High-precision spectroscopic study of heliumlike iron. Phys Rev A 29(6):3143. http://link.aps.org/doi/10.1103/PhysRevA.29.3143
163. Rudolph JK, Bernitt S, Epp SW, Steinbrügge R, Beilmann C, Brown GV, Eberle S, Graf A, Harman Z, Hell N, Leutenegger M, Müller A, Schlage K, Wille HC, Yavaş H, Ullrich J, Crespo López-Urrutia JR (2013) X-Ray resonant photoexcitation: linewidths and energies of K alpha transitions in highly charged Fe ions. Phys Rev Lett 111(10):103002. http://link.aps.org/doi/10.1103/PhysRevLett.111.103002
164. Beiersdorfer P, Brown GV (2015) Experimental study of the x-ray transitions in the heliumlike isoelectronic sequence: updated results. Phys Rev A 91(3):032514. http://link.aps.org/doi/10.1103/PhysRevA.91.032514
165. MacLaren S, Beiersdorfer P, Vogel DA, Knapp D, Marrs RE, Wong K, Zasadzinski R (1992) Precision measurement of Ka transitions in heliumlike Ge30+. Phys Rev A 45(1):329. http://link.aps.org/doi/10.1103/PhysRevA.45.329
166. Briand JP, Indelicato P, Tavernier M, Liesen D, Beyer HF, Liu B, Warczak A, Desclaux JP (1984) Observation of hydrogenlike and heliumlike krypton spectra. Zeitschrift für Physik A 318:1
167. Indelicato P, Tavernier M, Briand JP, Liesen D (1986) Experimental study of relativistic correlations and QED effects in heliumlike krypton ions. Zeitschrift für Physik D 2: 249
168. Epp SW, Steinbrügge R, Bernitt S, Rudolph JK, Beilmann C, Bekker H, Müller A, Versolato OO, Wille HC, Yavaş H, Ullrich J, Crespo López-Urrutia JR (2015) Single-photon excitation of $K\alpha$ in heliumlike Kr^{34+}: Results supporting quantum electrodynamics predictions. Phys Rev A 92(2):020502. http://link.aps.org/doi/10.1103/PhysRevA.92.020502

169. Briand JP, Indelicato P, Simionovici A, San Vicente V, Liesen D, Dietrich D (1989) Spectroscopic study of hydrogenlike and heliumlike xenon ions. Europhys Lett 9(3):225. http://iopscience.iop.org/0295-5075/9/3/007
170. Kramida A, Ralchenko Y, Reader J, NIST ASD Team (2015) NIST Atomic Spectra Database (ver. 5.3). National Institute of Standards and Technology, Gaithersburg. [Online]. Available: http://physics.nist.gov/asd 20 Dec 2015
171. Bosselmann P, Staude U, Horn D, Schartner KH, Folkmann F, Livingston AE, Mokler PH (1999) Measurements of $2s\,^2S_{1/2} - 2p\,^2P_{1/2,3/2}$ transition energies in lithiumlike heavy ions. II. Experimental results for Ag^{44+} and discussion along the isoelectronic series. PRA 59(3):1874
172. Feili D, Bosselmann P, Schartner KH, Folkmann F, Livingston AE, Träbert E, Ma X, Mokler PH (2000) Measurements of $2s\,^2S_{1/2} - 2p\,^2P_{1/2,3/2}$ transition energies in lithiumlike heavy ions. III. Experimental results for Sn^{47+} and Xe^{51+}. Phys Rev A 62(2):022501. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.62.022501
173. Büttner R, Kraus B, Schartner KH, Folkmann F, Mokler PH, Möller G (1992) EUV spectroscopy of beam-foil excited 14.25 MeV/u Xe^{52+}?Xe^{49+} ions. Zeitschrift für Physics D 22:693. http://link.springer.com/article/10.1007/BF01437250
174. Bernhardt D, Brandau C, Harman Z, Kozhuharov C, Böhm S, Bosch F, Fritzsche S, Jacobi J, Kieslich S, Knopp H, Nolden F, Shi W, Stachura Z, Steck M, Stöhlker T, Schippers S, Müller A (2015) Electron-ion collision spectroscopy: Lithium-like xenon ions. Phys Rev A 91(1):012710. http://link.aps.org/doi/10.1103/PhysRevA.91.012710
175. Podpaly Y, Clementson J, Beiersdorfer P, Williamson J, Brown GV, Gu MF (2009) Spectroscopy of $2s_{1/2} - 2p_{3/2}$ transitions in W^{65+} through W^{71+}. Phys Rev A 80(5):052504. http://link.aps.org/doi/10.1103/PhysRevA.80.052504
176. Zhang X, Nakamura N, Chen C, Andersson M, Liu Y, Ohtani S (2008) Measurement of the QED energy shift in the $1s[sup2]2p[sub3/2] - -1s[sup2]2s[sub1/2]$ x-ray transition in Li-like $^{208}Pb^{79+}$. Phys Rev A 78(3):032504 http://link.aps.org/abstract/PRA/v78/e032504http://dx.doi.org/10.1103/PhysRevA.78.032504
177. Beiersdorfer P, Osterheld AL, Scofield JH, Crespo Löpez-Urrutia JR, Widmann K (1998) Measurement of QED and hyperfine splitting in the $2s_{1/2} - 2p_{3/2}$ X-ray transition in Li-like $^{209}Bi^{80+}$. Phys Rev Lett 80(14):3022. http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.80.3022
178. Beiersdorfer P, Osterheld A, Elliott SR, Chen MH, Knapp D, Reed K (1995) Structure and Lamb shift of 2s1/2-3p3/2 levels in lithiumlike Th^{87+} through neonlike Th^{80+}. Phys Rev A 52(4):2693. http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.71.3939
179. Nakano Y, Takano Y, Ikeda T, Kanai Y, Suda S, Azuma T, Bräuning H, Bräuning-Demian A, Dauvergne D, Stöhlker T, Yamazaki Y (2013) Resonant coherent excitation of the lithiumlike uranium ion: a scheme for heavy-ion spectroscopy. PRA 87(6):060501. http://link.aps.org/doi/10.1103/PhysRevA.87.060501
180. Brandau C, Kozhuharov C, Müller A, Shi W, Schippers S, Bartsch T, Böhm S, Böhme C, Hoffknecht A, Knopp H, Grün N, Scheid W, Steih T, Bosch F, Franzke B, Mokler PH, Nolden F, Steck M, Stöhlker T, Stachura Z (2003) Precise determination of the 2s1/2–2p1/2 splitting in very heavy lithiumlike ions utilizing dielectronic recombination. PRL 91(7):073202. http://link.aps.org/doi/10.1103/PhysRevLett.91.073202
181. Beiersdorfer P, Chen H, Thorn DB, Träbert E (2005) Measurement of the two-loop Lamb shift in lithiumlike U^{89+}. Phys Rev Lett 95(23):233003. http://link.aps.org/abstract/PRL/v95/e233003
182. Mooney T, Lindroth E, Indelicato P, Kessler E, Deslattes RD (1992) Precision measurements of K and L transitions in xenon: experiment and theory for the K, L and M levels. Phys Rev A 45(3):1531. http://link.aps.org/doi/10.1103/PhysRevA.45.1531
183. Indelicato P, Lindroth E (1992) Relativistic effects, correlation, and QED corrections on K? transitions in medium to very heavy atoms. Phys Rev A 46(5):2426

184. Indelicato P, Boucard S, Lindroth E (1998) Relativistic and many-body effects in K, L, and M shell ionization energy for elements with $10 \leq Z \leq 100$ and the determination of the 1s Lamb shift for heavy elements. Eur Phys J D 3(1):29
185. Deslattes RD, Kessler EG Jr, Indelicato P, de Billy L, Lindroth E, Anton J (2003) X-ray transition energies: new approach to a comprehensive evaluation. Rev Mod Phys 75(1):35. http://www.physics.nist.gov/PhysRefData/XrayTrans/index.html, http://link.aps.org/doi/10.1103/RevModPhys.75.35

QED Effects and Challenges

6

Anton N. Artemyev

Contents

Introduction 244
Lamb Shift 246
History 246
Modern Explanation 250
QED Effects 253
Energy Levels 253
g-Factor 255
Challenges 258
Extremely Strong Fields 259
Summary 261
References 262

Abstract

Quantum electrodynamics (QED) is the most accurate and the best confirmed theory in modern physics. This chapter is devoted to the description of the QED effects in atoms and molecules. Starting from the famous Lamb's experiment with hydrogen, we finish with the most recent experiments with heavy ions. We will demonstrate the cases where the QED effects are extremely important for the comparison of the theoretical predictions with the experiment. Finally, we will provide brief review of the most important QED challenges at present.

Keywords

Lamb shift • QED effects in atoms and molecules

A.N. Artemyev
Institute of Physics, Kassel University, Kassel, Germany
e-mail: anton.artemyev@physik.uni-kassel.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_26

Introduction

It is generally accepted that the famous work of P. Dirac [1] has laid the foundation of the QED. In this work for the first time has been formulated the Dirac equation describing the relativistic electron. Instead of the nonrelativistic Schrödinger equation, with the Hamiltonian given by

$$H_{\rm Schr} = \frac{\boldsymbol{p}^2}{2m_e} + V_{\rm ext}(\boldsymbol{r})\,, \tag{1}$$

came the Lorenz-invariant Dirac equation with the Hamiltonian

$$H_{\rm D} = c\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta m_e c^2 + V_{\rm ext}(\boldsymbol{r})\,. \tag{2}$$

Here m_e is the electron's mass, $V_{\rm ext}(\boldsymbol{r})$ is the external field (Coulomb field of the nucleus for the atomic electron), c is the speed of the light, $\boldsymbol{p}$ is the momentum operator, and $\boldsymbol{\alpha}$ and β are the Dirac matrices, defined using the 2×2 unitary matrix I and $\boldsymbol{\sigma}$ Pauli matrices by the formulas:

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{3}$$

$$\beta = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}. \tag{4}$$

Note that since the complete energy in relativistic theory includes the rest-energy $m_e c^2$, the one-to-one correspondence between Hamiltonians requires additional term $m_e c^2$ in Eq. (1) (or subtraction of this term from Eq. (2)). This term, of course, just shifts all the energy levels by the constant value without making any physical changes.

Let us discuss now the physical differences between the nonrelativistic and relativistic cases. First of all, it is immediately seen from Eqs. (2), (3) and (4) that the wave function of the relativistic electron has four components. P. Dirac has found that 4 is the minimal dimension, allowing to write Lorenz-invariant Hamiltonian for the particle with the spin $1/2\hbar$. Second, the Hamiltonian (2) without any potential or with the spherically symmetric potential $V_{\rm N}$ does not commute neither with the orbital angular momentum operator $\boldsymbol{L}$ nor with the spin operator $\boldsymbol{S}$ but only with total momentum operator $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{S}$. Physically it means that even in the spherically symmetric fields, the electron does not have any well definite values of angular (l) or spin (s) momenta. Only the value of the total angular momentum j conserves. In spherically symmetric external fields, the wave function of the state with well-defined total angular momentum j always consists of two components with opposite parities. One of these components has the orbital momentum $l = j - 1/2$ and another has $l = j + 1/2$. In the nonrelativistic limit, $c \to \infty$ remains only one of them. Therefore, the relativistic one-electron states are

classified as nl_j, where n is the principal quantum number, j is the total angular momentum, and l is the orbital momentum of this state in nonrelativistic limit. Of course, the four-component wave functions with different angular dependence make the relativistic calculations much more complicated and time consuming than the same calculations for the nonrelativistic particle.

However, besides the technical difficulties, the Dirac Hamiltonian brings also completely new physics in comparison with the nonrelativistic case. Let us consider the spectra of the eigenvalues of Schrödinger and Dirac Hamiltonians. It is well known that the spectrum of the former Hamiltonian is limited from below. There exists the lowest possible value of the energy, which the electron may have. However, the energies of the Dirac Hamiltonian are not bounded from below. In the case of the attractive nuclear potential, the energy spectrum contains, like the spectrum of Schrödinger's Hamiltonian (1), the discrete levels of bound states with the energy $0 < \varepsilon < m_e c^2$ (Strictly speaking, in extremely strong fields this condition may transform to $-m_e c^2 < \varepsilon < m_e c^2$. It will be discussed in details in section "Challenges".) and the positive-energy continuum spectrum $\varepsilon > m_e c^2$. However, it also has negative-energy continuum of the states with the energies $\varepsilon < -m_e c^2$. The existence of negative continuum immediately arises the following questions:

- What is the physical meaning of the negative-energy states?
- How can the positive-energy states exist? Why do not they decay into negative continuum with the simultaneous release of energy?

The answers to these questions gives the "*Dirac sea*" formalism [2, 3]. According to this formalism, all the negative-continuum states are occupied by electrons. So spontaneous decay of the positive-energy state into negative continuum is forbidden by the Pauli principle. However, one can provide to the electron from negative continuum enough energy (usually $> m_e c^2$) to excite it into positive-energy state. The remaining in the negative-continuum vacancy behaves as the positively charged particle and was named positron. Sooner or later, this vacancy can be occupied by a positive-energy particle. The vacancy (positron) and the particle itself (electron) in this case disappear with release of energy. This process was called annihilation. In total the discovery of the positron brought to modern physics the concept of antimatter. Very soon this concept was successfully confirmed by the experimental observation of positrons [4].

Since that time QED is one of the most successful physical theories. Up to now no serious discrepancies between the QED predictions and the experimental results have been found. Based on the first principles of our knowledge about space and time, QED was chosen as a template for the construction of the theories describing other types of interactions. In this chapter we are going to provide the reader with some examples of successful applications of QED to the explanation of the effects in atoms and molecules. We will also point out the problems that are not yet solved.

The chapter is organized as follows. In the next section, we are going to write about Lamb's experiment and its explanation, which stimulated the appearance of QED in its modern form. In section "QED Effects" we provide several examples of highly accurate QED calculations and their comparison with experiments. We will show there the importance of QED effects in modern physics. Finally, in section "Challenges" we will discuss several problems, not yet solved by theory.

The relativistic units ($\hbar = m_e = c = 1$) are used throughout the chapter if not stated otherwise. Sometimes we will leave $\hbar$, m_e, and c to make the formulas more evident, like we did it in Eqs. (1) and (2). For one-electron atomic states, we use the notation nl_j, where n, l, and j denote the principal quantum number, orbital angular momentum (in nonrelativistic limit), and total angular momentum, respectively.

Lamb Shift

In this section we will discuss the discovery, which forced the development of QED and led to its formulation in modern form. This discovery has been done by W. Lamb and R. Retherford and was named after W. Lamb as the "*Lamb shift*."

History

Let us start from history. In addition to the existence of antimatter, the Dirac equation

$$H_{\mathrm{D}}\Psi = \varepsilon\Psi \tag{5}$$

with the Hamiltonian (2) and Coulomb field of the nucleus

$$V_{\mathrm{N}}(r) = -\frac{\alpha Z}{r}\,, \tag{6}$$

where $\alpha \approx 1/137$ and Z is the charge of the nucleus, could explain the difference between the energy levels with the same principal quantum number n and different values of the total angular momentum j (so-called fine splitting). Equation (5) with the Coulomb potential (6) can be solved analytically and yields the following formula for the energy levels:

$$\varepsilon_{nj} = \left(1 + \frac{(\alpha Z)^2}{\left[\sqrt{(j+1/2)^2 - (\alpha Z)^2} + n - j - 1/2\right]^2}\right)^{-1/2} m_e c^2. \tag{7}$$

One can see that the energy of the levels with the same values of n and j, for example, $2s_{1/2}$ and $2p_{1/2}$, is exactly the same. However, unlike the nonrelativistic

case, where the energy depends only on the principal quantum number, the relativistic electron energy is different for the levels with the same n but different j, for example, $2p_{1/2}$ and $2p_{3/2}$. The value of this splitting perfectly coincided with the experimental measurements known at that time. So approximately during 15 years after its derivation, the Dirac equation explained all experimental results and there was no need for changes in the existing theory.

Of course, it does not mean that the theory stopped to develop. Already in that time, the theoreticians tried to answer the question: is there any physics beyond the Dirac equation? So, for example, they tried to describe how the electron interacts with its own field. However, the infinite result of these calculations convinced them that they were on the wrong way. And there were no experimental results which could clearly confirm the existence of any physics beyond the Dirac theory. Although, as we will see later, some of the experiments clearly pointed to this unknown effects, they were not considered seriously.

WWII paused the development of QED but forced the progress, in particular, in the development of the microwave technique, which was used for the radiolocation. It made possible later the experiments which made a breakthrough in physics. In particular, in 1947 W. Lamb and R. Retherford published the results of their experiment with the hydrogen atoms, which strongly influenced the development of QED and finally led to the formulation of QED in its modern form [5]. The initial idea was to measure the shift between the $2s_{1/2}$ and $2p_{1/2}$ levels in magnetic field. As we have just discussed before, the Dirac theory predicts that the energy of these two levels in pure Coulomb field of the nucleus must be the same. The former level is metastable, with the lifetime $\tau \approx 1/7\,\mathrm{s}$. The latter is short living, with $\tau \sim 10^{-9}\,\mathrm{s}$. So if one operates with a beam of excited atoms, one can be sure that initially all atoms are in $2s_{1/2}$ states. In their experiment Lamb and Retherford directed the beam of the excited hydrogen atoms through the region with adjustable microwave radiation and magnetic field to the detector, which could count only the excited atoms. The microwave radiation could excite the atom from metastable $2s_{1/2}$ to $2p_{1/2}$ state. So when the frequency of the microwave radiation is equal to the energy difference between levels, all the atoms should be first excited in $2p_{1/2}$ state and then decay before reaching the detector.

Unexpectedly, Lamb and Retherford found that the energy difference between $2s_{1/2}$ and $2p_{1/2}$ states does not vanish even without magnetic field. In the absence of the magnetic field, this difference was equal approximately 1,000 MHz. Later this difference was called the Lamb shift. Just to demonstrate how tiny this effect is, let us compare it with the binding energy of the states. As we know, the binding energy of the hydrogen levels can be approximately expressed in atomic units as $1/(2n^2) = 0.125$ a.u. Transferring 1,000 MHz to atomic units, we obtain $1{,}000\,\mathrm{MHz} \approx 1.6 \times 10^{-7}$ a.u., i.e., six orders of magnitude smaller. The smallness of the effect explains why it was not found before, although some experiments pointed to it. In his Nobel Lecture in 1955 [6], Lamb noticed that the experiments, performed almost ten years before his and Retherford's one by W. Houston [7] and R. Williams [8], "*indicated a discrepancy* (between theory and experiment ANA) *which should have been taken seriously*."

After the publication of Lamb's results, it became definitively clear that the theory needed further development. The Lamb shift became the main topic of the Shelter Island Conference in 1947. Most theoreticians came to the conclusion that the origin of the Lamb shift was in the interaction of the electron with its own field, known also as "*self-energy*." However, the theory provided for this correction an infinite result. It was Bethe who first demonstrated how one should perform the renormalization procedure and what the physical meaning of this procedure is. In his work [9] he found a simple and clear way to perform the mass renormalization. The detailed procedure of renormalization is described in the corresponding chapter of this book. So here we will just emphasize its main features.

Using the nonrelativistic theory of radiation, Bethe has written the formula for the energy shift of the electron due to emission and reabsorption of the photon (self-energy). Then he noticed that the energy shift already contains the change of the electron kinetic energy "*due to the fact that electromagnetic mass is added to the mass of the electron. This electromagnetic mass is already contained in the experimental electron mass; the contribution (3) (corresponding to the change of the mass ANA) to the energy should therefore be disregarded*" [9]. In this phrase Bethe formulated the main physical idea of the renormalization procedure. The physical constants, like the electron mass or charge, may contain in their theoretical description infinite integrals. However, the theory must use the "pure" values of these constants, i.e., subtract the contribution of the infinities, already taken into account. In particular, in his work [9] Bethe subtracted from the expression for the energy shift due to self-energy the expression corresponding to the infinite change of the electromagnetic mass of the electron. Due to this subtraction, the initially linearly divergent (with respect to the upper limit) integral $\sim \int_0^K dk$ converted to the logarithmically divergent one, $\sim \int_0^K dk/k$. Then Bethe supposed that the upper limit K must be large enough and that the fully relativistic treatment will provide convergent result. He has chosen the upper limit $K = m_e c^2$ as the most natural for the nonrelativistic case and obtained 1,040 MHz for the Lamb shift between $2s_{1/2}$ and $2p_{1/2}$ levels. One year later T. Welton [10] gave a qualitative description of the electron self-energy, explaining it by the fluctuations of the field, felt by the electron, and arrived to the same result as Bethe for the nonrelativistic case.

After publication of Bethe's results, many people were involved in relativistic self-energy calculations. However, due to the large number of ambiguities in the procedure, their calculations led to different results. So in the end of 1947, V. Weisskopf wrote to J. Oppenheimer [11] "*I am somewhat doubtful that this problem can be solved, and I think it is a limit in principle of the theory*." However, in 1949 this work has been done. First N. Kroll and W. Lamb [12] found the fully relativistic Lamb shift to be equal to 1,052 MHz, which was very close to Bethe's 1,040 MHz. Then S. Tomonaga et al. [13] calculated all the effects of the same order of magnitude and obtained the formula for the Lamb shift of the level with principal quantum number n, orbital momentum l, and total angular momentum j:

$$\Delta E_{njl} = \frac{4\alpha(\alpha Z)^4}{3\pi n^3}\left[L_{nl} + \frac{19}{30}\delta_{l0} + \frac{3}{8}\frac{j(j+1) - l(l+1) - 3/4}{l(l+1)(2l+1)}(1-\delta_{l0})\right] m_e c^2. \tag{8}$$

Here L_{nl} are integrals of the wave function with the proper operators. For example, for the hydrogen $L_{20} = 7.03$ and $L_{21} = 0.03$, which yields for the Lamb shift

$$\Delta E_{2s_{1/2}} - \Delta E_{2p_{1/2}} = 1{,}051 \text{ MHz}\,. \tag{9}$$

In this result is also included the shift of the levels due to the deformation of the Coulomb field of the nucleus by the perturbed electron-positron field, the so-called *vacuum polarization* correction. This correction in the lowest order in (αZ) was first evaluated by E. Uehling in [14] within the frameworks of the Dirac theory. The nature of this contribution can be easily understood, if one considers the charge density induced by the solutions of the Dirac equation:

$$\rho(\boldsymbol{r}) = \frac{e}{2}\left[\sum_{\varepsilon_n<0}|\psi_n(\boldsymbol{r})|^2 - \sum_{\varepsilon_n>0}|\psi_n(\boldsymbol{r})|^2\right]. \tag{10}$$

Here ψ_n are the eigenfunctions of the Dirac Hamiltonian with the energies ε_n. In the absence of the external field, the solutions of Dirac equation are exactly symmetric. One can see it from the fact that the change of the energy sign in the Dirac equation for the free electron

$$(\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta)\psi_n = \varepsilon_n\psi_n \tag{11}$$

is equivalent to the simultaneous sign change of the momentum and interchange of two upper components with two lower. So the probability density, generated by the electron and the positron solutions is exactly the same, and the sums in Eq. (10) cancel each other. In the presence of the external field, this symmetry disappears. The positive-energy (electron) states are attracted to the nucleus, and the discrete spectrum appears. On the contrary, the negative-energy states (positrons) are repelled from the nucleus. It leads to an induced charge density, which generates the shift of the energy levels from the predictions of the Dirac theory. Usually, this effect is smaller than the self-energy correction. So in the case of hydrogen, it makes a 2.3 % contribution to the total Lamb shift. However, with the increase of the nuclear charge, the vacuum polarization correction grows faster than the self-energy one and becomes comparable with the latter for a nuclear charge close to $Z = 100$.

The understanding and explanation of the Lamb shift was one of the greatest triumphs of QED. The further intensive development of the theory led to the final formulation of QED in its modern form to the end of the 50th of the twentieth century. In the next section, we will provide some examples of successful application of QED predictions. However we would like to make a first brief description of QED effects from the point of view of modern theory.

Modern Explanation

Modern QED for bound particles is a perturbative theory, based on the Wick's expansion of the S-matrix or Green function (depending on the approach used) in powers of the fine-structure constant $\alpha \approx 1/137$. Each term of the perturbative expansion in α can be conveniently depicted by the corresponding Feynman diagram [15]. The exact rules, allowing to write the calculation formula for the given diagram, may depend on the approach used for the derivation. However, as a general rule, each diagram consists of external and internal lines. External lines denote the real incoming or outgoing particle, described by wave functions. Internal lines correspond to virtual particles, described by propagators. The four momentums at each vertex of the diagram must be conserved. Each vertex implies four-dimensional time-space (or energy-momentum) integration with the proper factors. Each closed loop implies the integration over the energy circulating inside the loop. In general, the Feynman diagram technique allows to visualize the complicated processes and essentially simplifies the calculation formula derivation.

In what follows, we will use straight double lines for the electrons moving in the external field of the nucleus and wavy lines for the photons. So the Lamb shift in hydrogen, discussed in the previous subsection, can be represented by the diagrams shown in Fig. 1. The diagram shown in Fig. 1a is known as the self-energy (SE) and the diagram (b) as vacuum polarization (VP).

Before proceeding further, let us define the main difference between the low-Z and middle- and high-Z atoms. In natural (a.k.a. relativistic) units, the Dirac Hamiltonian of the free electron does not contain any constant. It is just $H_D = \boldsymbol{\alpha}\cdot\boldsymbol{p}+\beta$. All terms are proportional to 1. The Coulomb potential of the nucleus enters here as $V_C = -\alpha Z/r$, i.e., multiplied by (αZ). For low-Z atoms this constant is small and serves allows to perform a additional perturbative expansion in powers of (αZ). Unlike low-Z case, for higher $Z \geq 20$, this parameter is no longer small, e.g., for uranium $\alpha Z \approx 0.7$ and all calculations must be performed without additional (αZ) expansion.

Historically, the Lamb shift was discovered and calculated first for the hydrogen atom. In that time only the order-by-order evaluation in powers of (αZ) was possible. Equation (8) represents the almost exact result for hydrogen but cannot provide accurate result for the case of hydrogen-like ion of, e.g., uranium ($Z = 92$). The all-order in (αZ) calculation of these diagrams became possible only with

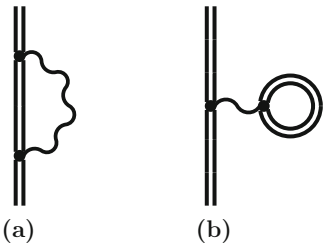


Fig. 1 Self-energy and vacuum polarization diagrams

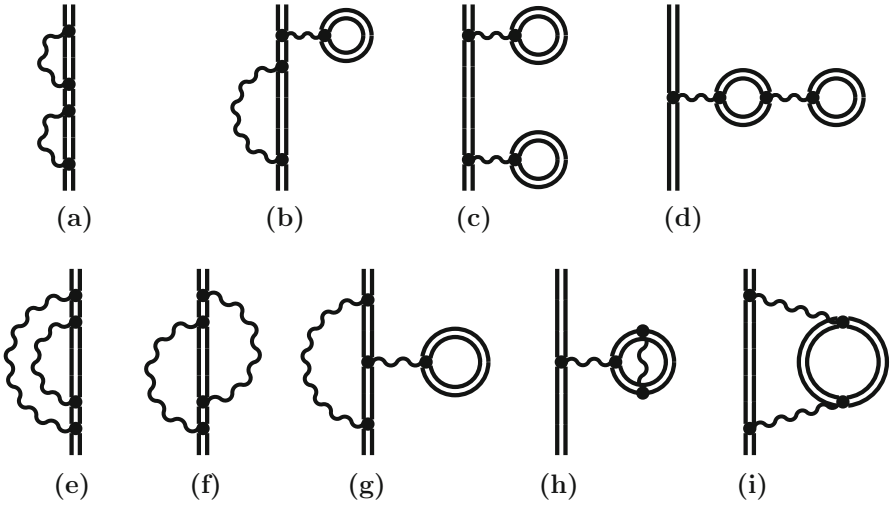


Fig. 2 Second-order one-electron diagrams

the development of computers. While the lowest order in (αZ) (Eq. 8) was known already in 1949, the first attempt of a complete evaluation of the SE correction for the high-Z case was performed by A. Desiderio and W. Johnson in [16]. The first precise evaluation of this correction was done by P. Mohr in [17,18]. The completion of (αZ) evaluation of the VP correction took another 15 years and was done by G. Soff and P. Mohr in 1988 [19] and independently by N. Manakov et al. in 1989 [20].

While the two diagrams shown in Fig. 1 almost completely define the Lamb shift in hydrogen, in the case of highly charged ions, other corrections are required. In the case of one-electron ions, in the second order of perturbation theory arise the diagrams shown in Fig. 2.

The calculation of this set of diagram to all orders in (αZ) is not finished yet. The contribution of the diagrams (h) and (i) is known only in the lowest order and define nowadays the accuracy of the theoretical prediction for the low-lying energy levels of high-Z atoms and ions. The recent progress in this field has been done by V. Yerokhin and V. Shabaev [21]. The authors evaluated the entire set of two-loop self-energy diagrams (a), (e), and (f).

In the case of several electrons, besides the one-electron diagrams already discussed shown in Figs. 1 and 2, diagrams, including two and more electrons become possible. The first and second orders of these diagrams are shown in Fig. 3. In the first order, in α there exists only one diagram of one-photon exchange. This diagram describes the exchange by the single virtual photon between two electrons and is shown in Fig. 3a. In the chapter devoted to the effective QED Hamiltonian, it is shown that the Dirac-Coulomb-Breit Hamiltonian, frequently used in many-electron calculations, can be derived from the first principles of QED.

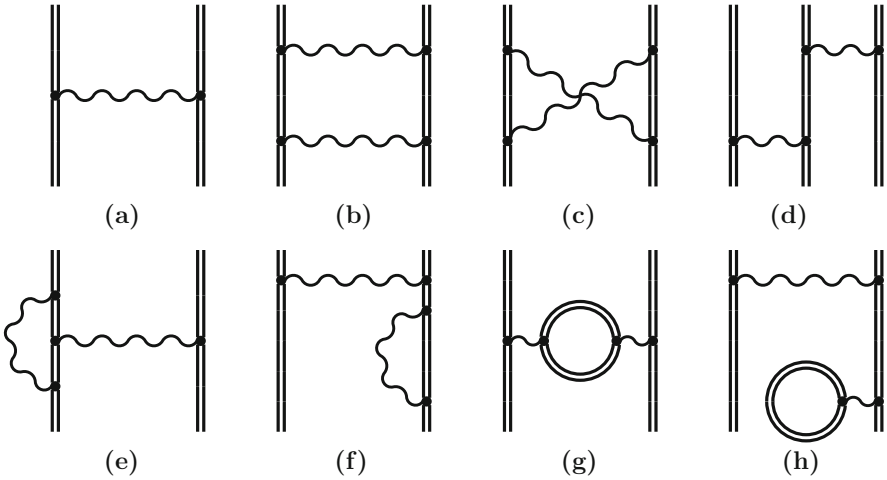


Fig. 3 Two- and three-electron diagrams

To this end one should consider the one-electron Dirac Hamiltonian (2) and the one-photon exchange diagram, taken in the Coulomb gauge and evaluated in the Breit approximation, i.e., neglecting the energy dependence of the photon propagator (retardation effect) and the contribution of the negative-energy continuum. For the details see the chapter ▸ Chap. 7, "Effective QED Hamiltonian" and the original work by V. Shabaev [22].

In the next order of the perturbation theory arise the diagrams describing the exchange by two virtual photons (diagrams Fig. 3b, c, and for more than 2 electrons (d)) as well as the screened SE ((e) and (f)) and VP ((g) and (h)). While the diagrams (b)–(d) within the Breit approximation are automatically included in the many-electron calculations using the Dirac-Coulomb-Breit Hamiltonian, the contributions of the negative-energy continuum as well as the retardation effect to these diagrams can be evaluated only within the rigorous QED approach. For example, the contribution of the diagram shown in Fig. 3c vanishes within the Breit approximation. It becomes nonzero only when the retardation effect and the negative continuum are taken into account. Of course, the screened QED corrections (e)–(h) can be evaluated only within the rigorous QED approach.

Nowadays the evaluation of the diagrams shown in Fig. 3 is completed for all low-lying states of He- and Li-like ions, ground state of Be-like ions, and several states of B-like ions in wide range of nuclear charge $10 \leq Z \leq 100$ [23–31]. However, the larger is the electron number, the larger number of diagrams arises and the higher order of perturbation theory is required in order to reach high precision. So the application of the rigorous QED approach is restricted by the number of electrons. Otherwise it is necessary to combine the advantages of many-electron approach with those of QED approach in order to reach high precision of theoretical predictions. See more in the ▸ Chap. 7, "Effective QED Hamiltonians".

To finish the section devoted to the Lamb shift, we would like to mention that historically Lamb shift was discovered for the atoms of hydrogen, where it is almost completely defined by the contribution of one-loop SE diagram, shown in Fig. 1a. With the development of the theory, it became clear that there are other effects contributing to the Lamb shift even in hydrogen. So now it is common to call as "*Lamb shift*" the difference between the exact energy of the level and the eigenvalue of the Dirac Hamiltonian (2), calculated using Eq. (7) for the one-electron case, or the Dirac-Coulomb-Breit Hamiltonian for the many-electron case. Besides already discussed QED contributions, Lamb shift contains also the contribution due to the nuclear recoil effect, nuclear polarization effect, and the deviation of the nuclear field from the pure Coulomb one because of the finite nuclear size. In the next section, we will provide some examples of the Lamb shift calculations and their comparison with the experiment.

QED Effects

Energy Levels

In this section we are going to provide several examples of successful application of QED theory. We start these examples from the QED evaluation of the energy levels of highly charged ions (HCI). As we have already discussed in the previous section (see Eq. 8), with the increase of the nuclear charge, the QED effects are growing as $(\alpha Z)^4$, while the binding energy just as $(\alpha Z)^2$. Hence the larger the nuclear charge Z is, the more pronounced the QED effects should be.

The high-precision experiments with HCI became possible in the last decades of the twentieth century. These measurements gave next impact to the theory development. In particular, new methods of all-order in (αZ) were created. The comparison of the theoretical predictions with the experiment provided new tests of QED in the region of extremely strong electric fields.

In the previous section, we have already discussed the Lamb shift in the hydrogen atom. Let us now consider the Lamb shift of the hydrogen-like ion of uranium. The various contributions to this Lamb shift are presented in the Table 1. The numbers in the parentheses denote the uncertainty of the last digits. For the comparison, we remind here that the ground state binding energy for the point-like nucleus can be evaluated by the simplified formula (7)

$$\varepsilon_{1s} = m_e c^2 \sqrt{1 - (\alpha Z)^2}\,, \tag{12}$$

$$\varepsilon_{1s}^{\text{binding}} = m_e c^2 - \varepsilon_{1s}\,, \tag{13}$$

and is equal 132,279.92(1) eV for uranium, with the uncertainty ± 0.01 eV due to the value of the fine-structure constant α.

From Table 1 one can see that the Lamb shift of the ground state in hydrogen-like uranium makes $\approx$ 0.35 % of the binding energy (compared with 0.0001 %

Table 1 Individual contributions to the Lamb shift in hydrogen-like uranium (in eV)

Contribution	Value	Reference
Finite nuclear size	198.54(19)	[32]
First-order QED	266.45	[33]
Second-order QED	1.26(33)	[34]
Nuclear recoil	0.46	[35]
Nuclear polarization	0.20(10)	[36, 37]
Total lamb shift theory	463.99(39)	
Experiment	460.2(4.6)	[38]

Table 2 Individual contributions to the Lamb shift of the $2p_{1/2} \to 2s_{1/2}$ transition energy in Li-like uranium (in eV)

Contribution	Value	Reference
Finite nuclear size	−33.30(3)	[32]
One-photon exchange	368.83	[32]
One-electron first-order QED	−42.93	[33]
Two-photon exchange within Breit approximation	−13.54	[27]
Two-photon exchange beyond Breit approximation	0.17	[27]
Screened QED	1.16	[25, 26]
One-electron second-order QED	0.22(6)	[39]
Three and more photon effects	0.14(7)	[32]
Nuclear recoil	−0.07	[35]
Nuclear polarization	0.03(1)	[36, 37]
Total Lamb shift theory	280.71(10)	
Experiment	280.645(15)	[40]
Experiment	280.59(10)	[41]
Experiment	280.52(10)	[42]

for hydrogen). One can also notice that the main uncertainty of the theoretical prediction arises from uncalculated contributions of the second order in α shown in Fig. 2. The calculation of each second-order diagram can be considered as a separate very complicated problem and is, of course, one of the challenges to the modern QED theory. To the knowledge of the author, there has not yet been made any attempt to complete the calculations of the second-order contribution. On the other hand, these calculations are not so urgently needed at the moment, since the experimental precision is one order of magnitude worse than the theoretical one.

The situation with the experimental accuracy changes, if one considers the $2p_{1/2} \to 2s_{1/2}$ or $2p_{3/2} \to 2s_{1/2}$ transitions in Li-like ions. In this case the experimental accuracy reaches 0.015 eV and provides a possibility for very accurate tests of QED. As an example we provide here the table for the individual contributions to the $2p_{1/2} \to 2s_{1/2}$ transition in Li-like uranium.

As one can see from Table 2, the main theoretical uncertainty in this case arises from two sources. Besides the uncalculated contributions of one-electron second-order diagrams, one should also consider the three and more photon

effects. These effects can be taken into account nowadays only within the Breit approximation. As a result the total theoretical uncertainty in the case of Li-like uranium is almost one order of magnitude higher than the experimental one. In total the comparison of theoretical predictions with experimental results allows to test QED on a 0.2 % level to the first order in α and on a 6.5 % level to the second order in α. We note here again that only the highly charged nuclei allow to test QED in non-perturbative regime in (αZ).

g-Factor

In the previous subsection, we have discussed the QED corrections to the energy levels of the few-electron ions. The evaluation of these corrections and their comparison with the experiment allow to test QED theory with very high accuracy. In this subsection we provide an example of the successful use of QED predictions for the accurate determination of the electron rest mass using precise measurements of the bound electron g-factor.

By definition, g-factor connects the magnetic moment operator of an electron $\boldsymbol{\mu}$ with its total angular momentum $\boldsymbol{J}$:

$$g = -\frac{\langle j_a m_a | \mu_z | j_a m_a \rangle}{\mu_B m_a}\,, \tag{14}$$

where $\mu_B = \frac{e\hbar}{2m_e}$ is the Bohr magneton, μ_z is the $z-$projection of the electron magnetic moment operator $\boldsymbol{\mu}$, and j_a and m_a denote total electron angular moment and its projection. The Dirac theory gives for the g-factor the value

$$g = \frac{2}{3}\left(1 + \frac{2\varepsilon_a}{m_e}\right)\,, \tag{15}$$

where ε_a is the Dirac energy of the electron (7). Other corrections should be calculated by means of QED. Theoretically various contributions to the g-factor can be calculated using their relations to the corresponding energy shifts, caused by the interaction with the homogeneous magnetic field $\boldsymbol{B}$ directed along the $z-$axis:

$$\Delta E = \Delta g \mu_0 B m_a\,. \tag{16}$$

The typical diagrams, which have to be calculated for the accurate evaluation of the matrix element in Eq. (14), are shown in Fig. 4. The wavy line ending with the triangle designates there the interaction with the external magnetic field.

In 2000 the results of the measurements of the g-factor in H-like ion of carbon by Häffner et al. were published [43]. The experimentalists could store a single ion $^{12}C^{5+}$ in the trap in the magnetic field and keep it there almost infinitely long. By measuring the relation between the cyclotron ν_c and Larmor precession ν_L frequencies, they could deduce the value for the bound electron g-factor using the formula

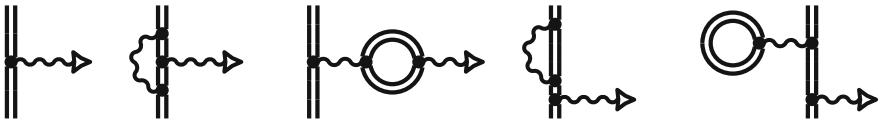


Fig. 4 Typical diagrams needed for the accurate evaluation of the g-factor

Table 3 Individual contributions to the g-factor in $^{12}C^{5+}$ and $^{16}O^{7+}$

Contribution		$^{12}C^{5+}$	$^{16}O^{7+}$	Reference
Dirac value		1.99872135439(1)	1.99772600306(2)	
Finite nuclear size		0.00000000041	0.00000000155(1)	
One-loop QED	$(\alpha Z)^0$	0.00232281947(1)	0.00232281947(1)	
	$(\alpha Z)^2$	0.00000074216	0.00000131940	[47]
	$(\alpha Z)^4$	0.00000009342	0.00000024007	[52]
	h.o.,SE	0.00000000828	0.00000003443(1)	[51]
	h.o.,VP-EL	0.00000000056	0.00000000224	[48]
	h.o.,VP-ML	0.00000000004	0.00000000016	[54]
≥two-loop QED	$(\alpha Z)^0$	0.00000351510	0.00000351510	[56]
	$(\alpha Z)^2$	0.00000000112	0.00000000200	[47]
	$(\alpha Z)^4$	0.00000000006	0.00000000008	[55]
	h.o.	0.00000000000(3)	0.00000000000(11)	[55]
Recoil	m_e/M	0.00000008770	0.00000011707	[50]
	h.o.	0.00000000008	0.00000000010	[49]
Total theory		2.00104159018(3)	2.00004702032(11)	
Experiment		2.001041596(5)		[43]
Experiment			2.0000470254(15)(44)	[44]

$$g = 2\frac{\nu_L}{\nu_c}\frac{Q}{e}\frac{m_e}{M}\,, \tag{17}$$

where Q and M are the charge and the mass of the ion, respectively. In the experiment [43] the frequency relation has been measured with the relative accuracy better than 10^{-9}. Later the highly accurate measurements have been also performed for the hydrogen-like ions $^{16}O^{7+}$ [44] and $^{28}Si^{13+}$ [45].

Already after the first experiment, it became clear [46] that it provides not only the precise test of the magnetic sector of QED but also gives a wonderful possibility to improve the known accuracy of the electron mass. If the matrix element in Eq. (14) is evaluated with high enough accuracy, one can invert Eq. (17) in order to evaluate m_e instead of g, because m_e in this case has the worst relative accuracy.

The evaluation of various contributions to the matrix element in the Eq. (14) has been performed by many authors [47–54] and finished by K. Pachucki et al. [55]. The individual contributions to the g-factor of bound electron in the hydrogen-like ions of carbon and oxygen are shown in Table 3.

As follows from the table, the calculations were performed using additional perturbation expansion over the powers of (αZ) for the QED corrections, which is of course possible for the values of nuclear charge less than 10. The "h.o." in the table denotes "higher orders" (in (αZ) for QED and in m_e/M for the recoil corrections, respectively). The recoil contribution corresponds to the fact that the nucleus has the finite mass and is of course in the lowest order proportional to the mass relation m_e/M.

Comparing the theoretical and experimental results, one can deduce from Eq. (17) the most accurate values of the electron mass. So the experiments with carbon and oxygen provide the following values for the electron mass in atomic mass units (1/12 of the ^{12}C mass):

$$m_e^{C^{5+}} = 0.00054857990932(29)$$
$$m_e^{O^{7+}} = 0.00054857990960(41)\,.$$

Recently the experiment with the hydrogen-like carbon has been repeated and the accuracy was improved by an order of magnitude [57]. From this experiment even more precise value of the electron mass was deduced:

$$m_e^{C^{5+}} = 0.000548579909067(14)(9)(2)\,,$$

where the first two errors are the statistical and systematic uncertainties of the measurement and the third one represents the theoretical uncertainty.

The g-factor investigations may provide not only the ways for the electron mass determination. The measurements of the g-factor in the high-Z ions may serve as a tool for the accurate determination of the fine-structure constant α. This possibility is based on the fact that the relative accuracy of α determination is proportional to the corresponding accuracy of g-factor and inversely proportional to the squared nuclear charge Z:

$$\frac{\delta\alpha}{\alpha} \sim \frac{1}{(\alpha Z)^2}\frac{\delta g}{g}\,. \tag{18}$$

In turn the accurate determination of α may answer to one of the fundamental questions of modern physics: does the space or time variation of the fundamental constants really exist? Nowadays scientists are looking for the answer to this question mostly in astrophysical observations. For more information about the perspectives of the g-factor investigations, we refer to the work of V. Shabaev et al. [58] and references therein.

In this section we have provided several examples of successful QED effect evaluation and their applications in modern physics. We hope that the reader obtained an impression about the QED effects and their magnitude and importance for modern physics. In the next section, we will tell about the challenges faced by the modern QED theory.

Challenges

In this section we are going to discuss the challenges to the modern QED theory. It is convenient to separate the problems, which in principle can be explained within the existing approaches but require either enormous computational resources or long formula derivation, from the problem, requiring the creation of completely new approaches. Let us call the former group as the "technical challenges" and the latter as the "conceptual challenges" and briefly discuss both groups. Of course, these two groups are closely mixed, so it is sometimes difficult to declare the exact border.

For the "technical challenges," let us ascribe first of all the calculations of the next orders of QED corrections, especially for the experiments, where they are really needed already now. As an example let us consider Table 2. From this table it is clear that in order to make accurate comparison of the theoretical prediction with the best experimental results, one has to complete the evaluation of the two-loop one-electron diagrams (see Fig. 2) to all orders in (αZ) and to evaluate at least the three-photon exchange diagrams within the rigorous framework of QED. These calculations will reduce the uncertainty of the theoretical result at least by one order of magnitude and make it comparable with the experimental precision.

The need for next-order correction calculations is even much stronger pronounced for the case of dipole-forbidden $2p_{3/2} \to 2p_{1/2}$ transition in Be- and B-like ions of argon. The experimental precision of the determination of these energies is [59, 60] 5×10^{-7} and 4×10^{-7} for Be- and B-like ions, respectively. The best to date QED calculations [61] provide the relative accuracy 1.5×10^{-4} for the B-like argon. These experiments potentially may allow to test the third and fourth orders of the perturbation theory in α. However, as it has been already several times pointed out here, even the calculations of the second order in α corrections are not yet finished.

Another challenge, which can also be treated as rather technical, although partially conceptual, consists in the development of the methods for QED calculations in many-electron systems, like heavy neutral atoms. As one can notice, all the successful calculations, which we have already discussed, have been performed for the systems with the number of the electrons $N_e \leq 5$. Of course, theoretically one can try to apply the existing approaches to the many-electron systems. However, in this case the number of diagrams grows drastically, which makes almost impossible the application of the existing methods to the systems with number of electrons greater than 10. The possible solutions of this problem are discussed in the ▸ Chap. 7, "Effective QED Hamiltonians" of this book. The effective QED Hamiltonians, taking into account the QED effects in the lowest orders, are already constructed (see, e.g., Ref. [62]). Now the challenge is the improvement of their accuracy, inclusion of next orders of QED, and enhancement of the case of several Coulomb centers (molecules). The latter task looks the most difficult, because up to now the QED calculations in molecules have been performed only in the simplest case of H_2 and D_2 and HD molecules using (αZ) expansion (see, e.g., [63]). There is still no technique for the molecular QED calculations in all orders in (αZ).

Extremely Strong Fields

To the author's opinion, the problems, discussed above, will be solved or at least can be solved within the next decade. Now we would like to discuss one challenge, which is rather conceptual – the QED in the extremely strong fields. Let us start from formula (7) for the energy levels of the electron in the field of the point-like nucleus with the charge Z. It is evident that this formula can be applied for all values of nuclear charge $Z \leq \frac{j+1/2}{\alpha}$. So for the states with total angular momentum $j = 1/2$, the formula is limited by $Z = 137$. For larger Z the expression under the square root in the denominator becomes negative and the energy gets the imaginary part. This problem can be overcome, if one uses the finite (not point-like) nucleus charge distribution. In this case the energy stays real, but at some critical value of the charge $Z_{\rm cr}$, the energy of the ground state reaches the negative-energy continuum and dives into it.

W. Pieper and W. Greiner [64] were the first to perform calculations with the simple model of the nuclear charge distribution. They found that the critical value of the charge is $Z = 173$. Around this value of the charge, the energy of the ground state reaches the value $-m_e c^2$, and further increment of the charge leads to the diving of this state into the negative-energy continuum. As we have already discussed above, all negative-continuum states in the "Dirac sea" are occupied. If we consider the ion or atom with the vacancy in the state, which sank in the negative-energy continuum, this vacancy can be occupied by an electron from the "Dirac sea." Physically it means the spontaneous pair creation, or in the other words instability of the electron-positron vacuum.

Unfortunately, the possibility for the experimental creation of stable critical field sources does not exist. The heaviest nuclei available for the synthesis now and in the nearest future have the charge far away from the critical one. The only way to create critical fields for a short time is the collision of two ions with the total charge exceeding the critical one. For example, two uranium nuclei have the total charge $Z_{U_2} = 184 > Z_{\rm cr}$. The energy of the collision should be chosen in such a way that the ions come close to each other but do not collide. It is said that for a short time the two colliding nuclei create the quasi-molecule. The energy levels of this quasi-molecule can reach the negative continuum and dive into it.

First estimations of the energy levels of two colliding ions were performed by G. Soff et al. [65]. They have found that at the critical internuclear distance $D_{\rm cr} = 35\,\text{fm}$, the ground state of the quasi-molecule U_2^{183+} reaches the border of the negative continuum. The calculations were performed within the monopole approximation, i.e., the total potential, created by two nuclei, was restricted to its spherically symmetric part. Later this value was improved by two independent calculations [66, 67]. The calculations with the full two-center potential yield for the critical internuclear distance the value $D_{\rm cr} = 34.72\,\text{fm}$.

After the publication of the work by Pieper and Greiner [64], the idea to observe the spontaneous vacuum decay became very popular. Plenty of investigations, both theoretical and experimental, were devoted to this problem. See Ref. [68]

and references therein for the review. Unfortunately, the experimental facilities nowadays allow only the experiments by bombarding the foil consisting of the heavy metal by the beam of the ions. In this case the target atoms have all the electrons, and the exact solution of the collisional problem within the framework of the existing theories is impossible. The best available approximation in this case is the single active electron approximation, where the effect of all but one electron is replaced by the effective screening potential. While this approximation works relatively good for the description of the K-shell vacancies production or the emitted electron spectrum, it fails if one tries to describe the positron production. At the future experimental Facility for Antiproton and Ion Research (FAIR) [69], which will be built soon in Darmstadt (Germany), new-generation experiments will become possible. The "dream system" consisting of colliding U^{92+} and U^{91+} ions with total kinetic energy between 5 and 6 MeV/u will be available for the experiment. This system contains just one electron, so the inter-electronic interaction will not mask any effects or complicate the theoretical calculations. On the other hand, the presence of the electron will provide information about quasi-molecule formation (via the photon emission due to the excitations and de-excitations). Let us now consider what kind of problem may face the theory trying to describe this process.

Since the relative velocity of the nuclei is at least one order of magnitude smaller than the average speed of bound electron, we are going to use the Born-Oppenheimer approximation. Within this approximation the nuclei move along the classical trajectories and the electron moves in their field. To the lowest order the description of the process requires the solution of the time-dependent Dirac equation, with the two-center potential V_{2C} which is the sum of the potentials generated by both nuclei when they are in the positions $\boldsymbol{R}_{1,2}(t)$:

$$i\dot{\psi}(\boldsymbol{r};t) = (\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta + V_{2C}(\boldsymbol{r};\boldsymbol{R}_1(t),\boldsymbol{R}_2(t)))\psi(\boldsymbol{r};t)\,. \tag{19}$$

The effective algorithms for the solutions of this equation have been recently elaborated [66, 70]. Using these algorithms one can investigate the processes where only the electrons are involved, like excitation and de-excitation, ionization, or charge transfer. Moreover, in Ref. [71] the algorithm was applied for the calculation of the charge transfer probabilities and K-shell vacancies production in the collisions of the bare ions with the neutral target (in single active electron approximation) and the reasonable agreement with the available experimental data was achieved.

However, as soon as it is necessary to evaluate the pair creation probability in the super critical fields, the existing methods must be reconsidered. It is well known that unlike the nonrelativistic Schrödinger equation, the Dirac equation can describe only a single particle. If one uses the concept of the Dirac sea, where all negative-energy states are initially occupied, it is not clear how to solve in this case the Dirac equation and if it will provide the physical results. On the other hand, being the perturbation theory, the QED requires some zeroth-order approximation. The main question is what to select as this zeroth approximation if the Dirac equation does not fit to this aim.

Table 4 First-order QED corrections to the ground state energy of U_2^{183+} quasi-molecule as a function of the internuclear distance D. Preliminary results

D [fm]	E_b	ΔE_{SE}	ΔE_{VP}	ΔE_{QED}
40	1.9030	0.01341	−0.00941	0.00400
80	1.4968	0.00782	−0.00404	0.00378
100	1.3884	0.00654	−0.00307	0.00347
160	1.1839	0.00437	−0.00171	0.00266
200	1.0954	0.00358	−0.00128	0.00229
500	0.7715	0.00144	−0.00040	0.00104
1,000	0.5641	0.000701	−0.000178	0.000523

The next interesting question arises if one starts to consider the QED corrections in the critical fields. Taking into account the scaling of the QED corrections to the energy levels in leading order as $\alpha(\alpha Z)^4$ and the fact that the overall sign of this correction is positive, one may expect that this correction will grow very fast and prevent the ground state from diving into the negative continuum. However, very soon the first preliminary estimations of the QED effects for the hypothetical superheavy nucleus with the charge $Z = 170$ [19, 72–74] demonstrated that the QED corrections in this case represent less than 1 % of the total binding energy and hence cannot prevent the diving of the ground state into the negative continuum. Our new calculations [75] performed for the U_2^{183+} quasi-molecule demonstrate (see Table 4) that although the total QED correction grows very fast with the decreasing of the internuclear distance D, this contribution remains to be less than 1 % of the total binding energy E_b and hence can only reduce the critical internuclear distance but cannot prevent the diving into the negative continuum.

On the other hand, all the QED correction calculations are based on the assumption that the level (or the group of levels) is isolated. When the level reaches the border of the negative-energy continuum and moreover dives into it, there is no warranty that the existing approaches provide reasonable results.

In total the QED of the supercritical fields states a lot of problems, of both technical and conceptual character, which should be solved within the next decade.

Summary

In summary, in this chapter we have provided a brief overview of the QED effects and their importance for modern physics. We have demonstrated various examples of the successful confirmation of the QED predictions by the experiment. Finally we have highlighted the challenges faced by the theory in modern time. We hope that the chapter will help the readers to become better informed about the modern state of QED.

Acknowledgements Stimulating discussions with Prof. V. M. Shabaev and Prof. P. Indelicato are gratefully acknowledged.

References

1. Dirac PAM (1928) The Quantum Theory of the Emission and Absorption of Radiation. Proc R Soc Lond Ser A 117:610
2. Dirac PAM (1929) The Quantum Theory of the Electron. Proc R Soc Lond Ser A 126:360
3. Dirac PAM (1931) A Theory of Electrons and Protons. Proc R Soc Lond Ser A 133:60
4. Anderson CD (1933) The Positive Electron. Phys Rev 43:491
5. Lamb WE Jr, Retherford RC (1947) Fine structure of the hydrogen atom by a microwave method. Phys Rev 72:241
6. Tiselius A (1964) Nobel lectures, physics 1942–1962. Elsevier, Amsterdam
7. Houston WV (1937) A New Method of Analysis of the Structure of H_α and D_α. Phys Rev 51:446
8. Williams RC (1938) The Fine Structures of H_α and D_α Under Varying Discharge Conditions. Phys Rev 54:558
9. Bethe HA (1947) The electromagnetic shift of energy levels. Phys Rev 72:339
10. Welton TA (1948) Some Observable Effects of the Quantum-Mechanical Fluctuations of the Electromagnetic Field. Phys Rev 74:1157
11. Schweber SS (1994) QED and the men who made it: Dyson, Feynman, Schwinger, and Tomonaga. Princeton series in physics. Princeton University Press, Princeton
12. Kroll NM, Lamb WE Jr (1949) On the self-energy of a bound electron. Phys Rev 75:388
13. Fukuda H, Miyamoto Y, Tomonaga S (1949) A Self-Consistent Subtraction Method in the Quantum Field Theory. Prog Theor Phys 4:47
14. Uehling EA (1935) Polarization Effects in the Positron Theory. Phys Rev 48:55
15. Feynman RP (1949) Space-time approach to quantum electrodynamics. Phys Rev 76:769
16. Desiderio AM, Johnson WR (1971) Lamb Shift and Binding Energies of K Electrons in Heavy Atoms. Phys Rev A 3:1267
17. Mohr PJ (1974) Self-Energy Radiative Corrections in Hydrogen-Like Systems. Ann Phys (New York) 88:26
18. Mohr PJ (1974) Numerical Evaluation of the $1S_{1/2}$-State Radiative Level Shift. Ann Phys (New York) 88:52
19. Soff G, Mohr P (1988) Vacuum polarization in a strong external field. Phys Rev A 38:5066
20. Manakov NL, Nekipelov AA, Fainshtein AG (1989) Vacuum polarization by a strong coulomb field and its contribution to the spectra of multiply-charged ions. Zh Eksp Teor Fiz 95:1167 [Sov Phys JETP 68:673 (1989)]
21. Yerokhin VA, Shabaev VM (2001) Two-loop self-energy correction in H-like ions. Phys Rev A 64:062507
22. Shabaev VM (1993) Schrödinger-like equation for the relativistic few-electron atom. J Phys B 26:4703
23. Lindgren I, Persson H, Salomonson S, Karasiev V, Labzowsky L, Mitrushenkov A, Tokman M (1993) Second-order QED corrections for few-electron heavy ions: reducible Breit-Coulomb correction and mixed self-energy–vacuum polarization correction. J Phys B 26:L503
24. Lindgren I, Persson H, Salomonson S, Sunnergren P (1995) QED Calculations on Two- and Three-Electron Ions. Physica Scripta T59:179
25. Artemyev AN, Beier T, Plunien G, Shabaev VM, Soff G, Yerokhin VA (1999) Vacuum polarization screening corrections to the energy levels of lithiumlike ions. Phys Rev A 60:45
26. Yerokhin VA, Artemyev AN, Beier T, Plunien G, Shabaev VM, Soff G (1999) Two-electron self-energy corrections to the $2p_{1/2} - 2s$ transition energy in Li-like ions. Phys Rev A 60:3522
27. Yerokhin VA, Artemyev AN, Shabaev VM, Sysak MM, Zherebtsov OM, Soff G (2000) Two-Photon Exchange Corrections to the $2p_{1/2} - 2s$ Transition Energy in Li-Like High-Z Ions. Phys Rev Lett 85:4699

28. Sapirstein J, Cheng KT (2001) Determination of the two-loop Lamb shift in lithiumlike bismuth. Phys Rev A 64:022502
29. Artemyev AN, Shabaev VM, Yerokhin VA, Plunien G, Soff G (2005) QED calculations of the $n = 1$ and $n = 2$ energy levels in He-like ions. Phys Rev A 71:062104
30. Artemyev AN, Shabaev VM, Tupitsyn II, Plunien G, Surzhykov A, Fritzsche S (2013) Ab initio calculations of the $2p_{3/2} - 2p_{1/2}$ fine-structure splitting in boronlike ions. Phys Rev A 88:032518
31. Malyshev AV, Volotka AV, Glazov DA, Tupitsyn II, Shabaev VM, Plunien G (2014) QED calculation of the ground-state energy of berylliumlike ions. Phys Rev A 90:062517
32. Kozhedub YS, Andreev OV, Shabaev VM, Tupitsyn II, Brandau C, Kozhuharov C, Plunien G, Stöhlker T (2008) Nuclear deformation effect on the binding energies in heavy ions. Phys Rev A 77:032501
33. Mohr PJ, Plunien G, Soff G (1998) QED corrections in heavy atoms. Phys Rep 293:227
34. Yerokhin VA, Indelicato P, Shabaev VM (2003) Evaluation of the two-loop self-energy correction to the ground state energy of H-like ions to all orders in $(Z\alpha)$. Eur Phys J D 25:203
35. Artemyev AN, Shabaev VM, Yerokhin VA (1995) Relativistic nuclear recoil corrections to the energy levels of hydrogenlike and high-Z lithiumlike atoms in all orders in αZ. Phys Rev A 52:1884
36. Plunien G, Soff G (1995) Nuclear polarization contribution to the Lamb shift in actinide nuclei. Phys Rev A 51:1119
37. Nefiodov AV, Labzowsky LN, Plunien G, Soff G (1996) Nuclear polarization effects in spectra of multicharged ions. Phys Lett A 222:227
38. Gumberidze A, Bosch F, Bräuning-Demian A, Hagmann S, Kühl T, Liesen D, Schuch R, Stöhlker T (2005) Atomic physics with highly-charged heavy ions at the GSI future facility: The scientific program of the SPARC collaboration. Nucl Instrum Methods Phys Res Sect B: Beam Interact Mater Atoms 233:28. Fast ion-atom collisions – proceedings of the eighth workshop on fast ion-atom collisions, eighth workshop on fast ion-atom collisions
39. Yerokhin VA, Indelicato P, Shabaev VM (2006) Nonperturbative Calculation of the Two-Loop Lamb Shift in Li-Like Ions. Phys Rev Lett 97:253004
40. Beiersdorfer P, Chen H, Thorn DB, Träbert E (2005) Measurement of the Two-Loop Lamb Shift in Lithiumlike U^{89+}. Phys Rev Lett 95:233003
41. Schweppe J, Belkacem A, Blumenfeld L, Claytor N, Feinberg B, Gould H, Kostroun VE, Levy L, Misawa S, Mowat JR, Prior MH (1991) Measurement of the Lamb Shift in Lithiumlike Uranium (U^{89+}). Phys Rev Lett 66:1434
42. Brandau C, Kozhuharov C, Müller A, Shi W, Schippers S, Bartsch T, Böhm S, Böhme C, Hoffknecht A, Knopp H, Grün N, Scheid W, Steih T, Bosch F, Franzke B, Mokler PH, Nolden F, Steck M, Stöhlker T, Stachura Z (2003) Precise Determination of the $2s_{1/2}$-$2p_{1/2}$ Splitting in Very Heavy Lithiumlike Ions Utilizing Dielectronic Recombination. Phys Rev Lett 91:073202
43. Häffner H, Beier T, Hermanspahn N, Kluge HJ, Quint W, Stahl S, Verdú J, Werth G (2000) High-Accuracy Measurement of the Magnetic Moment Anomaly of the Electron Bound in Hydrogenlike Carbon. Phys Rev Lett 85:5308
44. Verdú J, Djekić S, Stahl S, Valenzuela T, Vogel M, Werth G, Beier T, Kluge HJ, Quint W (2004) Electronic g Factor of Hydrogenlike Oxygen $^{16}O^{7+}$. Phys Rev Lett 92:093002
45. Sturm S, Wagner A, Schabinger B, Zatorski J, Harman Z, Quint W, Werth G, Keitel CH, Blaum K (2011) g Factor of Hydrogenlike $^{28}Si^{13+}$. Phys Rev Lett 107:023002
46. Beier T, Häffner H, Hermanspahn N, Karshenboim SG, Kluge HJ, Quint W, Stahl S, Verdú J, Werth G (2001) A new determination of the electron's massNuclear Mass Corrections to the Electron g Factor. Phys Rev Lett 88:011603
47. Grotch H (1970) Nuclear Mass Corrections to the Electron g Factor. Phys Rev Lett 24:39
48. Beier T (2000) The g_j factor of a bound electron and the hyperfine structure splitting in hydrogenlike ions. Phys Rep 339:79

49. Martynenko AP, Faustov RN (2001) The g Factors of Bound Particles in Quantum Electrodynamics. Zh Eksp Teor Fiz 120:539. [JETP 93(3):471–476 (2001)]
50. Shabaev VM, Yerokhin VA (2002) Recoil Correction to the Bound-Electron g Factor in H-Like Atoms to All Orders in αZ. Phys Rev Lett 88:091801
51. Yerokhin VA, Indelicato P, Shabaev VM (2002) Self-Energy Correction to the Bound-Electron g Factor in H-like Ions. Phys Rev Lett 89:143001
52. Pachucki K, Jentschura UD, Yerokhin VA (2004) Nonrelativistic QED Approach to the Bound-Electron g Factor. Phys Rev Lett 93:150401
53. Pachucki K, Jentschura UD, Yerokhin VA (2005) Erratum: Nonrelativistic QED Approach to the Bound-Electron g Factor Phys Rev Lett 94:229902
54. Lee RN, Milstein AI, Terekhov IS, Karshenboim SG (2005) Virtual light-by-light scattering and the g factor of a bound electron. Phys Rev A 71:052501
55. Pachucki K, Czarnecki A, Jentschura UD, Yerokhin VA (2005) Complete two-loop correction to the bound-electron g factor. Phys Rev A 72:022108
56. Mohr PJ, Taylor BN (2005) CODATA recommended values of the fundamental physical constants: 2002. Rev Mod Phys 77:1
57. Sturm S, Kohler F, Zatorski J, Wagner A, Harman Z, Werth G, Quint W, Keitel CH, Blaum K (2014) High-precision measurement of the atomic mass of the electron. Nature 506:467
58. Shabaev VM, Glazov DA, Oreshkina NS, Volotka AV, Plunien G, Kluge HJ, Quint W (2006) g-Factor of Heavy Ions: A New Access to the Fine Structure Constant. Phys Rev Lett 96:253002
59. Draganić I, López-Urrutia JRC, DuBois R, Fritzsche S, Shabaev VM, Orts RS, Tupitsyn II, Zou Y, Ullrich J (2003) High precision wavelength measurements of QED-sensitive forbidden transitions in highly charged argon ions. Phys Rev Lett 91:183001
60. Mäckel V, Klawitter R, Brenner G, López-Urrutia JRC, Ullrich J, Physica Scripta T156:014004 (2013) Laser spectroscopy of highly charged argon at the Heidelberg electron beam ion trap
61. Artemyev AN, Shabaev VM, Tupitsyn II, Plunien G, Yerokhin VA (2007) QED Calculation of the $2p_{3/2} - 2p_{1/2}$ Transition Energy in Boronlike Argon. Phys Rev Lett 98:173004
62. Shabaev VM, Tupitsyn II, Yerokhin VA (2013) Model operator approach to the Lamb shift calculations in relativistic many-electron atoms. Phys Rev A 88:012513
63. Korobov VI, Hilico L, Karr JP (2014) Theoretical transition frequencies beyond 0.1 ppb accuracy in H_2^+, HD^+, and antiprotonic helium. Phys Rev A 89:032511
64. Pieper W, Greiner W (1969) Interior electron shells in superheavy nuclei. Zeitschrift für Physik A Hadrons and Nuclei 218:327
65. Soff G, Müller B, Greiner W (1978) Spectroscopy of Electronic States in Superheavy Quasimolecules. Phys Rev Lett 40:540
66. Tupitsyn II, Kozhedub YS, Shabaev VM, Deyneka GB, Hagmann S, Kozhuharov C, Plunien G, Stöhlker T (2010) Relativistic calculations of the charge-transfer probabilities and cross sections for low-energy collisions of H-like ions with bare nuclei. Phys Rev A 82:042701
67. Artemyev AN, Surzhykov A, Indelicato P, Plunien G, Stöhlker T (2010) Finite basis set approach to the two-centre Dirac problem in Cassini coordinates. JPB 43:235207
68. Müller-Nehler U, Soff G (1994) Electron Excitations in Superheavy Quasimolecules. Phys Rep 246:101
69. Henning W (ed) (2001) FAIR conceptual design report: an international accelerator facility for beams of ions and antiprotons. GSI
70. McConnell SR, Artemyev AN, Mai M, Surzhykov A (2012) Solution of the two-center time-dependent Dirac equation in spherical coordinates: Application of the multipole expansion of the electron-nuclei interaction. Phys Rev A 86:052705
71. Tupitsyn II, Kozhedub YS, Shabaev VM, Bondarev AI, Deyneka GB, Maltsev IA, Hagmann S, Plunien G, Stöhlker T (2012) Relativistic calculations of the K-K charge transfer and K-vacancy production probabilities in low-energy ion-atom collisions. Phys Rev A 85:032712
72. Soff G, Müller B, Rafelski J (1974) Precise Values for Critical fields in Quantum Electrodynamics. Z Naturforsch A 29:1267

73. Soff G, Schlüter P, Müller B, Greiner W (1982) Self-Energy of Electrons in Critical Fields. Phys Rev Lett 48:1465
74. Persson H, Lindgren I, Salomonson S, Sunnergren P (1993) Accurate vacuum polarization contributions. Phys Rev A 48:2772
75. Artemyev AN, Surzhykov A (2015) Quantum electrodynamical corrections to energy levels of diatomic quasimolecules. Phys Rev Lett 114:243004

Effective QED Hamiltonians

7

Anton N. Artemyev

Contents

Introduction 268
Dirac-Coulomb-Breit Equation and QED Corrections 270
First-Order QED Corrections 272
Vacuum Polarization Contribution 273
Self-Energy Contribution 275
Systematic Approach 278
Beyond Effective QED Hamiltonian 282
Summary 284
References 284

Abstract

An effective QED Hamiltonian on the one hand should account for the QED corrections and on the other hand should be constructed in such a way that many-electron self-consistent calculations with this Hamiltonian remain to be efficient. In this chapter various approaches to the construction of the effective QED Hamiltonian are discussed. Finally, the scheme for the construction of the Hamiltonian which takes into account the lowest- order QED corrections is provided.

Keywords

QED effects in atoms and molecules • Effective Hamiltonian

A.N. Artemyev (✉)
Institute of Physics, Kassel University, Kassel, Germany
e-mail: anton.artemyev@physik.uni-kassel.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_27

Introduction

The constantly increasing power of computers allows one to perform the more and more accurate calculations of many-electron systems. Traditionally, the relativistic calculations in many-electron systems are based on the self-consistent solution of the Dirac-Coulomb-Breit (DCB) equation. Modern computers allow to solve this equation with accuracy, comparable with (or better than) the quantum electrodynamic (QED) corrections' contribution. Moreover, modern experiments often require the exact knowledge of QED corrections. Just as an example of such experiments, one can consider the measurements of parity nonconservation in neutral cesium [1, 2]. Only accurate calculation of QED contributions could reduce the disagreement between the theoretical predictions and the experimental results from 2.5σ to 1.1σ, where σ is the standard deviation.

Several methods are available to perform the fast, accurate, and effective solution of the DCB equation. Among them one can distinguish the multi-configurational Dirac-Fock (MCDF) method, the relativistic configuration-interaction (CI) method, and the relativistic many-body perturbation (RMBPT) theory [3–13]. Using these methods on modern computers, one can solve the DCB equation very precisely. However, as soon as the accuracy on the level of QED corrections is reached, there is no sense in its further improvement, before the QED corrections are taken into account.

In turn, the QED of bound electrons is a perturbative theory. The ab initio calculation of the QED corrections represents a separate very complicated computational problem. The calculations of the QED corrections to the energy levels of the bound electrons are nowadays completed only up to the second order of the perturbation theory in the fine-structure constant α. So the accurate ab initio calculations of many-electron systems, where the accounting for the electron-electron correlations requires much higher orders of the perturbation theory, within the rigorous QED approach become impossible.

Concluding from mentioned above, in order to perform high-precision calculations in many-electron systems, one needs to invent the way to combine the methods for the solution of the DCB equation, allowing to account for the most part of the electron-electron correlations, with the rigorous QED approach, which allows to evaluate the QED effects. Theoretically, the simplest way to combine these two methods was found in Ref. [14], where the DCB Hamiltonian has been derived within the rigorous QED approach and the additional term to the Hamiltonian, which takes into account the QED effects in the lowest order, has been found. However, due to the absence of the effective computational algorithms for all QED effects, the direct implementation of this method to the self-consistent procedure of the DCB equation solution is impossible. The evaluation of the QED corrections in this case will be necessary to perform at each step of self-consistent procedure. It will dramatically increase the computational time and make the calculations unrealizable.

Before discussing the effective QED Hamiltonian, we would like to mention another method which allows to combine the accurate solution of the DCB equation with the QED corrections. According to this approach, one should perform the many-electron calculations with the DCB Hamiltonian without any QED effects, representing the DCB Hamiltonian in the form

$$H_{\rm DCB} = H_0 + \lambda \Delta V \,, \tag{1}$$

where H_0 is the unperturbed Hamiltonian, which can be used in the rigorous QED calculations as well, the perturbation ΔV describes the interelectronic interaction, and λ is an arbitrary parameter. For $\lambda = 1$ the Hamiltonian (1) coincides with the exact one. In order to improve the convergence of the perturbation series, H_0 may contain some effective screening potential, partially accounting for the screening effects. In this case ΔV should contain the corresponding counterterm. The calculations should be performed for several values of the parameter λ, and the result of the calculations should be then represented as the series in powers of this parameter. After that the few first terms of this expansion should be replaced by the same terms, calculated within the rigorous QED approach. This method has been used in Refs. [15, 16] in order to evaluate the $2p_{3/2} \to 2p_{1/2}$ transition energy in five-electron ions. It allows to include the QED corrections order by order in rigorous way. However, the computational time needed to evaluate these corrections grows very fast with the increment of the electron's number. Moreover, the larger is the electron number, the higher orders of perturbation theory in ΔV become necessary to reach the same relative accuracy. It makes this method not applicable for the systems with large number of electrons.

For the case of the large electron number, one should better try to include into the initial DCB Hamiltonian the terms describing the QED corrections, at least approximately, and then to perform the self-consistent calculations. Unfortunately, it is not so easy to do due to the nonlocal nature of some of the QED corrections. There exist plenty of attempts to write some additional terms to the DCB Hamiltonian, which allow to take into account the QED effects and do not affect the efficiency of the self-consistent procedure. In this chapter we will discuss some of these approaches and describe the way for the construction of the effective QED Hamiltonian.

The chapter is organized as follows: in the next section, we remind the reader the form of the DCB Hamiltonian and diagrams for the QED corrections. In section "First-Order QED Corrections" the first-order QED corrections are discussed in detail. It is shown which part of QED corrections can be easily combined with the DCB Hamiltonian. The first attempts to account for all QED corrections are also described in this section. In section "Systematic Approach" the systematic approach to the construction of the effective QED Hamiltonian is discussed. Finally, in the last section, we will briefly discuss the QED effects, which are up to now beyond the effective Hamiltonian approach.

The relativistic units ($\hbar = m_e = c = 1$) are used throughout the chapter.

Dirac-Coulomb-Breit Equation and QED Corrections

As it has been already noticed in the Introduction, the good starting point for the relativistic many-electron calculations is the Dirac-Coulomb-Breit equation:

$$\hat{H}_{\rm DCB}\Psi = E\Psi, \tag{2}$$

$$\hat{H}_{\rm DCB} = \Lambda^{(+)}\left(\sum_i h_i^D + \sum_{i<j} V_{ij}\right)\Lambda^{(+)}. \tag{3}$$

Here the summation is performed over all occupied electron states, $\Lambda^{(+)}$ is the projector on the positive energy states, h^D is one-electron Dirac Hamiltonian, and V is the interelectronic interaction operator:

$$h^D = \boldsymbol{\alpha}\cdot\mathbf{p} + \beta + V_{\rm N}(r)\,, \tag{4}$$

$$V_{ij} = \frac{\alpha}{r_{ij}} - \alpha\left[\frac{\boldsymbol{\alpha}_i\cdot\boldsymbol{\alpha}_j}{r_{ij}} + \frac{1}{2}(\boldsymbol{\nabla}_i\cdot\boldsymbol{\alpha}_i)(\boldsymbol{\nabla}_j\cdot\boldsymbol{\alpha}_j)r_{ij}\right]. \tag{5}$$

In this equation

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix} \tag{6}$$

$\boldsymbol{\sigma}$ is the vector, composed of Pauli matrices,

$$\beta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \tag{7}$$

$V_{\rm N}(r)$ is the nuclear potential, and $\alpha \approx 1/137$ is the fine-structure constant.

It is mentioned in the Introduction that this equation can be self-consistently solved by means of the relativistic configuration-interaction (CI) method, multi-configurational Dirac-Fock (MCDF) method, or relativistic many-body perturbation theory (RMBPT). In section "Systematic Approach" it will be shown that the Hamiltonian (3) can be derived from the first principles of QED, using the exact formula for the one-photon exchange diagram contribution shown in Fig. 1 (taken in Coulomb gauge) and neglecting the energy dependence of the photon propagator.

However, within the rigorous QED approach arise also the diagrams which are not taken into account by the Hamiltonian (3). So in the first order of the perturbation theory in fine-structure constant α, one should evaluate also the so-called self-energy (SE, Fig. 2) and vacuum polarization (VP, Fig. 3) corrections. The wavy line in the diagrams denotes the virtual photon exchange, and the double line designates

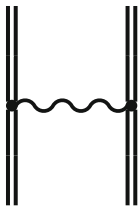


Fig. 1 One-photon exchange diagram

Fig. 2 Self-energy diagram

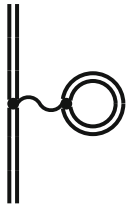


Fig. 3 Vacuum polarization diagram

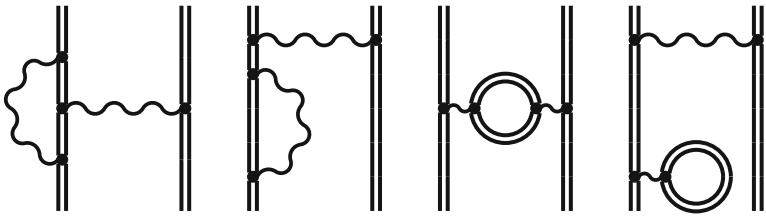


Fig. 4 SE and VP screening diagrams

the electron, moving in the external field of the nucleus. The former diagram (SE) describes the interaction of the electron with quantized electromagnetic field and the latter (VP) – with the quantized electron-positron field – disturbed by the presence of the external charge (in the absence of the external charge, the VP correction vanishes). The contribution of these two diagrams is also known as the radiative correction to the energy levels.

In the next order of the perturbation theory for the few-electron atom arise one-electron two-loop diagrams, as well as so-called SE screening and VP screening diagrams, shown in Fig. 4. The contribution of one-electron two-loop diagrams is relatively small and is up to now beyond the effective Hamiltonian approach. This contribution is discussed in section "Beyond Effective QED Hamiltonian". The contribution of the radiative screening corrections, on the other hand, very often

defines the accuracy of entire calculations. Moreover, in some cases the first-order radiative corrections may cancel each other, since SE and VP contributions have opposite signs. For example, in case of $2p_{3/2} \to 2p_{1/2}$ transition in B-like fermium ($Z = 100$), the QED contribution to the transition energy is completely defined by the screening SE and VP diagrams [16].

The ab initio evaluation of SE and VP corrections to the energy levels of bound electron is a quite difficult and time-consuming procedure. It requires the renormalization of divergent terms and evaluation of the Green function for the electron, moving in the external field of the nucleus (or several nuclei in case of molecule). The first accurate calculations of these corrections became possible relatively recently, in the 70th/80th year of the twentieth century [17–20]. Since that time, despite of tremendous progress of computational facilities, only the calculation of the entire set of second-order diagrams is almost close to be finished (the evaluation of mixed SE-VP two-loop one-electron diagrams in not yet completed to all orders in αZ). So one can conclude that ab initio QED approach does not fit to the direct incorporation to the procedure of self-consistent solution of the DCB equation.

On the other hand, the accuracy of many-electron calculations with the DCB Hamiltonian has already reached the level of QED contributions. Since the direct QED methods are too complicated to be incorporated into the solution procedure, one needs to find another way to take QED corrections into account. Numerous attempts have already been undertaken to do that. In the next sections, we will discuss how QED contribution can be expressed in terms of local potentials. We will see that while for the VP contribution it can be done in quite natural way, the search of the effective potential, approximating SE contribution is a serious problem. Finally, in section "Systematic Approach" we will explain the systematic approach to the construction of the effective QED Hamiltonian, developed recently in Ref. [21].

First-Order QED Corrections

In this section we are going to demonstrate how the first-order QED corrections should be calculated within the rigorous QED approach and how can they be incorporated into the many-electronic calculations. Let us start the discussion from the naive, but intuitively clear way to estimate the QED corrections. This method is worth to be mentioned here, since it allows to estimate the QED contribution very fast, without long calculations, and with relatively good accuracy. The approach is based on the observation that the main QED contribution during the rigorous calculations arises at the distances between electron and nucleus, close to the Compton wavelength $\lambda\!\!\!^{-}_c$.

According to this approach, one should first of all solve the DCB equation (2) by any of the available methods. Then one should evaluate the charge density for each one-electron state at the Compton wavelength distance from the nucleus and find the effective nuclear charge, which will produce the same density at this distance for

the hydrogen-like case. Finally, one should perform the interpolation of tabulated QED correction (e.g., from here: [22]) to the obtained value of the effective charge. This method is suitable not only for the single-Coulomb-center atomic problem. In the case of molecular description by means of the linear combination of atomic orbital (LCAO) method, this approximation can be also useful. If one needs only to estimate the approximate QED contribution to the many-electron calculations, this method is more than enough. To the best of the author's knowledge, this method is incorporated in many program packages used for many-electron calculations.

More accurate QED evaluations require more sophisticated methods. Before describing these methods, let us briefly discuss how one should calculate the QED corrections within the rigorous approach.

Vacuum Polarization Contribution

Let us first discuss the vacuum polarization diagram, shown in Fig. 3. Within the QED framework, this correction can be described in terms of the local potential:

$$U_{\mathrm{VP}}(\mathbf{r}_1) = \frac{\alpha}{2\pi i} \int d\mathbf{r}_2 \, \frac{1}{r_{12}} \int d\omega \, \mathrm{Tr}[G(\omega; \mathbf{r}_2, \mathbf{r}_2)] \,, \tag{8}$$

where $G(\omega; \mathbf{r}_2, \mathbf{r}_2)$ denotes the one-electron Green function. The expression (8) is ultraviolet divergent and must be renormalized. The simplest way to renormalize it is to use the expansion of the Green function:

$$G(\omega; \mathbf{x}, \mathbf{x}) = G_0(\omega; \mathbf{x}, \mathbf{x}) + \int d\mathbf{y} \; G_0(\omega; \mathbf{x}, \mathbf{y}) V_{\mathrm{N}}(\mathbf{y}) G(\omega; \mathbf{y}, \mathbf{x}) \tag{9}$$

where G_0 denotes the Green function in the absence of the external field. Applying this equation consequently, one can obtain the expansion of Eq. (8) in terms containing zero, one, two, etc. interactions with the external nuclear field. According to the Furry theorem, the terms with an even number of such interaction vanish. The resulting diagram equation is shown in Fig. 5. In this figure the double lines denote, as usual, electron in external field, while the single lines mean the free-electron propagators. The line with the cross at the end designates the interaction with the external field. In this expansion, only the first nonzero term, known as the Uehling

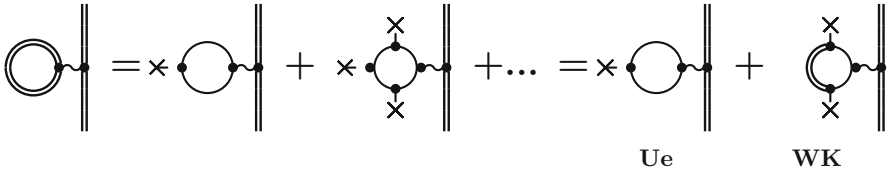


Fig. 5 Diagram equation for VP correction. "UE" and "WK" denote the diagrams, corresponding to Uehling and Wichmann-Kroll contributions, respectively

term [23], is infinite. The charge renormalization yields the following expression for the Uehling potential:

$$U_{\rm VP}^{\rm UE}(r_1) = -\frac{\alpha Z}{r_1}\frac{2\alpha}{3\pi}\int\limits_0^\infty dr_2\, 4\pi r_2 \rho(r_2) \int\limits_1^\infty dt\, \left(1+\frac{1}{2t^2}\right) \times \frac{\sqrt{t^2-1}}{4t^3}\Big(\exp(-2|r_1-r_2|t) - \exp(-2(r_1+r_2)t)\Big)\,. \qquad (10)$$

Here Z and ρ denote the nuclear charge and the density of the nuclear charge distribution, respectively. ρ is normalized as follows:

$$\int\limits_0^\infty dr\; r^2\rho(r) = 1\,. \qquad (11)$$

It is important to note that in case of several nuclei, the Uehling potential represents the sum of the Uehling potentials generated by each separate nuclei. So all the conclusions for the case of single center (atoms) can easily be extended to the many-center case (molecule).

Moreover, the Uehling potential depends only on the potential, generated by the nucleus (or nuclei). Even if one uses in the zeroth order of perturbation theory an effective potential $V_{\rm eff} = V_{\rm N} + V_{\rm scr}$, where $V_{\rm scr}$ partially describes the interelectronic interaction, the resulting Uehling term contains only the contribution of the nuclear part $V_{\rm N}$ of the total effective potential. The contribution of screening potential $V_{\rm scr}$ is canceled due to the presence of the counterterm $-V_{\rm scr}$ in the perturbation.

The remaining part of the terms in the diagram equation in Fig. 5 can be summed up again. This contribution is known as the Wichmann-Kroll contribution [24]. It can be shown that this contribution is not divergent. As one can see from Fig. 5, the calculation of this contribution requires the knowledge of the Green function in the external field (double line in the diagram). So, unlike the UE term, this term in the molecular case cannot be considered as the sum of atomic terms. Typically, the WK contribution is calculated using the partial wave renormalization technique [19]: in order to reduce the number of the space integrations, the WK term is expressed as the difference between the entire VP contribution and non-renormalized UE term: $U_{\rm WK} = U_{\rm VP}^{\rm non-ren} - U_{\rm UE}^{\rm non-ren}$. Both VP and UE terms are then expanded in the series of the partial waves with the definite angular quantum number κ: $U = \sum\limits_\kappa U_\kappa$. In this case, each partial term U_κ becomes finite for both VP and UE; however, the total sum over κ diverges also for both VP and UE parts but not for their difference. If one considers the differences between VP and UE partial contributions, one can see that the sum over κ of these differences, corresponding to the WK term, converges very fast. Usually one needs to take into account the partial waves with $|\kappa| \le 10$ for the accurate evaluation of the WK term.

The procedure, described above, is of course very complicated, to be directly included into many-electron calculations. However, this calculation can be performed only once before starting the calculations; the resulting potential can be tabulated and then incorporated into the DCB Hamiltonian (3). There exists also the possibility to use the analytical approximation of the WK potential, provided in the Ref. [20]. Concluding the discussion of the VP correction, we would like to emphasize that this correction can be described in terms of the local potential and therefore can be incorporated into the DCB Hamiltonian without loss of the computational efficiency.

Self-Energy Contribution

Unfortunately, the calculation of the SE correction cannot be easily reduced to the contribution of the local potential. In the general form, the SE operator can be defined as

$$\langle a|\Sigma(\epsilon)|b\rangle = \frac{i\alpha}{2\pi}\int_{-\infty}^{\infty} d\omega \sum_n \frac{\langle an|\alpha^\mu \alpha^\nu D_{\mu\nu}(\omega)|nb\rangle}{\epsilon - \omega - \varepsilon_n(1-i0)}\,. \tag{12}$$

Here the summation is performed over all the solutions of the one-electron Dirac equation, ε_n is the energy of the state $|n\rangle$, and $D_{\mu\nu}(\omega)$ is the photon propagator. The expression (12) is divergent and becomes finite after the mass renormalization. From Eq. (12) one can see that the SE contribution cannot be expressed in terms of the local potential, because the summation over the intermediate states $|n\rangle$ cannot be performed before the radial integrations (the states $|n\rangle$ in Eq. (12) depend on different radial variables).

There exist several approaches to the accurate evaluation of the SE correction within the rigorous QED approach [17, 18, 25–32]. However, none of them can be considered as the universal algorithm for the fast and efficient SE evaluation, suitable for the many-electronic calculations. Therefore, let us discuss the approximations of the SE contributions using the local potentials.

The first idea on how to represent the SE correction in the local potential form belongs to Welton [33]. According to his work, SE arises due to the perturbations of the classical electron trajectory induced by the fluctuations of the electromagnetic field of the vacuum.

He starts from the observation that the quantum-mechanical zero-point variation of the radiation field in vacuum leads to the electric and magnetic fields. Then considering the motion of the electron in the static electric field $V(\mathbf{r})$, he notices that due to the perturbation of the trajectory, the electron feels the field

$$V(\mathbf{r}+\delta\mathbf{r}) = (1 + \delta\mathbf{r}\cdot\nabla + \frac{1}{2}(\delta\mathbf{r}\cdot\nabla)^2 + \ldots)V(\mathbf{r})\,. \tag{13}$$

Averaging this expression over all possible values of $\delta\mathbf{r}$, he finds the average value of the potential "seen" by the electron:

$$\overline{\langle V \rangle} = (1 + \frac{1}{6}\overline{(\delta\mathbf{r})^2}\Delta + \ldots)V(\mathbf{r})\,, \tag{14}$$

which leads to the correction to the external static potential, proportional to the Laplacian of itself. Calculating this correction for the Coulomb field of the hydrogen nucleus $V_C = -\alpha/r$, he obtained the same result as Bethe [34] in his famous paper about the nonrelativistic Lamb shift calculations.

For the practical purposes, one must take into account that the average value of $(\delta\mathbf{r})^2$ in Eq. (14) is infinite and must be renormalized. Since the values for the SE corrections are known, one can use for the potential, simulating SE effect, the expression:

$$V_{\rm SE} = A\Delta V(\mathbf{r}) \tag{15}$$

and define the value of the constant A from the known results for the SE calculations. As a variation of this method, used in Ref. [5], one can evaluate the SE contribution as

$$\Delta E_{\rm SE} = \sum_n \frac{\langle n|\Delta V|n\rangle_{\rm DF}}{\langle n|\Delta V|n\rangle_{\rm H-like}} \Delta E_{\rm SE,H-like} \tag{16}$$

in order to estimate the contribution of screening diagrams. Here the sum runs over all occupied one-electron states, and the subscripts $_{\rm DF}$ and $_{\rm H-like}$ correspond to the calculations with Dirac-Fock and hydrogen-like wave functions.

Later on various attempts to approximate the Lamb shift by the short-range potentials of different form have been performed [35–37]. In particular, Pyykkö and Zhao in Ref. [35] provide a good review of previously existed method for the approximation of the SE contribution. Then they offer to approximate the SE effect by means of the potential

$$V_{\rm SE}(r) = B\exp(\beta r^2) \tag{17}$$

where two parameters B and β are chosen in such a way that they reproduce the reference values for the two quantities: the SE correction to the given energy level and the change due to SE of the M1 hyperfine integral:

$$I_{-2} = \int_0^\infty dr\ (gf + fg)\,. \tag{18}$$

Here $g(r)$ and $f(r)$ are the large and small component radial functions, respectively. After the evaluation of the parameters B and β for different values of nuclear charge

Z, the authors performed their fit with the simple quadratic expression

$$B(Z) = B_0 + B_1 Z + B_2 Z^2 \tag{19}$$

$$\beta(Z) = \beta_0 + \beta_1 Z + \beta_2 Z^2 \tag{20}$$

and obtain the very simple potential suitable for the estimation of the SE correction for arbitrary energy level and nuclear charge. Finally, the authors offer the extension of the method, which fits more calibration data. For example, the potential of the form

$$V_{\mathrm{SE}} = B_1 \exp(-\beta_1 r^2) + B_2 \exp(-\beta_2 r^2) \tag{21}$$

with $B_1 = 1{,}876.1$, $\beta_1 = 6{,}700$, $B_2 = 2{,}345.1$, and $\beta_2 = 1.675 \times 10^6$ reproduces for $2s$ state of H-like Bi the SE corrections to the energy levels, $g-$factor and hyperfine splitting. They conclude that although the elaboration of the approximate SE potentials is only in the beginning and despite the nonphysical shape of the potential (21), this approach may be of practical use for the estimation of the SE corrections to the valence energy levels of neutral atoms.

In Ref. [36] Flambaum and Ginges derived an approximate expression for a radiative potential which can be used to calculate QED corrections in many-electron atoms with an accuracy of a few percent. To this end the authors start from the case of low nuclear charge $(\alpha Z) << 1$ and derive approximate radiative potential:

$$V_{\mathrm{rad}} = V_{\mathrm{UE}} + V_{\mathrm{g}} + V_{\mathrm{f}}^{\lambda}\,, \tag{22}$$

where V_{UE} is the Uehling potential (10) and

$$V_{\mathrm{g}}(r) = \frac{\alpha}{4\pi} i\beta\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}\left[V_{\mathrm{N}}(r)\left(\int_1^\infty dt\ \frac{1}{t^2\sqrt{t^2-1}}\exp(-2tr) - 1\right)\right], \tag{23}$$

$$V_{\mathrm{f}}^{\lambda}(r) = -\frac{\alpha}{\pi}V_{\mathrm{N}}(r)\int_1^\infty dt\ \frac{\exp(-2tr)}{\sqrt{t^2-1}}\left[\left(1-\frac{1}{2t^2}\right)\left(\ln(t^2-1)+\ln\frac{4}{\lambda^2}\right)-\frac{3}{2}+\frac{1}{t^2}\right] \tag{24}$$

are the magnetic and electric form factor contributions, respectively. The derivation of Eqs. (22), (23) and (24) assumes the exchange by only the high-energy photon. Therefore, in Eq. (24) appears the cutoff parameter λ, which has been chosen by authors to be equal to $(\alpha Z)^2$ (the order of magnitude of the binding energies in the atom). Equation (22) was then completed by the approximative expression for the Wichmann-Kroll potential (see section "Vacuum Polarization Contribution") and used for the evaluation of the QED corrections.

Although all the methods described above have the physical meaning, the accuracy of the results, obtained within these approaches in the arbitrary case, remains unclear. The more systematic approach to the construction of the effective Hamiltonian in framework of QED will be discussed in the next section.

Systematic Approach

All the methods, described above, may provide in some cases very accurate results or may not. Anyway they do not allow to conclude which part of QED corrections has already been taken into account and what should still be calculated. On the contrary, within the systematic approach, it is immediately clear which part of diagrams has already been taken into account and what should be calculated in order to improve the accuracy of the result.

In 2013 Shabaev et al. offered the new approach to the construction of the effective QED Hamiltonian [21], taking into account the QED effects in lowest order. In this section a brief description of their method is provided.

Let us begin with the definition of the active space, in which the Hamiltonian acts. The theoretical description of middle- and high-Z atoms within rigorous QED approach usually starts from the solution of one-electron Dirac equation with the Hamiltonian given by Eq. (4):

$$h^D \psi_i = \epsilon_i \psi_i \,. \tag{25}$$

Therefore, it will be natural to define the active space to be formed from the Slater determinants, constructed from the one-electron unperturbed positive-energy solutions of Dirac equation ψ_i:

$$\Psi_n = \begin{vmatrix} \psi_{n_1}(\mathbf{r}_1) & \cdots & \psi_{n_N}(\mathbf{r}_1) \\ & \cdots & \\ \psi_{n_1}(\mathbf{r}_N) & \cdots & \psi_{n_N}(\mathbf{r}_N) \end{vmatrix} , \tag{26}$$

where N is the number of electrons and $\{n_i\}$ is the set of numbers, corresponding to the n-th basis function. Strictly speaking, as it has been shown in Ref. [14], there is a restriction to the total energy of these unperturbed many-electron state Ψ_n: it should not exceed the pair-creation threshold. However, to the lowest order of QED effects, one can neglect this restriction.

Now let us show how the DCB Hamiltonian (3) can be obtained from the QED principles. In order to do it, let us consider the contribution of the one-photon exchange diagram (Fig. 1). The evaluation of this diagram within the rigorous QED approach [38] leads to the formula:

$$h^{\rm int} = \sum_{k\neq l, m\neq n} |kl\rangle\langle kl| \frac{1}{2}[I(\varepsilon_k - \varepsilon_m) + I(\varepsilon_l - \varepsilon_n)]|mn\rangle\langle mn| \,, \tag{27}$$

where $|k\rangle$, $|l\rangle$, $|m\rangle$, $|n\rangle$ are the positive-energy solutions of Dirac equation with the Hamiltonian (4), $I(\omega) = e^2\alpha_1^\mu\alpha_2^\nu D_{\mu\nu}(\omega, |\mathbf{r}_1 - \mathbf{r}_2|)$, and $D_{\mu\nu}$ is the photon propagator. The operator (27) describes the interaction between only two electrons. For many-electron atom, one should then perform summation over all pairs of atomic electrons:

$$H^{\rm int} = \sum_{i<j} h_{ij}^{\rm int}\,, \tag{28}$$

where $h_{ij}^{\rm int}$ is the operator (27), acting to the one-electron wave functions, depending on $\mathbf{r}_i$ and $\mathbf{r}_j$.

Now, if one takes into account the exact form of the photon propagator in Coulomb gauge, one can obtain for $I(\omega)$ the following expression:

$$I(\omega) = \alpha\left(\frac{1}{r_{12}} - \frac{(\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2)\exp(i\omega r_{12})}{r_{12}} + \left[(\boldsymbol{\alpha}_1\cdot\boldsymbol{\nabla}_1),\left[(\boldsymbol{\alpha}_2\cdot\boldsymbol{\nabla}_2),\frac{\exp(i\omega r_{12})-1}{\omega^2 r_{12}}\right]\right]\right). \tag{29}$$

Neglecting the energy dependence and setting $\omega = 0$, one immediately obtains from Eq. (29) the interelectronic interaction operator (5). This Hamiltonian accounts for all nonrelativistic and lowest- order relativistic effects. It has been shown in Ref. [14] that in order to reach the same accuracy with the photon propagator in Feynman gauge, one should partially consider the higher-order (two and more) photon exchange diagrams.

Now let us discuss the contribution of the radiative corrections. The direct calculation of the contribution of SE and VP diagram within the two-time Green function method [38] leads to the symmetric form of one-electron QED operator:

$$\begin{aligned} h^{\rm QED} &= h^{\rm SE} + h^{\rm VP} \\ &= \sum_{\varepsilon_k>0,\varepsilon_l>0} |\psi_k\rangle\langle\psi_k| \left[\frac{1}{2}\left(\Sigma^{\rm SE}(\varepsilon_k) + \Sigma^{\rm SE}(\varepsilon_l)\right) + U^{\rm VP}\right] |\psi_l\rangle\langle\psi_l|\,. \end{aligned} \tag{30}$$

Here $\Sigma^{\rm SE}(\varepsilon)$ and $U^{\rm VP}$ are renormalized operators of SE and VP, defined by Eqs. (12) and (8), respectively.

For the total QED Hamiltonian, one needs to perform the summation over all occupied one-electron states:

$$H^{\rm QED} = \sum_i (h_i^{\rm SE} + h_i^{\rm VP})\,. \tag{31}$$

As a result in the lowest QED approximation, one obtains the Hamiltonian

$$H = \Lambda^{(+)}\left[\sum_i (h_i^{\rm D} + h_i^{\rm QED}) + \sum_{i<j} h_{ij}^{\rm int}\right]\Lambda^{(+)}\,. \tag{32}$$

In principle, this Hamiltonian already can be used in calculations. However, since up to now fast and effective algorithms for the evaluation of SE correction for the arbitrary bound state do not exist, one has to pay special attention to the SE contribution.

Before we proceed to the SE contribution, let us remind that as it has been pointed out in section "First-Order QED Corrections", the leading term describing the VP diagram contribution is given by the Uehling potential Eq. (10). The remaining Wichmann-Kroll term can be with the good accuracy approximated by formulas from Ref. [20] or calculated directly [19, 39]. Since the calculation of VP contribution can be reduced to the simple potential, we will restrict now our consideration to the SE correction.

Because of the absence of effective algorithms for arbitrary state, the operator $h^{\rm SE}$ should be reduced to its matrix elements between restricted number of low-lying bound states. This restriction, however, highly increases the interaction range of the SE operator and excludes the highly excited bound and continuum states out of consideration. As a result such a Hamiltonian may lead to wrong entire SE correction. In order to improve this situation, in Ref. [21] the following two-step procedure for the modification of $h^{\rm SE}$ was proposed.

First of all, it was proposed to split the entire SE operator in two parts. The first part represented by local operator $V^{\rm SE}_{\rm loc}(r)$ and the difference $h^{\rm SE} - V^{\rm SE}_{\rm loc}(r)$:

$$h^{\rm SE} = V^{\rm SE}_{\rm loc}(r) + (h^{\rm SE} - V^{\rm SE}_{\rm loc}(r))\,. \tag{33}$$

The local potential should have a short range and act mostly within the sphere with the radius, equal to the Compton wavelength $\lambda\!\!\!^{-}_c$. Since the operator $h^{\rm SE}$ conserves the angular quantum numbers κ and m, the local operator can be chosen different for every set of one-electron functions with the same angular quantum numbers. Namely, if one defines the projector to the one-electron states with fixed κ as

$$P_\kappa(\mathbf{n},\mathbf{n}') = \begin{pmatrix} \sum\limits_m \Omega_{\kappa m}(\mathbf{n})\Omega^\dagger_{\kappa m}(\mathbf{n}') & 0 \\ 0 & \sum\limits_m \Omega_{-\kappa m}(\mathbf{n})\Omega^\dagger_{-\kappa m}(\mathbf{n}') \end{pmatrix}, \tag{34}$$

where $\Omega_{\kappa m}$ is spherical spinor, the local potential can be defined as

$$V^{\rm SE}_{\rm loc} = \sum_\kappa V^{\rm SE}_{{\rm loc},\kappa} P_\kappa\,. \tag{35}$$

The exact form of $V_{{\rm loc},\kappa}$ can be chosen more or less arbitrary. It is only important that it satisfies the conditions above: it must be operator with a range restricted by distances close to Compton wavelength. In [21] this operator was chosen as

$$V_{{\rm loc},\kappa} = A_\kappa \exp(-r/\lambda\!\!\!^{-}_c)\,. \tag{36}$$

The constants A_κ were selected in such a way that $V_{\mathrm{loc},\kappa}$ reproduces the exact SE shift for the lowest one-electron Hydrogen-like state at given κ.

For the remaining operator $h^{\mathrm{SE}} - V_{\mathrm{loc}}^{\mathrm{SE}}(r)$, it is offered to use the matrix elements representation. However, in order to limit the range of the operator, its active space is restricted to the basis functions $\phi_i(r)$, different from Dirac wave functions $\psi_i(r)$. The functions ϕ_i must be chosen as close as possible to the functions ψ_i inside the sphere with the radius $\lambda\!\!\!^{-}_c$ and decrease much faster than ψ_i outside the sphere. The required behavior can be satisfied by the multiplication of the functions $\psi_i(r)$ by the factor

$$\rho_l(r) = \exp[-2\alpha Z(r/\lambda\!\!\!^{-}_c)/(l+1)]\,, \tag{37}$$

where $l = |\kappa + 1/2| - 1/2$ is the orbital angular momentum. However, the simple choice $\phi_i(r) = \rho_{l_i}(r)\psi_i(r)$ leads to numerical difficulties because of similar behavior of the wave functions $\psi_i(r)$ with the same l at small distances. So the basis functions ψ_i were chosen as

$$\phi_i(r) = \frac{1}{2}[I - (-1)^{s_i}\beta]\rho_{l_i}(r)\psi_i(r)\,, \tag{38}$$

where $s_i = n_i - l_i$, n_i is the principal quantum number, I is unitary matrix, and β is standard Dirac matrix (7). The number of the wave functions ϕ_i was restricted to $n_i \le 3$ for $l_i = 0$ and $n_i \le 4$ for $l_i = 1, 2$.

With this choice the SE operator can be represented as:

$$h^{\mathrm{SE}} = V_{\mathrm{loc}}^{\mathrm{SE}} + \sum_{i,k=1}^{n} |\phi_i\rangle B_{ik}\langle k|\,, \tag{39}$$

where the matrix B_{ik} must be determined in such a way that the matrix elements of the operator (39) coincide with the exact ones when evaluated with H-like wave functions ψ. This condition leads to the following system of equations:

$$B = (D^t)^{-1} S D^{-1}\,, \tag{40}$$

where the matrices D and S are defined as follows:

$$D_{ij} = \langle\phi_i|\psi_j\rangle\,, \tag{41}$$

$$S_{ij} = \langle\psi_i|\frac{1}{2}[\Sigma(\epsilon_i) + \Sigma(\epsilon_j)] - V_{\mathrm{loc}}^{\mathrm{SE}}|\psi_j\rangle\,. \tag{42}$$

Finally, the model Hamiltonian, which takes into account the SE effect in the lowest order, has a form:

$$h^{\rm SE} = V_{\rm loc}^{\rm SE} + \frac{1}{4}\sum_{i,k}\sum_{j,l}[I-(-1)^{s_i}]\rho_{l_i}(r)|\psi_i\rangle[(D^t)^{-1}]_{ij}$$
$$\times\langle\psi_j|\left\{\frac{1}{2}[\Sigma(\epsilon_j)+\Sigma(\epsilon_l)]-V_{\rm loc}^{\rm SE}\right\}|\psi_l\rangle$$
$$\times(D^{-1})_{lk}\langle\psi_k|\rho_{l_k}(r)[I-(-1)^{s_k}\beta\,, \tag{43}$$

where the summations run over ns states with the principal quantum number $n \le 3$ and over $np_{1/2}$, $np_{3/2}$, $nd_{3/2}$, and $nd_{5/2}$ states with $n \le 4$.

The values of the $\Sigma(\epsilon)$ operator diagonal and non-diagonal matrix elements, used in Eq. (43), have been calculated and tabulated in [21]. The calculations have been performed using hydrogen-like wave functions in wide range of the nuclear charges Z from 10 to 120. Using these tabulated values, it is easy to construct the model operator, which takes into account the QED effects in the lowest order.

Some examples of the successful application of the effective QED Hamiltonian are provided already in Ref. [21], where the SE correction was calculated for some alkali metal atoms and Li- and Cu-like ions and compared with previous ab initio calculations. Later on this approach was used in Ref. [13] to evaluate the QED contribution to the energy levels of Be-like iron.

Beyond Effective QED Hamiltonian

Before finishing the discussion of the effective QED Hamiltonians, let us name the QED effects, which are not taken into account by all the methods, discussed above. The contribution of these effects is known to be small even for the case of low-lying states of heavy few-electron ions, where the QED effects are most pronounced. However, these contribution should be at least mentioned in this chapter.

First of all, it is the negative-energy states contribution. The Hamiltonian (3) contains the projector operators to the positive-energy states. So these states are excluded from the consideration from the very beginning. Partially, these states are taken into account via SE and VP operators. In particular, the VP contribution cannot appear without the negative-energy states, since it arises due to the asymmetry between positive- and negative-energy solutions caused by the presence of the external charge. However, the complete estimation of the negative-energy states contribution requires the accounting for these states also in all two- and more-photon exchange diagrams.

Another important effect neglected by the Hamiltonian (3) is frequency dependence of the photon propagator, or the retardation effect. Within the rigorous QED framework, the interaction between electrons is described by the virtual photon exchange (see Eq. (29)). If the initial and final one-electron states have the same energy, the virtual photon has zero energy, otherwise it must carry the difference between the initial and final states energy from one electron to another. Since the energy of the one-electron states changes at each step of the self-consistent

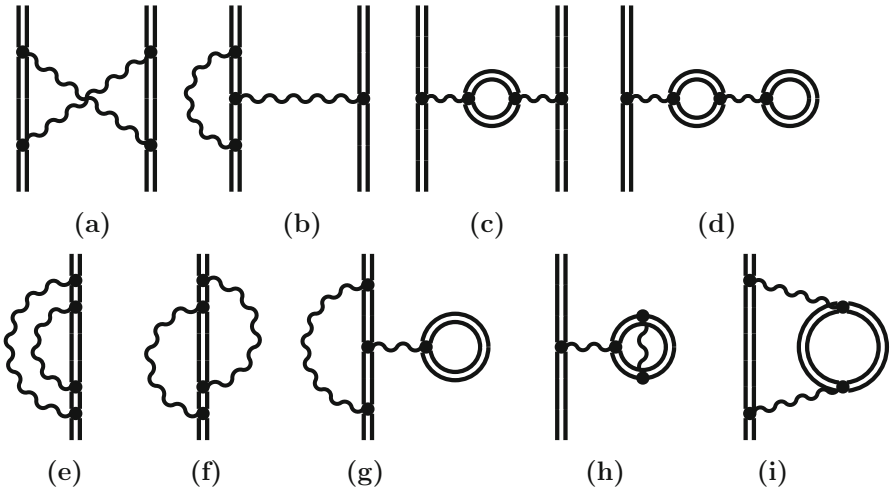


Fig. 6 Example of diagrams, not included in the effective QED Hamiltonian

procedure and because the active space is restricted usually by the low-energy one-electron solutions, so the energy difference is usually $\Delta E << mc^2$, one can neglect the retardation effects during the solution of DCB equation. However, the energy dependence contribution can be evaluated after the solutions of the DCB equation are found.

At the level of Feynman diagrams, one can point to many of them, which are not taken into account by the effective Hamiltonian. The set of the second-order QED diagrams, which is not taken into account by the effective Hamiltonian, is shown in Fig. 6. Of course, all higher- order diagrams, which have any building block, taken from the Fig. 6 are also neglected. The diagram (a) from this figure is known as the crossed two-photon exchange diagram. Its contribution is usually small and vanishes in the Breit approximation. The contributions of the so-called vertex self-energy diagram as well as the vacuum polarization correction to the photon propagator (diagrams (b) and (c) in the Fig. 6) are also neglected within the effective Hamiltonian approach, although their contribution may reach up to 15 % of the total screening QED diagrams, shown in Fig. 4. So if one needs the highly accurate calculations, one should account for these diagrams. Within the effective QED Hamiltonian approach, these diagrams could theoretically be taken into account by some corrections to the interelectronic interaction operator (5). However, up to now no attempts of such corrections have been performed.

Unlike the two-electron second-order diagrams, the contribution of the one-electron two-loop diagrams (d)–(i) is usually very small. It can be important only for the low-lying states of highly ($Z \gtrsim 80$) charged atoms. The contribution of the diagrams (d) and (h) can be expressed (at least in lowest order in αZ) in the form of local potential, which can be introduced into the total effective QED Hamiltonian. However, since all the diagrams (d)–(i) have the same order of magnitude but

different signs, it is more natural to take into account the entire contribution of these diagrams. Due to the relative smallness of this effect, one can do it using the known results for the hydrogen-like functions [40].

Summary

To summarize, we have discussed in this chapter the construction of the Hamiltonian, which allows to perform many-electron calculations taking into account the lowest-order QED effects. We have found that the main difficulties arise if one wants to account for the contribution of self-energy diagram. We discussed several approaches, which allow to insert this effect into many-electron Hamiltonian. We also demonstrated which QED effects remain up to now beyond the effective QED Hamiltonian approach and must be estimated by other means.

References

1. Wood CS, Bennett SC, Cho D, Masterson BP, Roberts JL, Tanner CE, Wieman CE (1997) Measurement of parity nonconservation and an anapole moment in cesium. Science 275(5307):1759
2. Bennett SC, Wieman CE (1999) Measurement if the 6S–>7S transition polarizability in atomic cesium and an improved test of the standard model. Phys Rev Lett 82(12):2484. [ibid. 82:4153(E) (1999); 83:889(E) (1999)]
3. Grant I (1970) Relativistic calculation of atomic structures. Adv Phys 19(82):747
4. Desclaux J (1975) A multiconfiguration relativistic DIRAC-FOCK program. Comput Phys Commun 9(1):31
5. Indelicato P, Gorveix O, Desclaux JP (1987) Multiconfiguration Dirac-Fock studies of two-electron ions: II radiative corrections and comparison with experiment. J Phys B 20(4):651
6. Indelicato P, Desclaux JP (1990) Multiconfiguration Dirac-Fock calculations of transition energies with QED corrections in three-electron ions. Phys Rev A 42(9):5139
7. Blundell SA, Johnson WR, Sapirstein J (1990) Improved many-body perturbation-theory calculations of the $n = 2$ states of lithiumlike uranium. Phys Rev A 41(3):1698
8. Indelicato P, Lindroth E (1992) Relativistic effects, correlation, and QED corrections on $K\alpha$ transitions in medium to very heavy atoms. Phys Rev A 46(5):2426
9. Ynnerman A, James J, Lindgren I, Persson H, Salomonson S (1994) Many-body calculation of the $2p_{1/2,3/2}$-$2s_{1/2}$ transition energies in Li-like ^{238}U. Phys Rev A 50(6):4671
10. Glazov D, Shabaev V, Tupitsyn I, Volotka A, Yerokhin V, Plunien G, Soff G (2004) Relativistic and QED corrections to the g factor of Li-like ions. Phys Rev A 70:062104
11. Tupitsyn I, Volotka A, Glazov D, Shabaev V, Plunien G, Crespo López-Urrutia J, Lapierre A, Ullrich J (2005) Magnetic-dipole transition probabilities in B-like and Be-like ions. Phys Rev A 72:062503
12. Parpia F, Fischer CF, Grant I (2006) GRASP92: a package for large-scale relativistic atomic structure calculations.Comput Phys Commun 175(11–12):745
13. Yerokhin VA, Surzhykov A, Fritzsche S (2014) Relativistic configuration-interaction calculation of Kα transition energies in berylliumlike iron. Phys Rev A 90:022509. doi:10.1103/PhysRevA.90.022509
14. Shabaev VM (1993) Schrödinger-like equation for the relativistic few-electron atom. J Phys B 26:4703

15. Artemyev AN, Shabaev VM, Tupitsyn II, Plunien G, Yerokhin VA (2007) QED calculation of the $2p_{3/2} - 2p_{1/2}$ transition energy in Boronlike Argon. Phys Rev Lett 98(17):173004
16. Artemyev AN, Shabaev VM, Tupitsyn II, Plunien G, Surzhykov A, Fritzsche S (2013) Ab initio calculations of the $2p_{\frac{3}{2}} - 2p_{\frac{1}{2}}$ fine-structure splitting in boronlike ions. Phys Rev A 88:032518. doi:10.1103/PhysRevA.88.032518
17. Mohr PJ (1974) Self-energy radiative corrections in hydrogen-like systems. Ann Phys (New York) 88:26
18. Mohr PJ (1974) Numerical evaluation of the $1S_{1/2}$-state radiative level shift. Ann Phys (New York) 88:52
19. Soff G, Mohr P (1988) Vacuum polarization in a strong external field. Phys Rev A 38(10):5066
20. Manakov NL, Nekipelov AA, Fainshtein AG (1989) Vacuum polarization by a strong coulomb field and its contribution to the spectra of multiply-charged ions. Zh Eksp Teor Fiz 95:1167. [Sov Phys JETP 68:673 (1989)]
21. Shabaev VM, Tupitsyn II, Yerokhin VA (2013) Model operator approach to the Lamb shift calculations in relativistic many-electron atoms. Phys Rev A 88:012513
22. Johnson WR, Soff G (1985) The lamb shift in hydrogen-like atoms, $1 \leq Z \leq 110$. At Data Nucl Data Tables 33(3):405
23. Uehling EA (1935) Polarization effects in the positron theory. Phys Rev 48:55
24. Wichmann EH, Kroll NM (1956) Vacuum polarization in a strong coulomb field. Phys Rev A 101(2):843
25. Cheng KT, Johnson WR (1976) Self-energy corrections to the K-electron binding in heavy and superheavy atoms. Phys Rev A 14(6):1943. doi:10.1103/PhysRevA.14.1943
26. Snyderman NJ (1991) Electron radiative self-energy of highly stripped heavy atoms. Ann Phys (New York) 211:43
27. Blundell SA, Snyderman NJ (1991) Basis-set approach to calculating the radiative self-energy in highly ionized atoms. Phys Rev A 44(3):R1427
28. Indelicato P, Mohr P (1992) Coordinate-space approach to the bound-electron self-energy. Phys Rev A 46:172
29. Lindgren I, Persson H, Salomonson S, Ynnerman A (1993) Bound-state self-energy calculation using partial wave renormalization. Phys Rev A 47(6):R4555
30. Persson H, Lindgren I, Salomonson S (1993) A new approach to the electron self energy calculation. Physica Scripta T46:125
31. Labzowsky L, Goidenko I, Tokman M, Pyykkö P (1999) Calculated self-energy contributions for an ns valence electron using the multiple-commutator method. Phys Rev A 59(4):2707
32. Yerokhin VA, Shabaev VM (1999) First order self-energy correction in hydrogen-like systems. Phys Rev A 60(2):800
33. Welton TA (1948) Some observable effects of the quantum-mechanical fluctuations of the electromagnetic field. Phys Rev 74(9):1157
34. Bethe HA (1947) The electromagnetic shift of energy levels. Phys Rev 72:339
35. Pyykkö P, Zhao LB (2003) Search for effective local model potentials for simulation of quantum electrodynamic effects in relativistic calculations. JPB 36(8):1469
36. Flambaum VV, Ginges JSM (2005) Radiative potential and calculations of QED radiative corrections to energy levels and electromagnetic amplitudes in many-electron atoms. Phys Rev A 72:052115. doi:10.1103/PhysRevA.72.052115
37. Thierfelder C, Schwerdtfeger P (2010) Quantum electrodynamic corrections for the valence shell in heavy many-electron atoms. Phys Rev A 82:062503. doi:10.1103/PhysRevA.82.062503
38. Shabaev VM (2002) Two-time Green's function method in quantum electrodynamics of high-Z few-electron atoms. Phys Rep 356(3):119
39. Persson H, Lindgren I, Salomonson S, Sunnergren P (1993) Accurate vacuum polarization contributions. Phys Rev A 48(4):2772
40. Yerokhin VA, Indelicato P, Shabaev VM (2008) Two-loop QED corrections with closed fermion loops. Phys Rev A 77:062510. doi:10.1103/PhysRevA.77.062510

Two-Time Greens Function Method

8

Anton N. Artemyev

Contents

Introduction 287
2N-Times Green's Function 289
Analytical Properties of Two-Time Green's Function 294
Energy Shift of a Single Level 299
Quasi-degenerate Levels 304
Summary 311
References 311

Abstract

In this chapter we describe the Two-Time Greens Function (TTGF) method developed by V. M. Shabaev. This method allows derivation of the formulas for the energy shift and other QED effects. Unlike the preceeding methods, the TTGF one is suitable not only in the case of single isolated, but also for the (quasi-)degenerate levels. Starting from the very basic principles and concepts of QED, we will demonstrate, how to derive basic formulas with the help of TTGF method and apply them to the case of many-electron systems.

Introduction

Quantum electrodynamics (QED) is the most accurate and the best confirmed theory in modern physics. Since its foundation in 1928 by the famous work of P. Dirac about relativistic electron [1], the theory has been intensively developed. The modern formulation has been finished in the 50th years of the twentieth century. In principle, to that time all necessary equations have already been derived and all necessary techniques, e.g., Feynman diagrams [2], elaborated. However,

A.N. Artemyev (✉)
Institute of Physics, Kassel University, Kassel, Germany
e-mail: anton.artemyev@physik.uni-kassel.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_28

the formulation of the basic principles and techniques does not necessary mean that all the problems and challenges, arising during the investigations, can be immediately solved. As an example we can mention here more common case of nonrelativistic quantum mechanics, where the basic equations and main principles have been formulated even earlier than those of QED. However, some problems of nonrelativistic quantum mechanics remain unsolved up to now. The Schrödinger equation has laid a foundation of many different branches of theory, like atomic and molecular physic, solid body physics, etc. Each branch has developed their own approaches, based on the Schrödinger equation and suitable for its own purposes. To deal with the new problems, one has to invent new methods, allowing at least approximate solution of them. The same occurs in QED.

Due to the much more complicated basic equations, QED cannot be successfully applied to a wide range of systems, unlike with nonrelativistic quantum mechanics. Nowadays QED is most successful in description of few-body systems, like atoms, ions, and exotic atoms (e.g., muonium, positronium, etc.). Usually, each class of problems requires for its solution the elaboration of the method, based on main principles of QED. The difference in methods can be caused, for example, by different values of the nuclear charge, different mass of particles, etc.

In this chapter we are going to concentrate on the atoms and ions with the middle and high values of nuclear charge. Although all methods, elaborated for such kind of systems, are usually valid for low-Z as well, the latter case allows more efficient methods due to additional perturbation expansion in powers of small (for low-Z systems) parameter αZ ($\alpha \approx 1/137$ is fine structure constant). Historically, the hydrogen atom was the first object investigated by means of QED. See the ▸ Chap. 6, "QED Effects and Challenges" of this book for more details. The experimental investigation of heavier systems, e.g., highly charged ions, has become possible since the late 60th years of the twentieth century. Almost simultaneously the first QED calculations in such systems were performed.

The first method for deriving the formal expressions for the QED energy shift of a bound-state level was formulated by Gell-Mann and Low [3] and by Sucher [4]. This method is based on the introduction of an adiabatically damped factor, $\exp(-\lambda|t|)$, in the interaction Hamiltonian and in the expression of the energy shift in terms of the so-called adiabatic S_λ-matrix elements, with final evaluation of $\lim_{\lambda\to 0}$. Due to its simple formulation, the Gell-Mann–Low–Sucher formula for the energy shift became very popular in the literature related to high-Z few-electron systems. However, the practical use of this method demonstrated that it had several serious disadvantages. One of them is the very complicated derivation of the formal expressions for the so-called reducible diagrams. As reducible diagrams we imply the diagrams where an intermediate-state energy of the atom coincides with the reference-state energy.

Another serious drawback of the Gell-Mann–Low–Sucher method is the fact that this method requires special investigation of the renormalization procedure since the adiabatic S_λ-matrix has ultraviolet divergences. The adiabatically damped factor, $\exp(-\lambda|t|)$, is non-covariant, and, therefore, the ultraviolet divergences cannot be

removed from S_λ if $\lambda \neq 0$. However, from the physical point of view, one may expect these divergences to cancel each other in the expression for the energy shift. Therefore, they may be disregarded in the calculation of the energy shift for a single level. For the case of degenerate levels, however, this problem remains since we cannot expect that the standard renormalization procedure makes the secular operator finite in the ultraviolet limit [5, 6].

It is also impossible within the framework of this method to deal with the quasi-degenerate states (the levels with the same symmetry and having close energy in zeroth-order approximation). Taking into account all the reasons, it becomes clear that although the Gell-Mann–Low–Sucher approach can be used for the solution of the wide class of the problems, the creation of more effective methods was necessary.

The development of the new approaches took some time. Finally, within approximately one decade, there appeared three theoretical methods, which significantly improved the Gell-Mann–Low–Sucher approach:

- the two-time Green's function (TTGF) method [7];
- the covariant-evolution-operator method [8];
- the spectral line profile method [9].

This chapter is devoted to the first of abovementioned methods, the TTGF. Because the current book is addressed to the quantum chemists, we will concentrate within this chapter only on the problem of the energy levels of few-electron system and leave untouched other problems, which can be solved by means of TTGF, like the scattering problem, g-factor calculations, hyperfine splitting of the energy levels, etc. We are going to stress out at least the most important ideas and features of the method. For the detailed derivation, we refer to the original paper [7] and references therein.

The relativistic units ($\hbar = m_e = c = 1$) and the Heaviside charge unit ($\alpha = e^2/4\pi$, $e < 0$) are used throughout the chapter if not stated otherwise. We try to keep the notations from the original work [7], so in case of any questions, one can easily find in this paper more detailed derivations.

2*N*-Times Green's Function

Before starting the description of the method, let us define the system which we are going to investigate. We are interested in the system consisting of N electrons moving in the field of the nucleus with the charge Z. For highly charged ions, it is considered that $Z \gg N$. In this case the average interaction of an electron with other electrons is smaller than its interaction with nucleus by the factor $1/Z$. The interaction with the quantized electromagnetic field in turn is by factor $\alpha \approx 1/137$

smaller than that with nucleus. Therefore, it is natural to suppose that the movement of the electrons in zeroth approximation is described by the Dirac equation:

$$(\boldsymbol{\alpha}\cdot\boldsymbol{p}+\beta+V_N(r))\psi_n(\boldsymbol{r})=\varepsilon_n\psi_n(\boldsymbol{r}), \tag{1}$$

where $\boldsymbol{\alpha}$ and β are the Dirac matrices, V_N is the potential of the nucleus, and $\boldsymbol{p}=-i\nabla$ is momentum operator. The potential can be conveniently represented as

$$V_N=-\frac{\alpha Z(r)}{r} \tag{2}$$

where $r-$dependent charge is introduced to take into account the effect of the charge distribution inside the nucleus. Without losing the generality, the potential V_N can be replaced by the effective potential $V_{\rm eff}$ which partially describes the interelectronic interaction. One should do it especially for the case of neutral atoms or ions with the number of electrons $N\approx Z$. In such systems the average interaction of the electron with other electrons is comparable with the interaction with nucleus and should be better taken into account in the zeroth order at least partially by means of $V_{\rm eff}$. In this case additional interaction $\delta V=V_N-V_{\rm eff}$ must be taken into account perturbatively. See for the details Refs. [10, 11], where four different effective potentials have been applied for the calculation of boron-like ions.

It can be shown that the complete information about the system with N electrons contains $2N-$times Green's function defined as

$$G(x_1'\cdots x_N';x_1\cdots x_N)=\langle 0|T\psi(x_1')\cdots\psi(x_N')\bar{\psi}(x_1)\cdots\bar{\psi}(x_N')|0\rangle\,. \tag{3}$$

Here $|0\rangle$ denotes vacuum, T is the symbol of $T-$ordered product, ψ is the electron-positron field operator in the Heisenberg representation, and $\bar{\psi}=\psi^\dagger\gamma_0$ is the Dirac conjugation. Note that x denotes time–space four vector; hence, this function depends on $2N$ different times.

In the interaction representation (see, e.g., [12, 13]) the Green's function is given by

$$\begin{aligned}
&G(x_1',\ldots x_N';x_1,\ldots x_N)\\
&=\frac{\langle 0|T\psi_{\rm in}(x_1')\cdots\psi_{\rm in}(x_N')\overline{\psi}_{\rm in}(x_N)\cdots\overline{\psi}_{\rm in}(x_1)\exp\{-i\int d^4z\,\mathcal{H}_I(z)\}|0\rangle}{\langle 0|T\exp\{-i\int d^4z\,\mathcal{H}_I(z)\}|0\rangle} \qquad (4)\\
&=\Big\{\sum_{m=0}^{\infty}\frac{(-i)^m}{m!}\int d^4y_1\cdots d^4y_m\;\langle 0|T\psi_{\rm in}(x_1')\cdots\psi_{\rm in}(x_N')\overline{\psi}_{\rm in}(x_N)\cdots\overline{\psi}_{\rm in}(x_1)\\
&\qquad\times\mathcal{H}_I(y_1)\cdots\mathcal{H}_I(y_m)|0\rangle\Big\}\\
&\times\Big\{\sum_{l=0}^{\infty}\frac{(-i)^l}{l!}\int d^4z_1\cdots d^4z_l\;\langle 0|T\mathcal{H}_I(z_1)\cdots\mathcal{H}_I(z_l)|0\rangle\Big\}^{-1} \qquad (5)
\end{aligned}$$

where

$$\mathcal{H}_I(x) = \frac{e}{2}\,[\overline{\psi}_{\rm in}(x)\gamma_\mu, \psi_{\rm in}(x)]A^\mu_{in}(x) - \frac{\delta m}{2}\,[\overline{\psi}_{\rm in}(x), \psi_{\rm in}(x)] \tag{6}$$

is the interaction Hamiltonian. The commutators in Eq. (6) refer to operators only. The first term in (6) describes the interaction of the electron–positron field with the quantized electromagnetic field, and the second one is the mass renormalization counterterm. The significant difference of the methods, developed for the middle- and highly charged atoms from that elaborated for the low-Z, is the fact that the interaction of the electrons with the Coulomb field of the nucleus is included in the unperturbed Hamiltonian. In other words we are working in the Furry picture. It is easy to understand why one should better include the Coulomb field of the nucleus into the zeroth order of the perturbation theory. As one can see from the Eq. (2), the Coulomb interaction with the nucleus is always accompanied by the factor αZ. While for low-Z systems this parameter is small and the perturbation series over this parameter converge very fast, in the case of high-Z this parameter is close to unity. It is clear that the expansion in powers of this parameter has no sense.

The Eq. (5) gives a key for the construction of the Green's function. To do it one expands the exponents in powers of the interaction Hamiltonians (second and third lines of Eq. (5)) and then uses the Wick's theorem (see, e.g., [12]). According to this theorem, the time-ordered product of field operators is equal to the sum of normal-ordered products with all possible contractions between the operators

$$\begin{aligned} T(ABCD\cdots) = N(ABCD\cdots) + N(A^aB^aCD\cdots) + N(A^aBC^aD\cdots) \\ + \text{ all possible contractions}, \end{aligned} \tag{7}$$

where N is the normal-ordered product operator and the superscripts denote the contraction between the corresponding operators. The normal-ordered product yields zero if acts to the vacuum, because the annihilation operators act before the creation ones. The contraction between neighboring operators is defined by

$$A^aB^a = T(AB) - N(AB) = \langle 0|T(AB)|0\rangle. \tag{8}$$

If the contracted operators are boson operators, they can be put one next to another. If the contracted operators are fermion operators, they also can be put one next to another, but in this case the expression must be multiplied by the parity of the permutation of the fermion operators. Since in the Green's function the vacuum expectation value is calculated, only the term with all operators contracted remains on the right-hand side of Eq. (7).

The contractions between the electron–positron fields and between the photon fields provide the following propagators:

$$\langle 0|T\psi_{\rm in}(x)\overline{\psi}_{\rm in}(y)|0\rangle = \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega\ \sum_n \frac{\psi_n(\boldsymbol{x})\overline{\psi}_n(\boldsymbol{y})}{\omega - \varepsilon_n(1-i0)}\exp\left[-i\omega(x^0-y^0)\right] \tag{9}$$

and (in Feynman gauge)

$$\langle 0|TA^{\mu}_{\rm in}(x)A^{\nu}_{\rm in}(y)|0\rangle = -ig^{\mu\nu}\int \frac{d^4k}{(2\pi)^4}\,\frac{\exp\left[-ik\cdot(x-y)\right]}{k^2+i0}\,. \tag{10}$$

In Eq. (9) the index n runs over all bound and continuum states, so effectively it implies the integration over all continuum states.

The denominator in Eq. (4) describes unobservable vacuum–vacuum transitions, and, as can be shown (see, e.g., [12]), it cancels disconnected vacuum–vacuum sub-diagrams from the numerator. Therefore, one can omit all diagrams containing disconnected vacuum–vacuum sub-diagrams in the numerator and replace the denominator by 1.

Since in practical (especially chemical) calculations one is usually interested in the value of the energy levels, it is convenient to make the Green's function explicitly dependent on the energy. It can be reached by the Fourier transform with respect to time variables:

$$\begin{aligned} &G\left((p'^0_1,\boldsymbol{x}'_1),\ldots,(p'^0_N,\boldsymbol{x}'_N);(p^0_1,\boldsymbol{x}_1),\ldots,(p^0_N,\boldsymbol{x}_N)\right)\\ &=(2\pi)^{-2N}\int_{-\infty}^{\infty} dx^0_1\cdots dx^0_N dx'^0_1\cdots dx'^0_N\\ &\quad\times\exp\left(ip'^0_1x'^0_1+\cdots+ip'^0_Nx'^0_N-ip^0_1x^0_1-\cdots-ip^0_Nx^0_N\right)\\ &\quad\times G(x'_1,\ldots,x'_N;x_1,\ldots,x_N). \end{aligned} \tag{11}$$

The next step in the evaluation of the Green's function is direct application of the perturbational expansion. To simplify this procedure, the Feynman diagram technique [2] is invented. Instead of long and laborious order-by-order expansion of T-exponent in Eq. (4) and evaluation of the corresponding contractions, one can draw the diagrams and write the corresponding formulas following the rules. So for the Green's function (11), the following Feynman rules can be established:

External electron line. $\frac{i}{2\pi}S(\omega;\boldsymbol{x},\boldsymbol{y})$, with

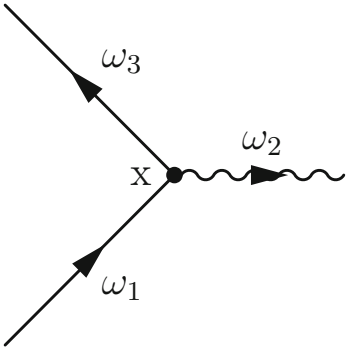
$$S(\omega;\boldsymbol{x},\boldsymbol{y})=\sum_n \frac{\psi_n(\boldsymbol{x})\bar{\psi}_n(\boldsymbol{y})}{\omega-\varepsilon_n(1-i0)}\,. \tag{12}$$

Internal electron line. $\frac{i}{2\pi}\int\limits_{-\infty}^{\infty} d\omega\, S(\omega; \boldsymbol{x}, \boldsymbol{y})$.

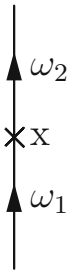
Disconnected electron line. $\frac{i}{2\pi}\delta(\omega_1 - \omega_2)\, S(\omega; \boldsymbol{x}, \boldsymbol{y})$.

Internal photon line. $\frac{i}{2\pi}\int\limits_{-\infty}^{\infty} d\omega\, D_{\mu\nu}(\omega; \boldsymbol{x} - \boldsymbol{y})$, where in Feynman gauge

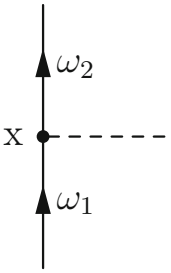
$$D_{\mu\nu}(\omega; \boldsymbol{x} - \boldsymbol{y}) = -g_{\mu\nu} \int \frac{d\boldsymbol{k}}{8\pi^3} \frac{\exp(i\boldsymbol{k} \cdot (\boldsymbol{x} - \boldsymbol{y}))}{\omega^2 - k^2 + i0}. \tag{13}$$



Vertex. $-2\pi i e \gamma^{\mu} \delta(\omega_1 - \omega_2 - \omega_3) \int d\boldsymbol{x}$.



Mass counterterm. $2\pi i \;\; \delta m \;\; \delta(\omega_1 - \omega_2) \int d\boldsymbol{x}$.



If in addition external potential V is present.
$2\pi i \;\; \gamma^0 \;\; \delta(\omega_1 - \omega_2) \int d\boldsymbol{x}\, V(\boldsymbol{x})$.

Additionally the contribution of the diagram should be multiplied by the symmetry factor $(-1)^{N+P}$, where N is the number of closed electron (in general case fermion) loops and P is the parity of the permutation of the outgoing electron coordinates with respect to the incoming ones.

This set of rules provides a possibility to construct the Green's function perturbative expansion order by order. After the Green's function is constructed, one can extract from it all necessary information about the energy levels of the system. However, to extract the information directly from $2N-$times Green's function is a very complicated task, because it depends on $2N$ relative times (energies). Therefore it was proposed to use instead of $2N-$times Green's function $G(t'_1, \cdots, t'_N; t_1, \cdots, t_N)$, the function, which depends only on two different times (energies):

$$\tilde{G}(t', t) \equiv G(t'_1 = t'_2 = \cdots = t'_N \equiv t'; t_1 = t_2 = \cdots = t_N \equiv t). \tag{14}$$

In other words, all initial (as well as all final) times in this function are equal. In the next sections, we will demonstrate that this function still contains all information about the energy levels of the system and show how it can be used for the calculations of the energy levels.

Analytical Properties of Two-Time Green's Function

As we noticed in the previous section, it is much easier to extract the information about the energy levels from the Green's function, which depends on only two time (energy) variables, than from the function, depending on $2N$ such variables. In this

section we will demonstrate that setting all initial (final) times to be equal each other, like it is done in Eq. (14), does not lead to any loss of information.

Let us introduce the Fourier transform of the two-time Green's function by

$$\begin{aligned}
&\mathcal{G}(E;\boldsymbol{x}'_1,\ldots\boldsymbol{x}'_N;\boldsymbol{x}_1,\ldots\boldsymbol{x}_N)\delta(E-E') \\
&\quad=\frac{1}{2\pi i}\frac{1}{N!}\int_{-\infty}^{\infty}dx^0dx'^0\ \exp\left(iE'x'^0-iEx^0\right) \\
&\qquad\times\langle 0|T\psi(x'^0,\boldsymbol{x}'_1)\cdots\psi(x'^0,\boldsymbol{x}'_N)\overline{\psi}(x^0,\boldsymbol{x}_N)\cdots\overline{\psi}(x^0,\boldsymbol{x}_1)|0\rangle, \qquad (15)
\end{aligned}$$

To study the analytical properties of the two-time Green's function, we derive the spectral representation for $\mathcal{G}$.

Let us insert into the matrix element in Eq. (15) the unitary operator I expressing it as an expansion over the complete set of the eigenstates of the Hamiltonian of the system:

$$I=\sum_n|n\rangle\langle n|, \qquad (16)$$

$$H|n\rangle=E_n|n\rangle. \qquad (17)$$

Using the time-shift transformation rule for the Heisenberg operators

$$\psi\left(x^0,\boldsymbol{x}\right)=\exp\left(iHx^0\right)\psi(0,\boldsymbol{x})\exp\left(-iHx^0\right) \qquad (18)$$

and the fact, that

$$\exp(aH)|n\rangle=\exp(aE_n)|n\rangle, \qquad (19)$$

we find

$$\begin{aligned}
&\mathcal{G}(E;\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_N;\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)\delta(E-E') \\
&=\frac{1}{2\pi i}\frac{1}{N!}\int_{-\infty}^{\infty}dx^0dx'^0\ \exp\left(iE'x'^0-iEx^0\right) \\
&\quad\times\Big\{\theta\left(x'^0-x^0\right)\sum_n\exp\left[i(E_0-E_n)(x'^0-x^0)\right]\langle 0|\psi(0,\boldsymbol{x}'_1)\cdots\psi(0,\boldsymbol{x}'_N)|n\rangle \\
&\quad\times\langle n|\overline{\psi}(0,\boldsymbol{x}_N)\cdots\overline{\psi}(0,\boldsymbol{x}_1)|0\rangle+(-1)^{N^2}\theta\left(x^0-x'^0\right) \\
&\quad\sum_n\exp\left[i(E_0-E_n)\left(x^0-x'^0\right)\right] \\
&\quad\times\langle 0|\overline{\psi}(0,\boldsymbol{x}_N)\cdots\overline{\psi}(0,\boldsymbol{x}_1)|n\rangle\langle n|\psi(0,\boldsymbol{x}'_1)\cdots\psi(0,\boldsymbol{x}'_N)|0\rangle\Big\}. \qquad (20)
\end{aligned}$$

Assuming, for simplicity, $E_0 = 0$ (it corresponds to choosing the vacuum energy as the origin of reference) and taking into account that

$$\int_{-\infty}^{\infty} dx^0 dx'^0\, \theta\left(x'^0 - x^0\right) \exp\left[-iE_n\left(x'^0 - x^0\right)\right] \exp\left[i\left(E'x'^0 - Ex^0\right)\right] = 2\pi\delta(E' - E)\frac{i}{E - E_n + i0}, \quad (21)$$

$$\int_{-\infty}^{\infty} dx^0 dx'^0\, \theta(x^0 - x'^0) \exp\left[-iE_n\left(x^0 - x'^0\right)\right] \exp\left[i\left(E'x'^0 - Ex^0\right)\right] = -2\pi\delta(E' - E)\frac{i}{E + E_n - i0}, \quad (22)$$

we obtain

$$\mathcal{G}(E; \boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sum_n \frac{\Phi_n(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N)\overline{\Phi}_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)}{E - E_n + i0} - (-1)^N \sum_n \frac{\Xi_n(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N)\overline{\Xi}_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)}{E + E_n - i0}. \quad (23)$$

Here

$$\Phi_n(\boldsymbol{x}_1, \ldots \boldsymbol{x}_N) = \frac{1}{\sqrt{N!}}\langle 0|\psi(0, \boldsymbol{x}_1)\cdots\psi(0, \boldsymbol{x}_N)|n\rangle, \quad (24)$$

$$\Xi_n(\boldsymbol{x}_1, \ldots \boldsymbol{x}_N) = \frac{1}{\sqrt{N!}}\langle n|\psi(0, \boldsymbol{x}_1)\cdots\psi(0, \boldsymbol{x}_N)|0\rangle. \quad (25)$$

In Eq. (23) the summation runs over all bound and continuum states of the system of the interacting fields. Let us now introduce the functions

$$A(E; \boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sum_n \delta(E - E_n)\Phi_n(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N)\overline{\Phi}_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N), \quad (26)$$

$$B(E; \boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \sum_n \delta(E - E_n)\Xi_n(\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N)\overline{\Xi}_n(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N), \quad (27)$$

which satisfy the conditions

$$\int_{-\infty}^{\infty} dE\ A(E; \boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_N; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \frac{1}{N!}\langle 0|\psi(0, \boldsymbol{x}'_1)\cdots\psi(0, \boldsymbol{x}'_N) \times \overline{\psi}(0, \boldsymbol{x}_N)\cdots\overline{\psi}(0, \boldsymbol{x}_1)|0\rangle, \quad (28)$$

$$\int_{-\infty}^{\infty} dE\ B(E;\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_N;\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) = \frac{1}{N!}\langle 0|\overline{\psi}(0,\boldsymbol{x}_N)\cdots\overline{\psi}(0,\boldsymbol{x}_1) \times \psi(0,\boldsymbol{x}'_1)\cdots\psi(0,\boldsymbol{x}'_N)|0\rangle. \tag{29}$$

With the help of these functions, we can rewrite Eq. (23) as

$$\mathcal{G}(E) = \int_0^{\infty} dE'\ \frac{A(E')}{E-E'+i0} - (-1)^N \int_0^{\infty} dE'\ \frac{B(E')}{E+E'-i0}, \tag{30}$$

where we have omitted the variables $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N,\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_N$ and have taken into account that $A(E') = B(E') = 0$ for $E' < 0$ since $E_n \geq 0$.

Due to charge conservation, only states with an electric charge of eN contribute to A in the sum over n in the right-hand side of Eq. (26), and only states with an electric charge of $-eN$ contribute to B in the sum over n in the right-hand side of Eq. (27). This can easily be shown by using the following commutation relations

$$[Q,\psi(x)] = -e\psi(x), \qquad [Q,\overline{\psi}(x)] = e\overline{\psi}(x), \tag{31}$$

where Q is the charge operator in the Heisenberg representation. Therefore, Eq. (30) can be written as

$$\mathcal{G}(E) = \int_{E_{\min}^{(+)}}^{\infty} dE'\ \frac{A(E')}{E-E'+i0} - (-1)^N \int_{E_{\min}^{(-)}}^{\infty} dE'\ \frac{B(E')}{E+E'-i0}, \tag{32}$$

where $E_{\min}^{(+)}$ is the minimal energy of states with electric charge eN and $E_{\min}^{(-)}$ is the minimal energy of states with electric charge $-eN$. Note that since we are working in Furry picture, the charge under considerations is provided only by electrons and positrons, because the nuclear charge is already taken into account as a source of the external field.

So far we considered $\mathcal{G}(E)$ for real E. However, for our purposes we are interested in the definition of $\mathcal{G}(E)$ in complex E plane. The mathematical technique, allowing to extend the domain of a given analytic function, is called *analytic continuation*. Let us first of all show that the Eq. (32) defines analytic function. This equation shows that the Green's function $\mathcal{G}(E)$ is the sum of Cauchy-type integrals. Since the integrals $\int_{E_{\min}^{(+)}}^{\infty} dE\ A(E)$ and $\int_{E_{\min}^{(-)}}^{\infty} dE\ B(E)$ converge (see Eqs. (28) and (29)), one can show with the help of standard mathematical methods that the equation

$$\mathcal{G}(E) = \int_{E_{\min}^{(+)}}^{\infty} dE'\ \frac{A(E')}{E-E'} - (-1)^N \int_{E_{\min}^{(-)}}^{\infty} dE'\ \frac{B(E')}{E+E'} \tag{33}$$

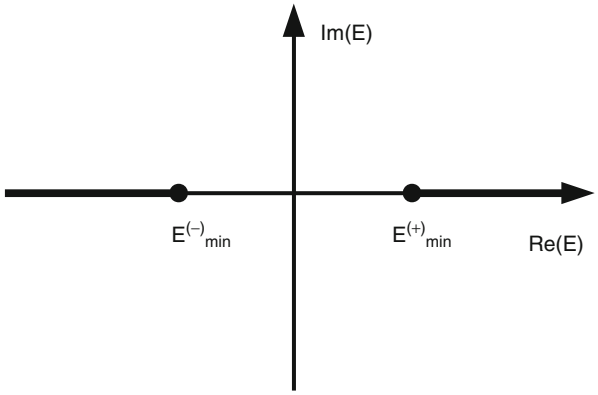


Fig. 1 Singularities of the two-time Green's function in complex E plane

defines an analytical function of E in the complex E plane with the cuts $(-\infty, E_{\rm min}^{(-)}]$ and $[E_{\rm min}^{(+)}, \infty)$ (see Fig. 1). This equation provides the analytical continuation of the Green's function to the complex E plane. According to (32), to get the Green's function for real E, we have to approach the right-hand cut from the upper half-plane and the left-hand cut from the lower half-plane.

In what follows we will be interested in bound states of the system. According to Eqs. (23), (24), (25), (26), (27), (28), (29), (30), (31), (32), and (33), the bound states correspond to the poles of the function $\mathcal{G}(E)$ on the right-hand real semiaxis. If the interaction between the electron–positron field and the electromagnetic field is switched off, the poles corresponding to bound states are isolated. Switching on the interaction between the fields transforms the isolated poles into branch points. This is caused by the fact that due to zero photon mass, the bound states are no longer isolated points of the spectrum, because zero-mass photon can carry arbitrary energy. The poles corresponding to the bound states lie on the upper boundary of the cut starting from the pole corresponding to the ground state. It is natural to assume that $\mathcal{G}(E)$ can be continued analytically under the cut, to the second sheet of the Riemann surface. As a result of this continuation the singularities of $\mathcal{G}(E)$ can be turned down.

In order to formulate the perturbation theory for calculations of the energy levels, we have to isolate the poles corresponding to the bound states from the related cuts. It can be done by introducing a nonzero photon mass μ which is generally assumed to be larger than the energy shift (or the energy splitting) of the level (levels) under consideration and much smaller than the distance to other levels. Later one can calculate $\lim\limits_{\mu\to 0}$ to obtain physical results. The singularities of $\mathcal{G}(E)$ with nonzero photon mass, including one- and two-photon spectra (i.e., states with one and two photons), are shown in Fig. 2. As one can see from this figure, introducing the photon mass makes the poles corresponding to the bound states to be isolated.

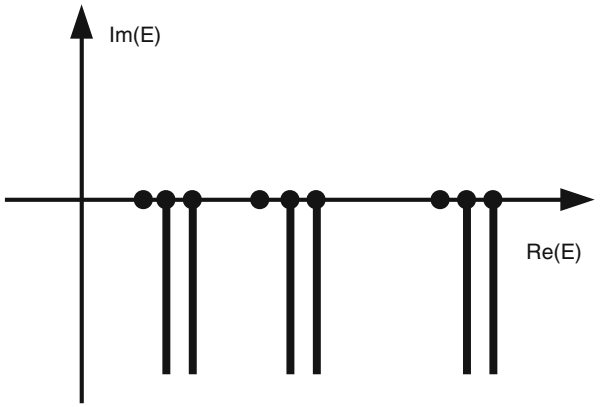


Fig. 2 Singularities of the two-time Green's function in complex E plane, taking into account nonzero photon mass

In every finite order of perturbation theory, the singularities of the Green's function $\mathcal{G}(E)$ in the complex E-plane are defined by the unperturbed Hamiltonian. In quantum mechanics this fact easily follows from the expansion of the Green's function $(E - H)^{-1} = (E - H_0 - \delta V)^{-1}$ in powers of the perturbation potential δV

$$(E - H)^{-1} = \sum_{n=0}^{\infty} (E - H_0)^{-1} \left[\delta V (E - H_0)^{-1}\right]^n . \tag{34}$$

As one can see from this equation, to n-th order of perturbation theory, the Green's function has poles of all orders till $n + 1$ at the unperturbed positions of the bound-state energies. This fact remains also valid in QED for $\mathcal{G}(E)$ defined above and allows us to evaluate the value of the bound-state energies in any order of perturbation theory.

Energy Shift of a Single Level

In previous section we have found the correspondence between the poles of the Green's function and the energy of the bound state. Let us evaluate in this section the energy shift $\Delta E_a = E_a - E_a^{(0)}$ of a single isolated level a of an N-electron atom due to the perturbative interaction. As we have agreed in the beginning of this chapter, we start in zeroth order from the noninteracting electrons described by the Dirac equation (1). So the unperturbed energy $E_a^{(0)}$ is equal to the sum of the one-electron Dirac energies:

$$E_a^{(0)} = \varepsilon_{a_1} + \cdots + \varepsilon_{a_N} , \tag{35}$$

which are defined by Eq. (1).

To construct the zeroth-order wave function in the simplest case, one has to consider the Slater determinant, constructed from the solutions of the Eq. (1):

$$u_a(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) = \frac{1}{\sqrt{N!}}\det\begin{vmatrix}\psi_1(\boldsymbol{x}_1) & \cdots & \psi_N(\boldsymbol{x}_1)\\ \cdots & \cdots & \cdots\\ \psi_1(\boldsymbol{x}_N) & \cdots & \psi_N(\boldsymbol{x}_N)\end{vmatrix}$$
$$= \frac{1}{\sqrt{N!}}\sum_P (-1)^P \psi_{Pa_1}(\boldsymbol{x}_1)\cdots\psi_{Pa_N}(\boldsymbol{x}_N), \tag{36}$$

where ψ_n are the one-electron Dirac wave functions defined by Eq. (1) and P is the permutation operator with the parity $(-1)^P$. In the most general case, the unperturbed wave function represents a linear combination of the one-determinant functions:

$$u_a(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N) = \sum_b C_a^b \frac{1}{\sqrt{N!}}\sum_P (-1)^P \psi_{Pb_1}(\boldsymbol{x}_1)\cdots\psi_{Pb_N}(\boldsymbol{x}_N). \tag{37}$$

In order to evaluate the energy shift, let us introduce the Green's function $g_{aa}(E)$ as

$$g_{aa}(E) = \langle u_a|\mathcal{G}(E)\gamma_1^0\cdots\gamma_N^0|u_a\rangle$$
$$\equiv \int d\boldsymbol{x}_1\cdots d\boldsymbol{x}_N\, d\boldsymbol{x}'_1\cdots d\boldsymbol{x}'_N\; u_a^\dagger(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_N)$$
$$\times\mathcal{G}(E,\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_N;\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N)\gamma_1^0\cdots\gamma_N^0 u_a(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N). \tag{38}$$

From the spectral representation for $\mathcal{G}(E)$ (see Eqs. (23), (24), (25), (26), (27), (28), (29), (30), (31), (32), and (33)), we obtain

$$g_{aa}(E) = \frac{A_a}{E-E_a} + \text{terms that are regular at } E\sim E_a, \tag{39}$$

where

$$A_a = \frac{1}{N!}\int d\boldsymbol{x}_1\cdots d\boldsymbol{x}_N\, d\boldsymbol{x}'_1\cdots d\boldsymbol{x}'_N\, u_a^\dagger(\boldsymbol{x}'_1,\ldots,\boldsymbol{x}'_N)\langle 0|\psi(0,\boldsymbol{x}'_1)\cdots\psi(0,\boldsymbol{x}'_N)|a\rangle$$
$$\times\langle a|\psi^\dagger(0,\boldsymbol{x}_N)\cdots\psi^\dagger(0,\boldsymbol{x}_1)|0\rangle u_a(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_N). \tag{40}$$

As discussed in the previous section, we assume here that a nonzero photon mass μ is introduced to isolate the pole corresponding to the bound state a from the related cut. To generate the perturbation series for E_a, it is convenient to use a contour integral formalism developed first in operator theory by Szökefalvi-Nagy and Kato [14], Kato [15, 16], Messiah [17], and Lepage [18]. Choosing a contour

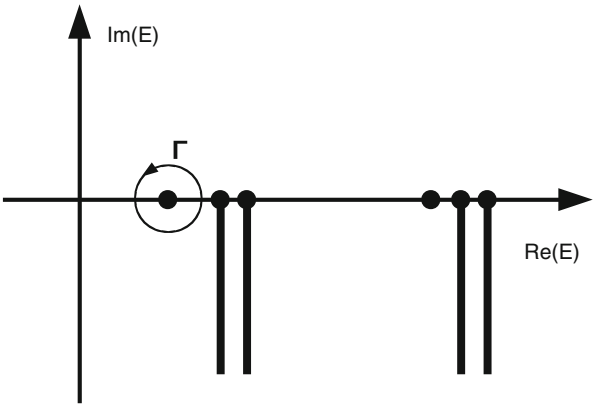


Fig. 3 Integration contour in complex E plane

Γ in the complex E plane in a way that it surrounds the pole corresponding to the level a and keeps outside all other singularities (see Fig. 3), we have

$$\frac{1}{2\pi i}\oint_{\Gamma} dE\ E g_{aa}(E) = E_a A_a, \tag{41}$$

$$\frac{1}{2\pi i}\oint_{\Gamma} dE\ g_{aa}(E) = A_a. \tag{42}$$

Here we have assumed that the contour Γ is oriented counterclockwise.

Dividing Eq. (41) by (42), we obtain

$$E_a = \frac{\frac{1}{2\pi i}\oint_{\Gamma} dE\ E g_{aa}(E)}{\frac{1}{2\pi i}\oint_{\Gamma} dE\ g_{aa}(E)} \tag{43}$$

It is convenient to transform Eq. (43) to a form that directly yields the energy shift $\Delta E_a = E_a - E_a^{(0)}$. To this end we need to evaluate the zeroth order of the Green's function (38) and extract it from the total function. Substituting the operators

$$\psi_{\text{in}}(0, \boldsymbol{x}) = \sum_{\varepsilon_n > 0} b_n \psi_n(\boldsymbol{x}) + \sum_{\varepsilon_n < 0} d_n^{\dagger} \psi_n(\boldsymbol{x}), \tag{44}$$

$$\overline{\psi}_{\text{in}}(0, \boldsymbol{x}) = \sum_{\varepsilon_n > 0} b_n^{\dagger} \overline{\psi}_n(\boldsymbol{x}) + \sum_{\varepsilon_n < 0} d_n \overline{\psi}_n(\boldsymbol{x}) \tag{45}$$

into Eqs. (24) and (25) instead of $\psi(0, \boldsymbol{x})$ and $\overline{\psi}(0, \boldsymbol{x})$, respectively, and considering the states $|n\rangle$ in (24) and (25) as unperturbed states in the Fock space, one can find

that every matrix element in the numerators of the fractions in (24) and (25) turns to unity due to the ortho-normalization of the states. Then summation over all possible permutations cancels the factor $1/N!$. Finally, selecting from entire sum only the term proportional to $1/(E - E_a^{(0)})$, we find

$$g_{aa}^{(0)} = \frac{1}{E - E_a^{(0)}} \,. \tag{46}$$

Defining $\Delta g_{aa} = g_{aa} - g_{aa}^{(0)}$, from (43) we obtain the formula for the energy shift [19]:

$$\Delta E_a = \frac{\dfrac{1}{2\pi i} \oint_\Gamma dE \, \left(E - E_a^{(0)}\right) \Delta g_{aa}(E)}{1 + \dfrac{1}{2\pi i} \oint_\Gamma dE \, \Delta g_{aa}(E)} \,. \tag{47}$$

The Green's function $\Delta g_{aa}(E)$ is constructed by perturbation theory, order by order:

$$\Delta g_{aa}(E) = \Delta g_{aa}^{(1)}(E) + \Delta g_{aa}^{(2)}(E) + \cdots \,, \tag{48}$$

where the superscript denotes the order in α (interaction constant). If we represent the energy shift as a series in α

$$\Delta E_a = \Delta E_a^{(1)} + \Delta E_a^{(2)} + \cdots \,, \tag{49}$$

the formula (47) yields for two lowest orders

$$\Delta E_a^{(1)} = \frac{1}{2\pi i} \oint_\Gamma dE \, \Delta E \, \Delta g_{aa}^{(1)}(E), \tag{50}$$

$$\Delta E_a^{(2)} = \frac{1}{2\pi i} \oint_\Gamma dE \, \Delta E \, \Delta g_{aa}^{(2)}(E) - \left(\frac{1}{2\pi i} \oint_\Gamma dE \, \Delta E \, \Delta g_{aa}^{(1)}(E) \right) \left(\frac{1}{2\pi i} \oint_\Gamma dE \, \Delta g_{aa}^{(1)}(E) \right), \tag{51}$$

where $\Delta E \equiv E - E_a^{(0)}$. One can of course continue this derivation for the higher orders of perturbation theory. However, the calculations in Furry picture require so intensive numerical calculations that modern computers allow the evaluation of only two lowest orders of QED corrections. Moreover, the calculation of the one-electron second-order QED corrections to all orders in αZ in the simplest case of the hydrogen-like ion is not completed yet. See the ▸ Chap. 6, "QED Effects and Challenges" of this book for the details. Therefore we also restrict the derivation by the second order.

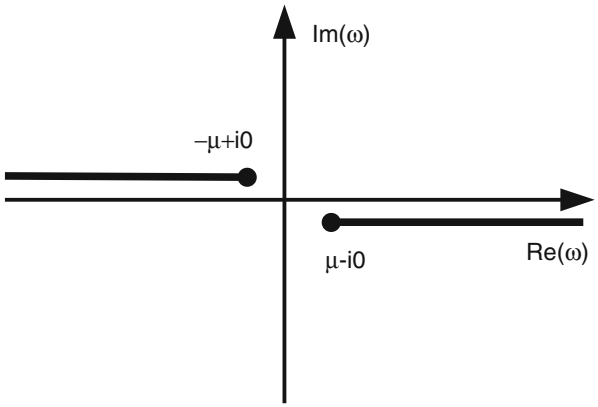


Fig. 4 Singularities of the photon propagator with nonzero mass μ

In the previous section, deriving Eqs. (43) and (47), we assumed a nonzero photon mass μ. It allows to move all the cuts outside the contour Γ (see Fig. 3) and also regularizes the infrared singularities of individual diagrams. In the Feynman gauge, the photon propagator with nonzero photon mass μ is

$$D_{\mu\nu}(\omega, \boldsymbol{x}-\boldsymbol{y}) = -g_{\mu\nu}\int \frac{d\mathbf{k}}{(2\pi)^3}\,\frac{\exp\left(i\mathbf{k}\cdot(\boldsymbol{x}-\boldsymbol{y})\right)}{\omega^2-\mathbf{k}^2-\mu^2+i0} \tag{52}$$

or, after integration,

$$D_{\mu\nu}(\omega, \boldsymbol{x}-\boldsymbol{y}) = g_{\mu\nu}\frac{\exp\left(i\sqrt{\omega^2-\mu^2+i0}\,|\boldsymbol{x}-\boldsymbol{y}|\right)}{4\pi|\boldsymbol{x}-\boldsymbol{y}|}, \tag{53}$$

where $\mathrm{Im}\sqrt{\omega^2-\mu^2+i0} > 0$. $D_{\mu\nu}(\omega, \boldsymbol{x}-\boldsymbol{y})$ is an analytical function of ω in the complex ω plane with cuts beginning at the points $\omega = -\mu + i0$ and $\omega = \mu - i0$ (see Fig. 4). The related expressions for the photon propagator with nonzero photon mass in other covariant gauges are presented, e.g., in [12].

We know from the previous section that the singularities of the two-time Green's function in the complex E plane are defined by the unperturbed Hamiltonian if it is constructed by perturbation theory. In particular, it means that in n-th order of perturbation theory, $g_{aa}(E)$ has poles of all orders till $n+1$ at the position of the unperturbed energy level under consideration. Therefore, in calculations by perturbation theory, it is sufficient to consider the photon mass as a very small parameter which provides a separation of the pole from the related cut. At the end of the calculations after taking into account a whole gauge invariant set of Feynman diagrams, we can put $\mu \to 0$. The possibility of taking the limit $\mu \to 0$ follows, in particular, from the fact that the contour Γ can be shrunk continuously to the point $E = E_a^{(0)}$ (see Fig. 3).

Summarizing the result of the previous sections, we have derived the recipe of evaluation of the energy shift of the bound level from the basic principles of QED. To do so, one has to

- Define which order of the perturbation theory and what kind of diagrams are considered;
- Construct the part of the 2N-time Green's function, corresponding to the selected diagrams, using Feynman rules from the section "$2N$-Times Green's Function";
- Evaluate the two-time Green's function and the energy shift using Eqs. (47), (50), and (51).

Some useful for practical calculations formulas, simplifying this procedure in the simplest case of isolated level, can be found in Ref. [7]. In the next section, we will demonstrate how to apply this method to the nontrivial case of quasi-degenerate levels of two-electron ions.

Quasi-degenerate Levels

In the previous sections, we have discussed the TTGF method. In this section we are going to demonstrate how to use this method for the case of quasi-degenerate states of helium-like ions. In particular we are interested in the $(1s2p_{1/2})_1$ and $(1s2p_{3/2})_1$ states, whose symmetry and total angular momentum are the same, and the unperturbed zeroth-order energies are separated only by fine splitting. We select this example because, on the one hand, it is a nontrivial case where all existed before QED methods failed and, on the other hand, the two-electron case is rather simple and allows to write all the formulas explicitly.

Let us start from the unperturbed wave functions. We define them as

$$u_1 = \sum_{m_a m_v} \langle j_a m_a j_v m_v | JM \rangle \frac{1}{\sqrt{2}} \sum_P (-1)^P |PaPv\rangle, \tag{54}$$

$$u_2 = \sum_{m_b m_w} \langle j_b m_b j_w m_w | JM \rangle \frac{1}{\sqrt{2}} \sum_P (-1)^P |PbPw\rangle, \tag{55}$$

where a and b are taken to represent $1s$ orbitals while v and w are taken to represent $2p_{1/2}$ and $2p_{3/2}$ orbitals, respectively; P is the permutation operator:

$$\sum_P (-1)^P |PaPv\rangle = |av\rangle - |va\rangle,$$

$|av\rangle \equiv |a\rangle|v\rangle$ is the product of the one-electron Dirac wave functions. We use here the jj coupling which is natural for high-Z few-electron atoms. The transition to the

wave functions corresponding to the LS-coupling scheme can be done using the following formula [20]:

$$\begin{pmatrix} |1s2p\ ^3P_1\rangle \\ |1s2p\ ^1P_1\rangle \end{pmatrix} = R \begin{pmatrix} |(1s2p_{\frac{1}{2}})_1\rangle \\ |(1s2p_{\frac{3}{2}})_1\rangle \end{pmatrix}, \tag{56}$$

where

$$R = \frac{1}{\sqrt{3}} \begin{pmatrix} \sqrt{2} & -1 \\ 1 & \sqrt{2} \end{pmatrix}. \tag{57}$$

According to the algorithm, given in the previous section, now we should define the four-time Green's function. The standard definition of this function in the external field of the nucleus is (compare with Eq. (3))

$$G(x_1', x_2'; x_1, x_2) = \langle 0|T\psi(x_1')\psi(x_2')\bar{\psi}(x_1)\bar{\psi}(x_2)|0\rangle. \tag{58}$$

This Green's function is constructed by perturbation theory after the transition to the interaction representation where it is given by (see Eq. (4))

$$G(x_1', x_2'; x_1, x_2) = \frac{\langle 0|T\psi_{\rm in}(x_1')\psi_{\rm in}(x_2')\overline{\psi}_{\rm in}(x_2)\overline{\psi}_{\rm in}(x_1)\exp\left[i\int d^4z\mathcal{H}_{\rm int}(z)\right]|0\rangle}{\langle 0|T\exp\left[i\int d^4z\mathcal{H}_{\rm int}(z)\right]|0\rangle} \tag{59}$$

This equation allows to construct G using Wick's theorem.

It is however more convenient to work with the Green's function in the mixed energy-coordinate representation, which is defined by

$$\begin{aligned} G(p_1'^0, \boldsymbol{x}_1', p_2'^0, \boldsymbol{x}_2'; p_1^0, \boldsymbol{x}_1, p_2^0, \boldsymbol{x}_2) = \frac{1}{(2\pi)^4}\int_{-\infty}^{\infty} dx_1^0\, dx_2^0\, dx_1'^0\, dx_2'^0 \\ \times\exp\left(ip_1'^0x_1'^0 + ip_2'^0x_2'^0 - ip_1^0x_1^0 - ip_2^0x_2^0\right) \\ \times G(x_1', x_2'; x_1, x_2). \end{aligned} \tag{60}$$

The Feynman rules for $G(p_1'^0, \boldsymbol{x}_1', p_2'^0, \boldsymbol{x}_2'; p_1^0, \boldsymbol{x}_1, p_2^0, \boldsymbol{x}_2)$ are given in section "$2N$-Times Green's Function".

Now we introduce the Green's function $g(E)$ as

$$\begin{aligned} g(E)\,\delta(E-E') = \frac{\pi}{i}\int_{-\infty}^{\infty} dp_1^0\, dp_2^0\, dp_1'^0\, dp_2'^0\,\delta\left(E - p_1^0 - p_2^0\right) \\ \times\delta\left(E' - p_1'^0 - p_2'^0\right)\, P_0\, G\left(p_1'^0, p_2'^0; p_1^0, p_2^0\right)\, \gamma_1^0\gamma_2^0\, P_0, \end{aligned} \tag{61}$$

where $P_0 = \sum_k u_k u_k^\dagger$ is the projector on the subspace of the unperturbed quasi-degenerate states under consideration. The main difference of this equation from Eq. (38) is the space of unperturbed wave functions. While in the case of isolated level the basis of this space is given by the single unperturbed function $|u_a\rangle$, in quasi-degenerate case one should consider the two-(and possibly more-)dimensional subspace.

It can be derived (see Ref. [7, 21] for details) that the system under consideration can be described by the set of two-dimensional Schrödinger-like equation ($k = 1, 2$),

$$H\psi_k = E_k \psi_k, \tag{62}$$

where

$$H = P^{-1/2} K P^{-1/2}, \tag{63}$$

$$K = \frac{1}{2\pi i} \oint_\Gamma dE\, E\, g(E), \tag{64}$$

$$P = \frac{1}{2\pi i} \oint_\Gamma dE\, g(E), \tag{65}$$

Γ is a contour in the complex E plane that surrounds the levels under consideration but does not encircles other levels, and E_k are the exact energies of the states under consideration. It is assumed that the contour Γ is oriented anticlockwise. The operator H in Eq. (62) is constructed by perturbation theory in α. Substituting

$$g(E) = g^{(0)}(E) + g^{(1)}(E) + g^{(2)}(E) + \cdots, \tag{66}$$

$$P = P^{(0)} + P^{(1)} + P^{(2)} + \cdots, \tag{67}$$

$$K = K^{(0)} + K^{(1)} + K^{(2)} + \cdots, \tag{68}$$

where the index indicates the order in α, we obtain

$$H^{(0)} = K^{(0)}, \tag{69}$$

$$H^{(1)} = K^{(1)} - \frac{1}{2} P^{(1)} K^{(0)} - \frac{1}{2} K^{(0)} P^{(1)}, \tag{70}$$

$$\begin{aligned} H^{(2)} = {} & K^{(2)} - \frac{1}{2} P^{(2)} K^{(0)} - \frac{1}{2} K^{(0)} P^{(2)} - \frac{1}{2} P^{(1)} K^{(1)} - \frac{1}{2} K^{(1)} P^{(1)} \\ & + \frac{3}{8} P^{(1)} P^{(1)} K^{(0)} + \frac{3}{8} K^{(0)} P^{(1)} P^{(1)} + \frac{1}{4} P^{(1)} K^{(0)} P^{(1)}. \end{aligned} \tag{71}$$

The solvability of Eq. (62) yields the basic equation for the calculation of the energy levels

$$\det(E - H) = 0. \tag{72}$$

As was noticed in Ref. [22], due to nonzero decay rates of excited states, the matrix H becomes complex, and the self-adjoint part of H should be understood in Eqs. (62) and (72):

$$H \equiv (1/2)(H + H^{\dagger}). \tag{73}$$

To the zeroth order in α, we have

$$g^{(0)}(E) = \sum_{s=1}^{2} \frac{|u_s\rangle\langle u_s|}{E - E_s^{(0)}}, \tag{74}$$

where $E_1^{(0)}$ and $E_2^{(0)}$ are the unperturbed energies of the $(1s2p_{1/2})_1$ and $(1s2p_{3/2})_1$ states, respectively. They are given by the sum of the one-electron Dirac–Coulomb energies:

$$E_1^{(0)} = \varepsilon_{1s} + \varepsilon_{2p_{1/2}}, \qquad E_2^{(0)} = \varepsilon_{1s} + \varepsilon_{2p_{3/2}}.$$

Substituting Eq. (74) into the definitions of K, P, and H, we find

$$K_{ik}^{(0)} = E_i^{(0)}\delta_{ik}, \tag{75}$$

$$P_{ik}^{(0)} = \delta_{ik}, \tag{76}$$

$$H_{ik}^{(0)} = E_i^{(0)}\delta_{ik}, \tag{77}$$

where $K_{ik} = \langle i|K|k\rangle$, $P_{ik} = \langle i|P|k\rangle$, and $H_{ik} = \langle i|H|k\rangle$.

In order to illustrate the TTGF method, we present below the detailed derivation of the correction to the quasi-degenerate energy levels $(1s2p_{1/2})_1$ and $(1s2p_{3/2})_1$ due to the one-photon exchange diagram (Fig. 5). On the one hand, this derivation is much less cumbersome than that for the second-order corrections. On the other hand, it demonstrates most essential features of the TTGF method applied to quasi-degenerate case. For simplicity, in the derivation below, we will assume that the unperturbed energy of the initial state i differs from that of the final state k: $E_i^{(0)} \neq E_k^{(0)}$ (in the case under consideration, it corresponds to $i \neq k$). However, all the final formulas can be shown to be valid also for the case $E_i^{(0)} = E_k^{(0)}$.

For the sake of brevity, we will use the short-hand notation for the summation over the Clebsch–Gordan coefficients in Eqs. (54) and (55):

$$F_i\,|i_1 i_2\rangle \equiv \sum_{m_{i_1} m_{i_2}} \langle j_{i_1} m_{i_1} j_{i_2} m_{i_2}|JM\rangle\,|i_1 i_2\rangle. \tag{78}$$

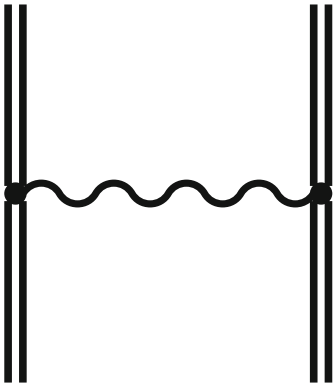


Fig. 5 One-photon exchange diagram

where $|i_1 i_2\rangle$ is either $|av\rangle$ or $|aw\rangle$. It is convenient also to use the notation for the operator of the electron-electron interaction:

$$I(\omega) = e^2 \sum_{\mu\nu} \alpha_1^\mu \, \alpha_2^\nu \, D_{\mu\nu}(\omega), \tag{79}$$

where $\alpha^\mu = \gamma^0 \gamma^\mu = (1, \boldsymbol{\alpha})$, and $D_{\mu\nu}$ denotes the photon propagator. For the matrix elements of the operator $I(\omega)$, we will use the short-hand notation

$$I_{ijkl}(\omega) = \langle ij \,|\, I(\omega) | kl \rangle. \tag{80}$$

According to the Feynman rules and the definition of $g(E)$, the contribution of the one-photon exchange diagram to the matrix elements $g_{ik}(E) = \langle i \,|\, g(E) | k \rangle$ is

$$\begin{aligned} g_{ik}^{(1)}(E) = F_i F_k \Big(\frac{i}{2\pi}\Big)^2 \int_{-\infty}^{\infty} dp_1^0 \, dp_1'^0 \sum_P (-1)^P \\ \times \frac{1}{(p_1'^0 - \varepsilon_{Pi_1} + i0)(E - p_1'^0 - \varepsilon_{Pi_2} + i0)} \\ \times \frac{I_{Pi_1 Pi_2 k_1 k_2}(p_1'^0 - p_1^0)}{(p_1^0 - \varepsilon_{k_1} + i0)(E - p_1^0 - \varepsilon_{k_2} + i0)} \, . \end{aligned} \tag{81}$$

In order to evaluate the contour integrals in Eqs. (64) and (65), we must rewrite the integrand in Eq. (81), separating the factors $1/(E - E_i^{(0)})$ and $1/(E - E_k^{(0)})$. To this end let us use the identities

$$\frac{1}{(p_1'^0 - \varepsilon_{Pi_1} + i0)(E - p_1'^0 - \varepsilon_{Pi_2} + i0)} = \frac{1}{E - E_i^{(0)}} \left(\frac{1}{p_1'^0 - \varepsilon_{Pi_1} + i0} + \frac{1}{E - p_1'^0 - \varepsilon_{Pi_2} + i0} \right), \tag{82}$$

$$\frac{1}{(p_1^0 - \varepsilon_{k_1} + i0)(E - p_1^0 - \varepsilon_{k_2} + i0)} = \frac{1}{E - E_k^{(0)}} \left(\frac{1}{p_1^0 - \varepsilon_{k_1} + i0} + \frac{1}{E - p_1^0 - \varepsilon_{k_2} + i0} \right). \tag{83}$$

With their help we obtain

$$K_{ik}^{(1)} = F_i F_k \frac{1}{2\pi i} \oint_\Gamma dE \frac{E}{(E - E_i^{(0)})(E - E_k^{(0)})} \Bigg[\left(\frac{i}{2\pi}\right)^2 \int_{-\infty}^{\infty} dp_1^0 \, dp_1'^0 \sum_P (-1)^P \times \left(\frac{1}{p_1'^0 - \varepsilon_{Pi_1} + i0} + \frac{1}{E - p_1'^0 - \varepsilon_{Pi_2} + i0} \right) \left(\frac{1}{p_1^0 - \varepsilon_{k_1} + i0} + \frac{1}{E - p_1^0 - \varepsilon_{k_2} + i0} \right) \times I_{Pi_1 Pi_2 k_1 k_2}(p_1'^0 - p_1^0) \Bigg]. \tag{84}$$

The expression in the square brackets is an analytical function of E inside the contour Γ, if the photon mass μ is chosen properly (see Refs. [21, 22]). Carrying out the E integration by Cauchy's theorem and taking into account that

$$\left(\frac{i}{2\pi}\right) \left(\frac{1}{x + i0} + \frac{1}{-x + i0} \right) = \delta(x), \tag{85}$$

we obtain

$$\begin{aligned} K_{ik}^{(1)} = F_i F_k \Bigg\{ & \frac{i}{2\pi} \int_{-\infty}^{\infty} dp_1^0 \sum_P (-1)^P \frac{E_i^{(0)} I_{Pi_1 Pi_2 k_1 k_2}(\varepsilon_{Pi_1} - p_1^0)}{E_i^{(0)} - E_k^{(0)}} \\ & \times \left(\frac{1}{p_1^0 - \varepsilon_{k_1} + i0} + \frac{1}{E_i^{(0)} - p_1^0 - \varepsilon_{k_2} + i0} \right) \\ & + \frac{i}{2\pi} \int_{-\infty}^{\infty} dp_1'^0 \sum_P (-1)^P \frac{E_k^{(0)} I_{Pi_1 Pi_2 k_1 k_2}(p_1'^0 - \varepsilon_{k_1})}{E_k^{(0)} - E_i^{(0)}} \\ & \times \left(\frac{1}{p_1'^0 - \varepsilon_{Pi_1} + i0} + \frac{1}{E_k^{(0)} - p_1'^0 - \varepsilon_{Pi_2} + i0} \right) \Bigg\} . \end{aligned} \tag{86}$$

In the same way, we find

$$\begin{aligned} P_{ik}^{(1)} = F_i F_k \Bigg\{ & \frac{i}{2\pi} \int_{-\infty}^{\infty} dp_1^0 \sum_P (-1)^P \frac{I_{Pi_1 Pi_2 k_1 k_2}(\varepsilon_{Pi_1} - p_1^0)}{E_i^{(0)} - E_k^{(0)}} \\ & \times \left(\frac{1}{p_1^0 - \varepsilon_{k_1} + i0} + \frac{1}{E_i^{(0)} - p_1^0 - \varepsilon_{k_2} + i0} \right) \\ & + \frac{i}{2\pi} \int_{-\infty}^{\infty} dp_1'^0 \sum_P (-1)^P \frac{I_{Pi_1 Pi_2 k_1 k_2}(p_1'^0 - \varepsilon_{k_1})}{E_k^{(0)} - E_i^{(0)}} \\ & \times \left(\frac{1}{p_1'^0 - \varepsilon_{Pi_1} + i0} + \frac{1}{E_k^{(0)} - p_1'^0 - \varepsilon_{Pi_2} + i0} \right) \Bigg\} . \end{aligned} \tag{87}$$

Substituting Eqs. (86) and (87) into Eq. (70), we get

$$H_{ik}^{(1)} = F_i F_k \left\{ \frac{i}{4\pi} \int_{-\infty}^{\infty} dp_1^0 \sum_P (-1)^P I_{P i_1 P i_2 k_1 k_2}(\varepsilon_{P i_1} - p_1^0) \right.$$
$$\times \left(\frac{1}{p_1^0 - \varepsilon_{k_1} + i0} + \frac{1}{E_i^{(0)} - p_1^0 - \varepsilon_{k_2} + i0} \right)$$
$$+ \frac{i}{4\pi} \int_{-\infty}^{\infty} dp_1'^0 \sum_P (-1)^P I_{P i_1 P i_2 k_1 k_2}(p_1'^0 - \varepsilon_{k_1})$$
$$\left. \times \left(\frac{1}{p_1'^0 - \varepsilon_{P i_1} + i0} + \frac{1}{E_k^{(0)} - p_1'^0 - \varepsilon_{P i_2} + i0} \right) \right\} . \tag{88}$$

Introducing the notations $\Delta_1 = \varepsilon_{P i_1} - \varepsilon_{k_1}$ and $\Delta_2 = \varepsilon_{P i_2} - \varepsilon_{k_2}$, we can rewrite Eq. (88) as follows:

$$H_{ik}^{(1)} = F_i F_k \frac{i}{8\pi} \int_{-\infty}^{\infty} d\omega \sum_P (-1)^P I_{P i_1 P i_2 k_1 k_2}(\omega) \left(\frac{1}{\omega + \Delta_1 + i0} + \frac{1}{\Delta_2 - \omega + i0} \right.$$
$$+ \frac{1}{\omega + \Delta_2 + i0} + \frac{1}{\Delta_1 - \omega + i0} + \frac{1}{\omega - \Delta_1 + i0} + \frac{1}{-\Delta_2 - \omega + i0}$$
$$\left. + \frac{1}{\omega - \Delta_2 + i0} + \frac{1}{-\Delta_1 - \omega + i0} \right)$$
$$= F_i F_k \frac{1}{4} \int_{-\infty}^{\infty} d\omega \sum_P (-1)^P I_{P i_1 P i_2 k_1 k_2}(\omega)$$
$$\times \Big[\delta(\omega + \Delta_1) + \delta(\omega - \Delta_1) + \delta(\omega + \Delta_2) + \delta(\omega - \Delta_2) \Big] . \tag{89}$$

Taking into account that $I(z) = I(-z)$, we finally obtain [22, 23]

$$H_{ik}^{(1)} = F_i F_k \frac{1}{2} \sum_P (-1)^P \left[I_{P i_1 P i_2 k_1 k_2}(\Delta_1) + I_{P i_1 P i_2 k_1 k_2}(\Delta_2) \right] . \tag{90}$$

The derivation of the second-order diagrams is rather lengthy and cannot be discussed within a single chapter of the book. Only publication of final formulas in this case would occupy more than two pages. Those who are interested in further application of this technique to the quasi-degenerate levels can find in Ref. [24] the detailed derivation and evaluation of all low-lying levels in helium-like ions. The evaluations are performed for the wide range of the nuclear charges $12 \le Z \le 100$ up to second order of the perturbation theory. The lowest uncalculated terms have the orders $\alpha^3(\alpha Z)^2$ and $\alpha^2(\alpha Z)^7$. To date, it is the most accurate evaluation of the energy levels of two-electron ions.

Summary

In summary we have presented in this chapter the two-time Green's function method – the powerful method allowing to derive the computational formulas from the basic principles of QED. We have discussed the main features of the method and demonstrated its application to the evaluation of the one-photon exchange diagram contribution to the energy shift of the quasi-degenerate levels.

Acknowledgements Stimulating discussions with Prof. V. M. Shabaev and Prof. P. Indelicato are gratefully acknowledged.

References

1. Dirac PAM (1928) The quantum theory of the electron. Proc R Soc Lond Ser A 117(778):610
2. Feynman RP (1949) Space-time approach to quantum electrodynamics. Phys Rev 76:769
3. Gell-Mann M, Low F (1951) Bound states in quantum field theory. Phys Rev 84(2):350
4. Sucher J (1957) S-Matrix formalism for level-shift calculations. Phys Rev 107(5):1448
5. Braun MA, Gurchumelia AD (1980) Relativistic adiabatic perturbation theory for degenerate levels. Theor Math Phys 45:199
6. Braun MA, Gurchumelia AD, Safronova UI (1984) Relativistic thory of atom. Nauka, Moscow
7. Shabaev VM (2002) Two-time Green's function method in quantum electrodynamics of high-Z few-electron atoms. Phys Rep 356(3):119
8. Lindgren I, Salomonson S, Åsen B (2004) The covariant-evolution-operator method in bound-state QED. Phys Rep 389(4):161
9. Andreev OY, Labzowsky LN, Plunien G, Solovyev DA (2008) QED theory of the spectral line profile and its applications to atoms and ions. Phys Rep 455:135
10. Artemyev AN, Shabaev VM, Tupitsyn II, Plunien G, Yerokhin VA (2007) QED Calculation of the $2p_3/2 - 2p_1/2$ transition energy in boronlike argon. Phys Rev Lett 98(17):173004
11. Artemyev AN, Shabaev VM, Tupitsyn II, Plunien G, Surzhykov A, Fritzsche S (2013) Ab initio calculations of the $2p_3/2 - 2p_1/2$ fine-structure splitting in boronlike ions. Phys Rev A 88:032518. 10.1103/PhysRevA.88.032518, http://link.aps.org/doi/10.1103/PhysRevA.88.032518
12. Itzykson C, Bernard Zuber J (1980) Quantum field theory. McGraw-Hill, New York
13. Bjorken JD, Drell SD (1964) Relativistic quantum mechanics. McGraw-Hill, New York
14. Szökefalvi-Nagy B (1946/1947) Perturbations des transformations autoadjointes dans l'espace de Hilbert. Comment Math Helv 19:347 (in French)
15. Kato T (1949) On the convergence of the perturbation method. Prog Theor Phys 4:514
16. Kato T (1966) Perturbation theory for linear operators. Die Grundlagen der mathematischen Wissenschaften in Einzeldarstellungen, vol 132. Springer, New York
17. Messiah A (1961) Quantum mechanics, vol 2. North-Holland, Amsterdam
18. Lepage GP (1977) Analytic bound-state solutions in a relativistic two-body formalism with applications in muonium and positronium. Phys Rev A 16(3):863
19. Shabaev VM (1988) In: Safronova UI (ed) Many-particle effects in atoms. AN SSSR, Nauchnyi Sovet po Spektroskopii, Moscow), p 15
20. Drake GWF (1988) Theoretical energies for the n = 1 and 2 states of the helium isoelectronic sequence up to Z = 100. Can J Phys 66:586
21. Shabaev VM (1994) Quantum electrodynamic theory of recombination of an electron with a highly charged ion. Phys Rev A 50(6):4521

22. Shabaev VM (1993) Schrödinger-like equation for the relativistic few-electron atom. J Phys B 26:4703
23. Mittleman MH (1972) Configuration-space Hamiltonian for heavy atoms and correction to the Breit interaction. Phys Rev A 5:2395
24. Artemyev AN, Shabaev VM, Yerokhin VA, Plunien G, Soff G (2005) QED calculations of the $n = 1$ and $n = 2$ energy levels in He-like ions. Phys Rev A 71:062104

Unifying Many-Body Perturbation Theory with Quantum Electrodynamics

9

Ingvar Lindgren and Paul Indelicato

Contents

Introduction . . . 314
Standard Perturbation Theory . . . 316
All-Order Methods: Coupled-Cluster Approach . . . 320
Multi-configuration Hartree/Dirac-Fock (MCHF/MCDF) . . . 322
Relativistic MBPT . . . 322
Fock Space . . . 323
QED Effects . . . 324
Full QED Treatment . . . 324
QED Methods . . . 324
The Covariant Evolution Operator . . . 325
Energy-Dependent Perturbation Theory: The Green's Operator . . . 327
Model-Space Contributions . . . 328
Generalized Bloch Equation . . . 332
Combining QED with Electron Correlation . . . 333
General . . . 333
Retardation and Virtual Pairs . . . 335
Self-Energy and Vertex Correction . . . 336
Summary . . . 338
References . . . 339

I. Lindgren (✉)
Department of Physics, University of Gothenburg, Gothenburg, Sweden
e-mail: ingvar.lindgren@physics.gu.se

P. Indelicato
Laboratorie Kastler Brossel, UPMC-Sorbonne Universités, Collège de France and CNRS, Paris, France

PSL Research University, Paris, France
e-mail: paul.indelicato@lkb.ens.fr

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_29

Abstract

A unified procedure for many-body perturbation theory and quantum electrodynamics has been constructed by the Gothenburg group, based upon the recently developed covariant-evolution-operator method. This is a form of time- or energy-dependent perturbation theory, where all perturbations, including relativity and quantum electrodynamics, are built into the wave function, which is a necessary requisite for a true unification. The procedure is based upon the use of the Coulomb gauge, where the dominating part of the electron-electron interaction is expressed by means of the energy-independent Coulomb interaction and only the weaker energy-dependent transverse part by means of the covariant field theoretical expression. This leads to a more effective way of treating quantum electrodynamics in combination with electron correlation than the traditional purely quantum electrodynamical procedures. This will make it possible to go beyond two-photon exchange, which until now has been possible only for the lightest elements. The procedure has been implemented on highly charged helium-like ions.

The procedure was primarily developed for static properties, but it is also applicable to dynamical processes, like scattering processes and electronic transition rates, as has recently been demonstrated.

Keywords

Covariant evolution operator • Electron correlation • Electron self energy • Green's operator • Model-space contribution • Perturbation theory • Quasi-degeneracy • Quantum electrodynamics • Relativity • Vacuum polarization • Vertex correction

Introduction

Many-body perturbation theory (MBPT) and its all-order variants, particularly various coupled-cluster approaches (CCAs), have been quite successful in describing the correlational effects of quantum chemical systems to a high degree of accuracy (see, for instance, Ref. [1]). Also relativity can here be handled in the standard "no-virtual-pair" approximation. On the other hand, effects of quantum electrodynamics (QED) beyond that approximation cannot be included – at least not in any systematic fashion – effects that become increasingly important as the demand for higher accuracy becomes more pronounced.

One systematic procedure is, of course, to perform a full QED calculation and to treat all effects, including electron correlation, in terms of fully covariant multi-photon interactions. This is, however, an extremely inefficient way of treating electron correlation in cases where this is important, and more efficient approaches have to be found.

Several procedures have been developed for pure QED calculations on electronic systems. The most frequently used is the S-matrix formulation [2]. Another approach is the two-time Green's function (TTGF) method, developed by Shabaev and coworkers [3] in St Petersburg, and the other approach is the Green's operator

procedure, based upon the covariant-evolution-operator (CEO), developed by the Gothenburg group [4]. For practical reasons, these procedures in their original form can only be applied to one- and two-photon exchange, which implies that the electron correlation is treated only to the lowest order. This is sufficient and works well for heavy, highly charged ions, where the electron correlation is comparatively weak, but less so for other systems, particularly in the light or medium-heavy region.

For lighter elements, it is possible to expand the electron interactions, including QED and relativity, in a perturbation expansion in powers of α and $Z\alpha$, where α is the fine-structure constant and Z is the nuclear charge, a procedure introduced and employed for a long time by Drake and coworkers [5, 6] and developed more recently to a high degree of sophistication by Pachuchi, Yerokhin, Sapirstein, and others [7–10]. This works very well for light elements but converges slowly for heavier ones.

In a standard MBPT expansion, QED contributions can be included in the form of first-order energy corrections, a procedure successfully applied particularly by the Notre Dame group [11]. For a genuine unification, however, it is required that the QED perturbations are contained in the many-body *wave function* – not just added to the final MBPT energy – which is not possible with the standard MBPT procedures.

Of the three methods for QED calculations, mentioned above, the first two (S-matrix and TTGF) can only give information about the energy, while the last method (Green's operator) can also yield contributions to the wave function. This latter property suggests that it could serve as a basis for a unification of MBPT and QED. Such a unified procedure has recently been developed and is presently being implemented by the Gothenburg group [12–15].

In QED calculations, covariant gauges, like the Feynman gauge, are normally used. The basic idea of the unified QED-MBPT approach presented here, however, is that the *non-covariant Coulomb gauge* is used, where the Coulomb interaction between the electrons is represented by the instantaneous interaction and only the weaker transverse part is time dependent. In combining this with the covariant-evolution procedure, the time-independent Coulomb interaction and time-dependent QED interactions can be mixed arbitrarily. This is a legitimate procedure that is much more effective in treating the combined effect than the conventional procedures.

One fundamental problem in combining MBPT and QED is that they have quite different structures. MBPT is based upon standard quantum mechanics with a single time – the same for all particles – while QED is based on relativistic field theory with individual time for each particle. In order to reconcile the two, one way is to sacrifice the full covariance and apply the *equal-time approximation* with the same time for all particles. This approximation can have serious effects in high-energy physics but fortunately seems to have very small effect on problems in quantum chemistry (see, for instance, the discussion in the Introduction of Ref. [15]).

By far, the most important interaction for normal quantum-chemistry systems is the Coulomb interaction. With the CEO-based procedure, this interaction can be included essentially to all orders, as in standard MBPT and CCA procedures, leading to a correlated wave function. Simultaneously QED perturbations can be included

in the expansion to some low order. The important thing here is that the QED effects are included into the wave function, not only added to the energy. Furthermore, the QED effects need only be considered to the degree that they are expected to be significant. Including only first-order QED effects in the correlated wave function will yield effects that go beyond what has been achieved so far in other approaches. In addition, the most important higher-order QED effects can in this procedure be included by iteration.

Ultimately, the Green's operator procedure, taken to all orders, can be shown to be equivalent to the Bethe-Salpeter equation for two-electron systems [12, 15], which demonstrates its relativistic covariance. It can be regarded as a perturbative procedure for this equation, where certain effects can be treated to higher orders than others (at some expense of the full covariance).

We shall begin this chapter with a review of the standard method of perturbation theory and then discuss how this can be combined with QED by means of the Green's operator (CEO) method. The covariant-evolution-operator (CEO) becomes (quasi)singular, when an intermediate state is (quasi)degenerate with the initial one. The elimination of these singularities leads to the *Green's operator*, which can be regarded as a time-dependent form of the wave operator of standard MBPT. Therefore, this procedure leads to a generalization of standard MBPT to a time- or energy-dependent regime. It is then possible to include the energy-dependent QED perturbations directly into the correlated MBPT wave function, which leads to a true unification. In this way, effects can be evaluated that have previously been beyond reach.

Standard Perturbation Theory

As an introduction, we shall review the standard nonrelativistic many-body perturbation theory (MBPT) (see, for instance, Ref. [16]). We partition the Hamiltonian for atomic and molecular systems into a *model Hamiltonian* and a *perturbation*:

$$H = H_0 + V, \tag{1}$$

where the model Hamiltonian H_0 is a sum of Schrödinger single-electron operators:

$$H_0 = \sum_n h_0(n) \tag{2}$$

and the perturbation V is mainly the electron-electron interaction. The electron orbitals used in the calculations are generated by the single-electron Schrödinger equation:

$$h_0\phi_i(\boldsymbol{x}) = \varepsilon_i\phi_i(\boldsymbol{x}), \tag{3}$$

where

$$h_0 = \frac{\hat{\mathbf{p}}^2}{2m} + v_{\text{ext}}(\boldsymbol{x}). \tag{4}$$

Here, $\hat{\mathbf{p}} = -\mathrm{i}\hbar\nabla$ is the momentum operator, $\hbar$ the Planck's constant divided by 2π, m the electron mass, and $v_{\mathrm{ext}}(\boldsymbol{x})$ the external (mainly nuclear) potential. This is known as the *Furry picture*.

In relativistic theory, the Schrödinger operator is replaced by the corresponding Dirac operator:

$$h_{\mathrm{D}} = c\boldsymbol{\alpha} \cdot \hat{\mathbf{p}} + \beta mc^2 + v_{\mathrm{ext}}, \tag{5}$$

where $\boldsymbol{\alpha}$, β are the Dirac operators.

In the many-body theory, we consider a number of *target states*, satisfying the many-body Schrödinger equation

$$H\Psi^{\alpha} = E^{\alpha}\Psi^{\alpha} \quad (\alpha = 1 \cdots d), \tag{6}$$

which form a *target space*. We define also a *model space* (with the projection operator P), composed of eigenstates of the model Hamiltonian, with the same dimensionality and with a large overlap with the target space. For each target state, there exists a *model state*, which in *intermediate normalization* is the projection on the model space:

$$\Psi_0^{\alpha} = P\Psi^{\alpha}. \tag{7}$$

These states are not necessarily orthogonal, but they are normally linearly independent, so that they span the entire model space.

The orbitals are separated into:

- core orbitals, occupied in all configurations of the model space;
- valence orbitals, occupied in some configurations of the mode space;
- virtual orbitals, unoccupied in the entire model space.

A state with no valence orbitals is said to be a *closed-shell state*, as opposed to an *open-shell state*.

A *wave operator* transforms the model states to the full-target states:

$$\Psi^{\alpha} = \Omega\Psi_0^{\alpha} \quad (\alpha = 1 \cdots d). \tag{8}$$

An *effective Hamiltonian* can be defined, so that it generates the exact energies, when operating on the model functions:

$$H_{\mathrm{eff}}\Psi_0^{\alpha} = PH\Omega\Psi_0^{\alpha} = E^{\alpha}\Psi_0^{\alpha}. \tag{9}$$

This can be separated into a zero-order part and a perturbative part:

$$H_{\mathrm{eff}} = PH_0P + W, \tag{10}$$

where

$$W = PV\Omega P \tag{11}$$

is known as the *effective interaction*.

It can then be shown that in intermediate normalization, the wave operator satisfies the *generalized Bloch equation* [17]:

$$\big[\Omega, H_0\big]P = \big(\Omega H_0 - H_0\Omega\big)P = \big[V\Omega - \Omega W\big]P. \tag{12}$$

This is valid for a general, multilevel (quasidegenerate) model space. If all model states are exactly degenerate (with energy E_0), this reduces to the original Bloch equation [18]:

$$\big(E_0 - H_0\big)\Omega P = \big[V\Omega - \Omega W\big]P. \tag{13}$$

The Bloch equations (13) and (12) can generate the Rayleigh-Schrödinger perturbation expansions for a degenerate and a quasidegenerate model space, respectively.

It is convenient to represent the various terms of the perturbation expansions in terms of *diagrams*, introduced by Feynman and others in quantum field theory in the 1940s and later used in nuclear as well as atomic MBPT (see, for instance, Ref. [16]).

In the diagrammatic representation, the Bloch equation can give rise to terms that are *linked* and *unlinked*. A diagram is called unlinked if it has a separate piece that is "closed," i.e., with the initial and final states in the model space. This is the case for the last term ΩW, if there is no connection between the two operators. The unlinked part is essentially canceled by similar terms originating from the first term, $V\Omega$, while the linked ones survive. This is the *linked-diagram theorem*.

It was conjectured by Brueckner in 1955 [19] that the unlinked terms in the Rayleigh-Schrödinger expansion must vanish for physical reasons, since they increase nonlinearly with the size of the system. This was a few years later formally demonstrated to all orders by Goldstone [20], using field theoretical methods. This has led to the *linked-diagram expansion*:

$$\big[\Omega, H_0\big]P = \big[V\Omega - \Omega W\big]_{\text{linked}}P, \tag{14}$$

which holds also for a quasidegenerate model space, provided that the model space is *complete*, i.e., containing all configurations that can be formed by the valence orbitals. The linked-diagram theorem also holds under certain conditions for an *incomplete* model space [21].

The connected (linked) part of the last term of the Bloch equation (14), which appears only for open-shell systems, is known as the *folded term*, the reason being that it is traditionally represented by folded Goldstone diagrams, as illustrated in Fig. 1. The Goldstone diagrams, normally used in conventional perturbation theory, are *time ordered*, in contrast to the originally introduced Feynman diagrams, which contain all possible time orderings. When all relative time orderings are considered,

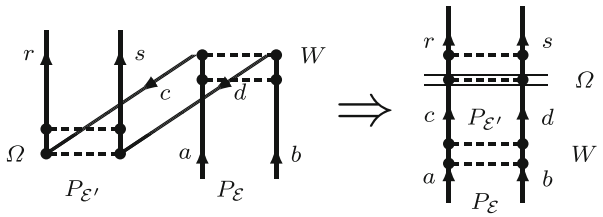


Fig. 1 Illustration of a "folded" diagram, described in the text. The *dashed lines* represent electrostatic interactions. The heavy, *solid lines* represent bound-state orbitals, generated in the Furry picture. The "backward" lines represent valence orbitals. In the figure to the right, the fold is indicated by a *double solid line*

the legs of the diagram of ΩW do factorize into a product of a wave-operator part and an effective-operator part [16]. In the relativistic treatment, it is more natural to employ Feynman diagrams, which automatically contain all time orderings, and then it is no longer motivated to draw the diagram in a folded way. Therefore, we shall here draw them straight, as indicated to the right in the figure. We shall follow the convention, though, and still refer to these diagrams as "folded."

The existence of the linked-diagram theorem has simplified the practical handling of the perturbation expansion enormously in that the number of terms that have to be considered is considerably reduced [16].

The cancelation of the nonlinear terms implies that the Rayleigh-Schrödinger expansion and the linked-diagram version are *size extensive*, i.e., with the energy in each order increasing linearly with the size of the system, in contrast, for example, to the Brillouin-Wigner perturbation expansion.

For a degenerate part of the model space of energy $\mathcal{E}$, with the projection operator $P_{\mathcal{E}}$, the Bloch equation becomes

$$\Omega P_{\mathcal{E}} = P_{\mathcal{E}} + \Gamma(\mathcal{E})\big[V\Omega - \Omega W\big]P_{\mathcal{E}}, \tag{15}$$

where

$$\Gamma(\mathcal{E}) = \frac{1}{\mathcal{E} - H_0} \tag{16}$$

is the *resolvent*. We shall later also need the *reduced resolvent*:

$$\Gamma_Q(\mathcal{E}) = Q\Gamma(\mathcal{E}) = \frac{Q}{\mathcal{E} - H_0}, \tag{17}$$

where

$$Q = 1 - P \tag{18}$$

is the projection operator for the orthogonal space outside the model space.

Omitting the last term of the Bloch equation (15) leads to the expansion without folds:

$$\Omega P_{\mathcal{E}} = \left[1 + \Gamma(\mathcal{E})V + \Gamma(\mathcal{E})V\Gamma(\mathcal{E})V + \cdots\right]P_{\mathcal{E}}. \tag{19}$$

This becomes (quasi)singular, when a state (intermediate or final) lies in the model space. The singularities are removed by the last term of the Bloch equation, $-\Omega W$, but there is a finite remainder, which is represented by the *linked* part of this term. This is referred to as the *model-space contribution*. (This is also referred to as the *reference-state contribution* in cases where only a single reference state is considered.) This is the folded part, mentioned above. We shall see that it plays an important role in the generalization to the energy-dependent formalism that we are presently considering.

All-Order Methods: Coupled-Cluster Approach

In the perturbation expansion, certain effects, such as the pair correlation, can be treated iteratively to arbitrary order by separating the wave operator by means of second quantization into one-body, two-body, etc., parts:

$$\Omega = \Omega_1 + \Omega_2 + \cdots . \tag{20}$$

This leads to the coupled equations:

$$\left[\Omega_n, H_0\right]P = \left[V\Omega - \Omega W\right]_{\text{linked},n} P. \tag{21}$$

Expanding the two-body part (without singles) leads to the *pair function*, illustrated in Fig. 2. This satisfies the *pair equation*:

$$\left[\Omega_2, H_0\right]P = \left[V\Omega_2 - \Omega_1 W_1\right]_{\text{linked},2} P, \tag{22}$$

where

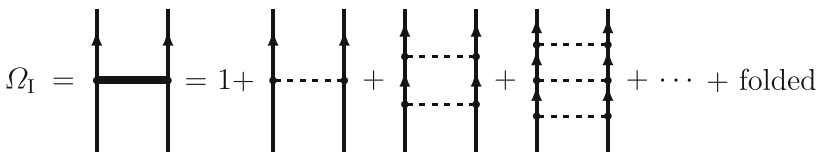


Fig. 2 Expanding the two-body part of the wave operator (without singles) leads to the *pair function*. The *solid vertical lines* represent bound-state orbitals, generated in the Furry picture (Eq. 3). The lines with upgoing *arrows* represent virtual or valence orbitals. The incoming lines (*without arrows*) represent core or valence orbitals

$$W_1 = PV\Omega_1 P \tag{23}$$

is the corresponding effective interaction. The pair function is then expanded in terms of the electron orbitals (3):

$$\Omega_1(1,2) = \sum_{i,j} c_{i,j}\phi_i(1)\phi_j(2). \tag{24}$$

Together with the *exponential Ansatz*

$$\Omega = \mathrm{e}^T = 1 + T + \tfrac{1}{2}T^2 + \cdots \tag{25}$$

the all-order procedure leads to the *coupled-cluster approach* with the *cluster operator* T being a sum of one- and two-body operators, $T = T_1 + T_2 + \cdots$. For the open-shell case, it is convenient to define the cluster operator by means of *normal ordering*, which removes the spurious contractions between valence orbitals [22] that otherwise may occur:

$$\Omega = \{\mathrm{e}^T\} = 1 + T + \tfrac{1}{2}\{T^2\} + \cdots . \tag{26}$$

Here, the curly brackets represent normal ordering. The diagrams of the cluster operator are for a complete model space (see above) *connected* and satisfied in the normal-ordered case the Bloch-type equation:

$$\big[T, H_0\big]P = \big[V\Omega - \Omega W\big]_{\text{connected}} P. \tag{27}$$

(The expansion can be connected also under somewhat more general conditions, as demonstrated by Mukherjee [21, 23].) Note the difference between *"linked"* and *"connected."* A linked diagram can be disconnected, if all parts are open (see, for instance, Ref. [16] for a detailed account).

The coupled-pair equation becomes in analogy with the pair equation in Eq. (22):

$$\big[T_2, H_0\big]P = \big[V\Omega_2 - \Omega_2 W_2\big]_{\text{connected},2} P. \tag{28}$$

Solving these equations self-consistently is known as the coupled-cluster doubles (CCD) approach. The wave operator is then approximated by

$$\Omega = 1 + T_2 + \tfrac{1}{2}\{T_2^2\} + \cdots . \tag{29}$$

One example of a coupled-pair contribution to the term $V\Omega_2$ is displayed in Fig. 3.

In the coupled-pair approximation, the most important quadrupole excitations, which are the most important ones beyond doubles, are included by means of pair functions only. The first CCD calculations were performed in the late 1970s [24–26].

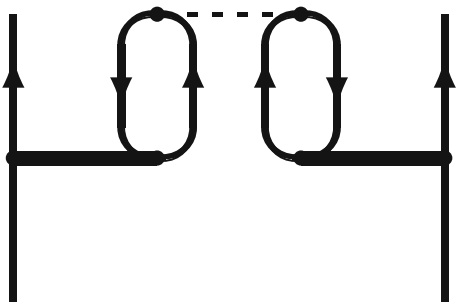


Fig. 3 An example of a coupled-pair contribution to $V\Omega_2$ in Eq. (22). The heavy horizontal line represents the pair function in Fig. 2

A better approximation than CCD is

$$T = T_1 + T_2, \tag{30}$$

known as coupled-cluster singles and doubles (CCSD), where the coupled equations for T_1 and T_2 are solved self-consistently. The first calculation of this kind was performed by Bartlett and Purvis in the early 1980s [27]. Singles have most pronounced effect for valence orbitals, which then become (approximate) Brueckner orbitals [28, 29]. Brueckner orbitals are characterized by the fact that there are no single excitations in any order of perturbation. (Compare Hartree-Fock orbitals for which there are no singles in *first* order [16, Chapter 12].)

For open-shell systems, a serious problem in perturbation theory is the appearance of so-called *intruder states*, i.e., states outside the group of target states that cross one or several target states as the perturbation is gradually turned on. Various schemes have now been developed to avoid or reduce this effect (see Ref. [1] for a recent review).

Multi-configuration Hartree/Dirac-Fock (MCHF/MCDF)

Alternatives to the perturbation expansion are the *multi-configuration Hartree/Dirac-Fock methods*, which are extensions of the standard self-consistent field methods, Hartree-Fock and Dirac-Fock. Here, the energy is minimized by varying the electron orbitals as well as the mixing of configurations, which implies that the electron correlation is included essentially to all orders [30]. In neither scheme, QED effects are included, but first-order QED effects can be added separately to the energy as in the MBPT case.

Relativistic MBPT

The standard relativistic MBPT is based upon the *Dirac-Coulomb-Breit approximation* [31]:

$$H = \Lambda_+ \Big[\sum_{i=1}^{N} h_D(i) + \sum_{i<j}^{N} \frac{e^2}{4\pi r_{ij}} + H_B \Big] \Lambda_+, \tag{31}$$

$$H_B = -\frac{e^2}{8\pi} \sum_{i<j} \Big[\frac{\boldsymbol{\alpha}_i \cdot \boldsymbol{\alpha}_j}{r_{ij}} + \frac{(\boldsymbol{\alpha}_i \cdot r_{ij})(\boldsymbol{\alpha}_j \cdot r_{ij})}{r_{ij}^3} \Big], \tag{32}$$

where h_D is the single-electron Dirac Hamiltonian (5) and $\boldsymbol{\alpha}_i$ is the vector part of the Dirac α matrices for particle i. The projection operators Λ_+ eliminate negative-energy states. This is known as the *no-(virtual)-pair approximation (NVPA)*. The operator H_B is known as the *Breit interaction*, containing magnetic interactions between the electrons as well as some relativistic retardation effects.

Fock Space

Several approaches have been suggested in order to include the correlation between negative-energy states (NES) into the relativistic Hamiltonian. Kutzelnigg has introduced a so-called *Fock-space approach*, starting from a normal-ordered Hamiltonian in the particle-hole formalism [32]. This procedure takes into account the effect of NES, using the instantaneous Coulomb interaction only:

$$V_C = \frac{1}{4\pi\epsilon_0 r_{12}}, \tag{33}$$

where r_{12} is the electron-electron separation. Some QED effects are also included indirectly. The procedure has been modified by Liu, who introduced a *charged-conjugated-contraction* (CCC) potential with the effect that the first-order QED corrections are included in a more consistent way [33, 34]. The first-order QED effects included in the extended procedure are shown in Fig. 4 (compare Fig. 5). The QED parts have to be properly renormalized.

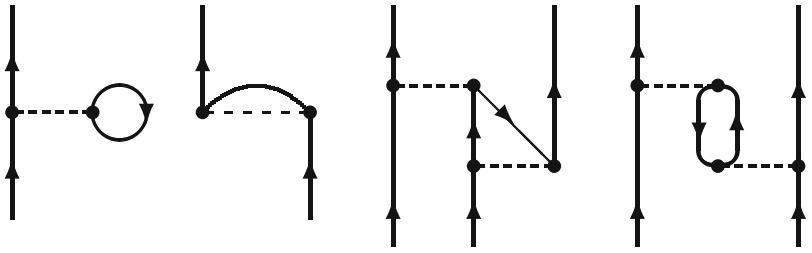


Fig. 4 First-order QED effects included in the extended Fock-space Coulomb-only procedure, from left electron vacuum polarization, single-electron self-energy, two-electron vertex correction, and two-electron vacuum polarization (photon self-energy). Again, *solid vertical lines* represent bound-state orbitals, generated in an external field (Ferry picture, Eq. 3)

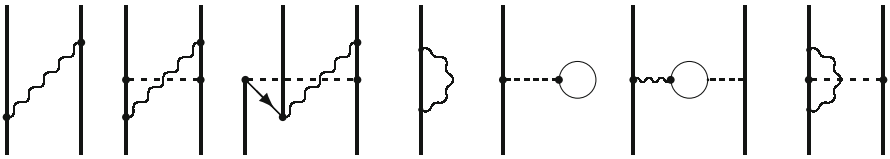


Fig. 5 Examples of low-order QED effects (single transverse photon)

QED Effects

The effects beyond NVPA are conventionally defined as *QED corrections*, which are of the order α^3 or higher (atomic units). Here, α is the fine-structure constant ($\approx 1/137$), and the energy unit is the atomic unit, $2hcR_y$, where h is Planck's constant, c the speed of light in vacuum, and R_y the Rydberg constant. These corrections involve retardation of the electromagnetic interaction and the effect of virtual electron-positron pairs, as well as the so-called *radiative corrections*, namely the electron self-energy and vacuum polarization, the photon self-energy (a form of vacuum polarization), and the vertex correction (see Fig. 5).

Full QED Treatment

QED Methods

A full QED treatment requires that also the *retardation* of the electromagnetic interaction is taken into account (beyond the level included in the Breit interaction), which has an effect of the same order as that of negative-energy states and of the radiative interactions. This can be achieved by means of the field theoretical expression for the interaction between the electrons and the radiation field:

$$\mathcal{H}(x) = -\hat{\psi}^\dagger(x)\, ec\alpha^\mu A_\mu \hat{\psi}(x). \tag{34}$$

Here, $\hat{\psi}$, $\hat{\psi}^\dagger$ are the electron-field operators, $\alpha^\mu = \beta\gamma^\mu$ the Dirac operators, and A_μ the electromagnetic field. An *adiabatic damping* is applied so that the perturbation vanishes as time goes to $\pm\infty$.

There exist several standard methods for QED calculations; the most frequently used methods are:

- *S-matrix formulation* [2],
- *Two-time Green's function*, developed by Shabaev *et al*. at St. Petersburg [3],
- *Covariant-evolution-operator method* (CEO), developed by the Gothenburg group [4].

All three methods are in practice limited to two-photon exchange, which implies that the electron correlation is treated only to the lowest order. The CEO method, however, can be modified so that also higher-order correlation can be included (Green's operator), as we shall describe in the following.

In traditional QED, the entire electron-electron interaction is treated by means of the covariant expression (34). This is, however, a very inefficient way of treating the electron correlation, which plays an important role in most quantum chemical systems. Therefore, we shall here discuss an alternative procedure, introduced by the Gothenburg group. Here, the dominating Coulomb interaction (33) between the electrons is treated by means of the standard Coulomb interaction and only the *transverse* part of the interaction by means of the field theoretical expression (34). This is a non-covariant procedure, based upon the Coulomb gauge, which, however, is quite legitimate and turns out to be a much more efficient way of treating the electron correlation in combination with QED. The Coulomb gauge is more complicated to use in QED than the more frequently used covariant gauges, like the Feynman gauge, but the problems involved have recently been solved by the Gothenburg group [35, 36], based upon works of Adkins [37, 38].

The Covariant Evolution Operator

We shall now introduce the covariant-evolution-operator method and see later how this can lead to the Green's operator that can be used to combine QED with electron correlation to arbitrary order (see [4, 13, 15]).

We start by considering the field theoretical single-particle *Green's function*, which in the single reference case can be defined [39] as

$$G(x, x_0) = \frac{\langle 0_{\mathrm{H}} | T[\hat{\psi}_{\mathrm{H}}(x)\hat{\psi}_{\mathrm{H}}^{\dagger}(x_0)] | 0_{\mathrm{H}} \rangle}{\langle 0_{\mathrm{H}} | 0_{\mathrm{H}} \rangle}, \tag{35}$$

where T is the *Wick time-ordering* operator and $\hat{\psi}_{\mathrm{H}}$, $\hat{\psi}_{\mathrm{H}}^{\dagger}$ are the electron-field operators in the Heisenberg representation ($\hbar = 1$),

$$\langle \Psi_{\mathrm{H}} | = \mathrm{e}^{\mathrm{i}Ht} \langle \Psi_{\mathrm{S}} |, \tag{36}$$

where $\langle \Psi_{\mathrm{S}} |$ is the Schrödinger representation.

The state $|0_{\mathrm{H}}\rangle$ represents the vacuum in this representation, i.e., the state with no particles or holes. Both the numerator and the denominator of the definition (35) can be singular in the limit of vanishing damping, but the *ratio* is always regular.

The standard, nonrelativistic evolution operator represents the time evolution of the nonrelativistic wave function or state vector:

$$\Psi(t) = NU(t, t_0)\,\Psi(t_0), \tag{37}$$

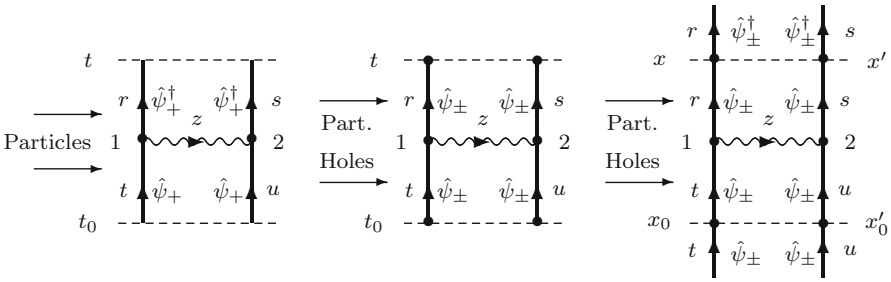


Fig. 6 Comparison between the standard evolution operator, the Green's function, and the covariant-evolution-operator for single-photon exchange in the equal-time approximation

where N is a normalization constant. This is illustrated for single-photon exchange in the first diagram of Fig. 6. Here, only particle states (positive energy) are involved, and time flows only in the positive direction. Therefore, this operator is *not* relativistically covariant. The second diagram represents the corresponding Green's function, where the free lines are replaced by electron propagators, allowing positive- and negative-energy states, making the concept covariant. The last diagram represents the corresponding *covariant-evolution-operator* (CEO), where in addition electron field operators are attached to the free ends, making the CEO an *operator*, while the Green's function is a *function*.

Generally, the single-particle covariant-evolution operator can be defined as

$$U^1_{\mathrm{Cov}}(t,t_0) = \iint \mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{x}_0\,\hat{\psi}^\dagger(x)\langle 0_{\mathrm{H}}|T[\hat{\psi}_{\mathrm{H}}(x)\hat{\psi}^\dagger_{\mathrm{H}}(x_0)]|0_{\mathrm{H}}\rangle\,\hat{\psi}(x_0), \tag{38}$$

using the same vacuum expectation as in the definition of the Green's function (35). In the *interaction picture*,

$$\langle\Psi_{\mathrm{I}}| = \mathrm{e}^{\mathrm{i}H_0 t}\langle\Psi_{\mathrm{S}}|, \tag{39}$$

where H_0 is the model Hamiltonian (1), this leads to the expansion

$$U^1_{\mathrm{Cov}}(t,t_0) = \sum_{n=0}^{\infty}\frac{1}{n!}\iint \mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{x}_0\left(\frac{-\mathrm{i}}{c}\right)^n\int \mathrm{d}^4x_1\cdots\int \mathrm{d}^4x_n$$
$$\times\,\hat{\psi}^\dagger(x)\,T\Big[\hat{\psi}(x)\,\mathcal{H}(x_1)\cdots\mathcal{H}(x_n)\,\hat{\psi}^\dagger(x_0)\Big]\hat{\psi}(x_0)\,\mathrm{e}^{-\gamma(|t_1|+|t_2|\cdots)}, \tag{40}$$

coupled to a one-body operator. Here, γ is the parameter of the adiabatic damping. Note that a contraction of TWO interactions of $\mathcal{H}$ is needed to form a single-photon exchange. Generally, when the evolution operator contains unpaired interactions of the type (34), it will fall in the *photonic Fock space*, where the number of photons is no longer conserved.

The CEO represents the time evolution of the *relativistic* state vector

$$\Psi_{\mathrm{Rel}}(t) = N\,U_{\mathrm{Cov}}(t,t_0)\,\Psi_{\mathrm{Rel}}(t_0) \tag{41}$$

in analogy with the nonrelativistic relation (37).

Energy-Dependent Perturbation Theory: The Green's Operator

We consider now the CEO for a *two-particle ladder* (Fig. 7), analogous to the Coulomb ladder in the pair function in Fig. 2 but with a general energy-dependent interaction, V, (cf. Eq. 19):

$$U_{\mathrm{Cov}}(t,-\infty)\,P_{\mathcal{E}} = \mathrm{e}^{-\mathrm{i}t(\mathcal{E}-H_0)}\big[1+\Gamma(\mathcal{E})\,V(\mathcal{E})+\Gamma(\mathcal{E})\,V(\mathcal{E})\,\Gamma(\mathcal{E})\,V(\mathcal{E})+\cdots\big]P_{\mathcal{E}}, \tag{42}$$

where $P_{\mathcal{E}}$ represents the projection operator for a degenerate part of the model space of energy $\mathcal{E}$ and $\Gamma(\mathcal{E})$ is the corresponding resolvent (16). We shall in the following always assume that the initial time is $t_0 = -\infty$, which implies – with the adiabatic damping – that we start from an unperturbed state.

Intermediate model-space states may lead to (quasi)singularities. It can be shown that the covariant evolution operator can be separated into a regular part $\mathcal{G}(t,-\infty)$, referred to as the *Green's operator*, and a part containing all the singularities:

$$\boxed{U(t,-\infty)P = \mathcal{G}(t,-\infty)\cdot PU(0,-\infty)P.} \tag{43}$$

Here, the heavy dot indicates that the Green's operator acts on the intermediate model-space state.

From the relations (41) and (43) we have

$$\Psi^{\alpha}(t) = N^{\alpha}U(t,-\infty)\Psi^{\alpha}(-\infty) = \mathcal{G}(t,-\infty)\cdot N^{\alpha}PU(0,-\infty)\Psi^{\alpha}(-\infty), \tag{44}$$

but $N^{\alpha}PU(0,-\infty)\Psi^{\alpha}(-\infty)$ is the model function (Eq. 7), $\Psi_0^{\alpha} = P\Psi^{\alpha}$, showing that *the Green's operator acts as a* ***time-dependent wave operator****:*

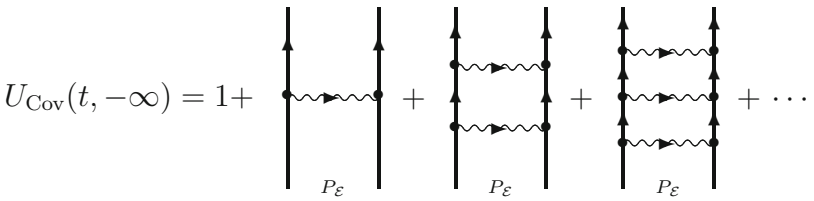


Fig. 7 Ladder diagrams of the covariant-evolution-operator (cf., Fig. 2)

$$\Psi^{\alpha}(t) = \mathcal{G}(t, -\infty)\Psi_0^{\alpha}. \tag{45}$$

For time $t = 0$, the Green's operator becomes equivalent to the standard MBPT wave operator (8):

$$\boxed{\Omega = \mathcal{G}(0),} \tag{46}$$

which demonstrates the close analogy between the QEO procedure and standard MBPT. The covariant *effective interaction* (Eq. 11) is found to be

$$\boxed{W = P\Big(\mathrm{i}\frac{\partial}{\partial t}\,\mathcal{G}(t)\Big)_{t=0}P.} \tag{47}$$

Model-Space Contributions

Lowest Orders

We shall now look closer at the singularities due to intermediate model-space states, which play an important role in the procedure we are now developing. The covariant evolution operator and the Green's operator can have the final state in the model space P as well as in the complementary space Q. Here, we shall restrict ourselves to the latter case with the final state in the Q space. Furthermore, we shall leave out the initial time $-\infty$ and include the energy parameter. The definition (43) can then be expressed as

$$U(t, \mathcal{E})P_{\mathcal{E}} = \mathcal{G}(t, \mathcal{E}')P_{\mathcal{E}'}U(0, \mathcal{E})P_{\mathcal{E}} \tag{48}$$

or

$$\mathcal{G}(t, \mathcal{E})P_{\mathcal{E}} = U(t, \mathcal{E})P_{\mathcal{E}} - \mathcal{G}(t, \mathcal{E}')P_{\mathcal{E}'}\big[U(0, \mathcal{E}) - 1\big]P_{\mathcal{E}}. \tag{49}$$

The last term is referred to as the *counterterm* that eliminates the singularities of the evolution operator, leaving a finite residue, known as the *model-space contribution*. Note that the parameter of the Green's operator of the counterterm is the energy of the *intermediate* model-space state ($\mathcal{E}'$), while the energy parameter of all factors of the evolution operator (42) are equal to the energy of the initial state ($\mathcal{E}$).

According to (42), the first-order Green's operator, which is equal to the first-order evolution-operator, is

$$\mathcal{G}^{(1)}(t, \mathcal{E})P_{\mathcal{E}} = U^{(1)}(t, \mathcal{E})P_{\mathcal{E}} = \mathcal{G}^{(0)}\,\Gamma(\mathcal{E})V(\mathcal{E})P_{\mathcal{E}}. \tag{50}$$

where $\mathcal{G}^{(0)} = \mathrm{e}^{-\mathrm{i}t(\mathcal{E}-H_0)}$. The second-order evolution operator can then be expressed as

$$U^{(2)}(t, \mathcal{E})P_{\mathcal{E}} = U^{(1)}(t, \mathcal{E})\,U^{(1)}(0, \mathcal{E})P_{\mathcal{E}} = \mathcal{G}^{(1)}(t, \mathcal{E})\,U^{(1)}(0, \mathcal{E})P_{\mathcal{E}}. \tag{51}$$

If there is an intermediate model-space state, this becomes

$$\mathcal{G}^{(1)}(t,\mathcal{E})P_{\mathcal{E}'}U^{(1)}(0,\mathcal{E})P_{\mathcal{E}}. \tag{52}$$

(Note, the parameter $\mathcal{E}$ of the Green's operator.)

From the definition (49), we have a counterterm

$$-\mathcal{G}^{(1)}(t,\mathcal{E}')P_{\mathcal{E}'}U^{(1)}(0,\mathcal{E})P_{\mathcal{E}} \tag{53}$$

(note, the parameter $\mathcal{E}'$), yielding the second-order Green's operator

$$\mathcal{G}^{(2)}(t,\mathcal{E})P_{\mathcal{E}} = U_0^{(2)}(t,\mathcal{E})P_{\mathcal{E}} + \left[\mathcal{G}^{(1)}(t,\mathcal{E}) - \mathcal{G}^{(1)}(t,\mathcal{E}')\right]P_{\mathcal{E}'}U^{(1)}(0,\mathcal{E})P_{\mathcal{E}}.$$

Here, $U_0 \equiv \mathcal{G}_0$ represents the evolution operator without any model-space state. But

$$P_{\mathcal{E}'}U^{(1)}(0,\mathcal{E})P_{\mathcal{E}} = P_{\mathcal{E}'}\Gamma_Q(\mathcal{E})V(\mathcal{E})P_{\mathcal{E}} = P_{\mathcal{E}'}\frac{V(\mathcal{E})}{\mathcal{E}-\mathcal{E}'}P_{\mathcal{E}}, \tag{54}$$

and we then have

$$\mathcal{G}^{(2)}(t,\mathcal{E})P_{\mathcal{E}} = \mathcal{G}_0^{(2)}(t,\mathcal{E})P_{\mathcal{E}} + \frac{\delta\mathcal{G}^{(1)}(t,\mathcal{E})}{\delta\mathcal{E}}P_{\mathcal{E}'}W^{(1)}P_{\mathcal{E}}, \tag{55}$$

where $W^{(1)}$ is the first-order effective interaction

$$W^{(1)} = PVP, \tag{56}$$

and

$$\frac{\delta\mathcal{G}^{(1)}(t,\mathcal{E})}{\delta\mathcal{E}} = \frac{\mathcal{G}^{(1)}(t,\mathcal{E}) - \mathcal{G}^{(1)}(t,\mathcal{E}')}{\mathcal{E}-\mathcal{E}'} \tag{57}$$

is a difference ratio that turns into a partial derivative in the case of complete degeneracy. The last term in (55) represents the second-order Green's operator with one model-space state

$$\mathcal{G}_1^{(2)} = \frac{\delta\mathcal{G}^{(1)}(t,\mathcal{E})}{\delta\mathcal{E}}P_{\mathcal{E}'}W^{(1)}P_{\mathcal{E}}. \tag{58}$$

The relation (55) is illustrated in Fig. 8, where the first diagram represents the part without model-space state, $\mathcal{G}_0^{(2)}$, and the second one that with one such state, $\mathcal{G}_1^{(2)}$. The latter corresponds to the "folded" diagram in standard MBPT (compare Fig. 1).

The second-order relation (55) can with somewhat simplified notations for $t = 0$ be expressed as

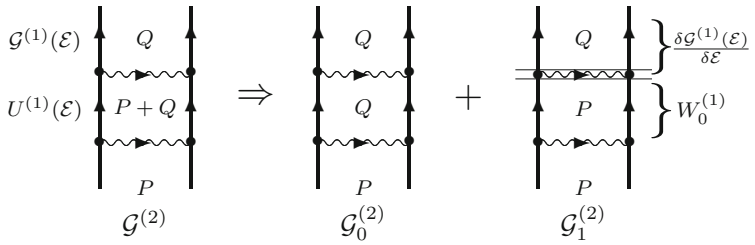


Fig. 8 Elimination of the singularity of the second-order evolution operator due to intermediate model-space states (Eq. 55). The last diagram is the finite remainder or "folded" diagram, corresponding to the folded diagram in the standard MBPT procedure (Fig. 1)

$$\mathcal{G}^{(2)} = \mathcal{G}_0^{(2)} + \frac{\delta \mathcal{G}^{(1)}}{\delta \mathcal{E}} W^{(1)} = \Gamma_Q V\, \mathcal{G}^{(1)} + \frac{\delta \mathcal{G}^{(1)}}{\delta \mathcal{E}} W^{(1)} \tag{59}$$

or

$$\mathcal{G}^{(2)} = \Gamma(\mathcal{E}) V\, \mathcal{G}^{(1)} - \Gamma_Q \mathcal{G}^{(1)} W^{(1)} + \Gamma_Q \frac{\delta V}{\delta \mathcal{E}} W^{(1)}. \tag{60}$$

We then see that this is quite analogous to the second-order expression obtained from the Bloch equation (15). The folded term of the Bloch equation corresponds to the second term in (59) or to $\mathcal{G}_1^{(2)}$ in (58), which has an extra term (60) due to the energy dependence of the interaction.

Higher Orders

We can generalize the treatment above, first to the general case with a single model-space state. The evolution operator with a single model-space state can generally be expressed as

$$U_1(t, \mathcal{E}) P_\mathcal{E} = U_0(t, \mathcal{E}) P_{\mathcal{E}'} \big[U_0(0, \mathcal{E}) - 1\big] P_\mathcal{E}, \tag{61}$$

where U_0 represents the operator without model-space states. According to (49), we have a counterterm:

$$U_1(t, \mathcal{E}) P_\mathcal{E} = -\mathcal{G}_0(t, \mathcal{E}') P_{\mathcal{E}'} \big[U_0(0, \mathcal{E}) - 1\big] P_\mathcal{E}, \tag{62}$$

where we again note the energy parameter $\mathcal{E}'$. This yields the Green's operator with a single model-space state

$$\mathcal{G}_1(t, \mathcal{E}) P_\mathcal{E} = \frac{\delta \mathcal{G}_0(t, \mathcal{E})}{\delta \mathcal{E}} W_0 P_\mathcal{E}, \tag{63}$$

where W_0 is the effective interaction without intermediate model-space states.

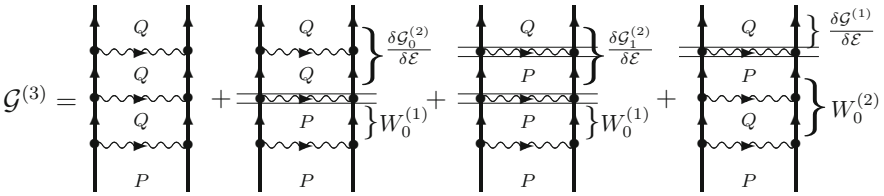


Fig. 9 Graphical representation of the third-order Green's operator (67)

The process can be continued, leading to

$$\begin{aligned}\mathcal{G}(t,\mathcal{E}) &= \mathcal{G}_0(t,\mathcal{E}) + \mathcal{G}_1(t,\mathcal{E}) + \mathcal{G}_2(t,\mathcal{E}) + \cdots \\ &= \mathcal{G}_0(t,\mathcal{E}) + \Big[\frac{\delta \mathcal{G}_0(t,\mathcal{E})}{\delta \mathcal{E}} + \frac{\delta \mathcal{G}_1(t,\mathcal{E})}{\delta \mathcal{E}} + \cdots\Big] W_0 P_{\mathcal{E}}\end{aligned} \tag{64}$$

or

$$\boxed{\mathcal{G}(t,\mathcal{E}) = \mathcal{G}_0(t,\mathcal{E}) + \frac{\delta \mathcal{G}(t,\mathcal{E})}{\delta \mathcal{E}} W_0.} \tag{65}$$

The corresponding equation for the effective interaction is

$$\boxed{W = W_0 + \frac{\delta W}{\delta \mathcal{E}} W_0.} \tag{66}$$

The third-order Green's operator becomes, using the formula above,

$$\begin{aligned}\mathcal{G}^{(3)} &= \mathcal{G}_0^{(3)} + \frac{\delta \mathcal{G}^{(2)}}{\delta \mathcal{E}} W^{(1)} + \frac{\delta \mathcal{G}^{(1)}}{\delta \mathcal{E}} W_0^{(2)} \\ &= \mathcal{G}_0^{(3)} + \frac{\delta \mathcal{G}_0^{(2)}}{\delta \mathcal{E}} W^{(1)} + \frac{\delta \mathcal{G}_1^{(2)}}{\delta \mathcal{E}} W_0^{(1)} + \frac{\delta \mathcal{G}^{(1)}}{\delta \mathcal{E}} W_0^{(2)},\end{aligned} \tag{67}$$

which is illustrated in Fig. 9. $W_0^{(n)}$ represents the effective interaction of order n without intermediate model-space states, and $\mathcal{G}_1^{(2)}$ represents the second-order Green's operator with one fold (58).

Using the expression (58), the third-order Green's operator can also be expressed as

$$\mathcal{G}^{(3)} = \mathcal{G}_0^{(3)} + \frac{\delta \mathcal{G}_0^{(2)}}{\delta \mathcal{E}} W^{(1)} + \frac{\delta \mathcal{G}^{(1)}}{\delta \mathcal{E}} W^{(2)} + \frac{\delta^2 \mathcal{G}^{(1)}}{\delta \mathcal{E}^2} \big(W^{(1)}\big)^2, \tag{68}$$

where $W^{(2)} = W_0^{(2)} + W_1^{(2)}$ is the second-order effective interaction with zero and one intermediate model-space state and

$$W_1^{(2)} = \frac{\delta W^{(1)}}{\delta \mathcal{E}} W^{(1)}. \tag{69}$$

This can be generalized to

$$\boxed{\mathcal{G}(t,\mathcal{E}) = \mathcal{G}_0(t,\mathcal{E}) + \sum_{n=1} \frac{\delta^n \mathcal{G}_0(t,\mathcal{E})}{\delta \mathcal{E}^n} W^n.} \tag{70}$$

and similarly

$$\boxed{W = W_0 + \sum_{n=1} \frac{\delta^n W_0}{\delta \mathcal{E}^n} W^n.} \tag{71}$$

Generalized Bloch Equation

It would be convenient to be able to evaluate the higher-order Green's operators iteratively, in analogy with the the procedure used in ordinary MBPT (12). This would require a Bloch equation valid also for energy-dependent perturbations. In order to derive such an equation, we consider first the third-order expression (68), which we want to express as

$$\mathcal{G}^{(3)} = \Gamma_Q V \mathcal{G}^{(2)} + \cdots . \tag{72}$$

From (59) we have

$$\Gamma_Q V \mathcal{G}^{(2)} = \Gamma_Q V \mathcal{G}_0^{(2)} + \Gamma_Q V \frac{\delta \mathcal{G}^{(1)}}{\delta \mathcal{E}} W^{(1)}, \tag{73}$$

and the expression (68) then becomes

$$\mathcal{G}^{(3)} = \Gamma_Q V \mathcal{G}^{(2)} + \frac{\delta(\Gamma_Q V)}{\delta \mathcal{E}} \mathcal{G}^{(1)} W^{(1)} + \frac{\delta(\Gamma_Q V)}{\delta \mathcal{E}} W^{(2)} + \frac{\delta^2(\Gamma_Q V)}{\delta \mathcal{E}^2} \left(W^{(1)}\right)^2. \tag{74}$$

This contains only derivatives of the last interaction $\Gamma_Q V$, since the previous ones are included in the first term, $\Gamma_Q V \mathcal{G}^{(2)}$. This can be generalized into a Bloch equation for time- or energy-dependent perturbation theory:

$$\boxed{\mathcal{G} = \mathcal{G}^{(0)} + \Gamma_Q V \mathcal{G} + \frac{\delta^* \mathcal{G}}{\delta \mathcal{E}} W,} \tag{75}$$

where the asterisk indicates that the differentiation is restricted to the *last factor* of $\Gamma_Q V$. This relation can also be written as

$$\boxed{\mathcal{G} = \mathcal{G}^{(0)} + \Gamma_Q V \mathcal{G} - \Gamma_Q \mathcal{G} W + \frac{\delta^* \mathcal{G}}{\delta \mathcal{E}} W} \tag{76}$$

with differentiation of the last *interaction*, V, only. This equation can be compared with the standard Bloch equation (15), and we see that the essential difference lies in the extra model-space term, involving the energy derivative of the interaction. The equation holds also when the interactions are different, provided that V in the second term is the last interaction and that $\mathcal{G}$ in the remaining terms is formed by the last interactions and W of the remaining ones.

The similarity between the Bloch equations for the Green's operator and the standard MBPT wave operator demonstrates that the perturbation expansion based upon the CEO or Green's function is completely compatible with the standard procedure. Therefore, it can serve as a convenient basis for a *unified procedure*, where QED and Coulomb interactions can, in principle, be mixed arbitrarily.

This completes the description of the procedure for energy-dependent perturbation theory, developed by the Gothenburg group, and we shall now see how this can be applied to the QED perturbations in combinations with electron correlation.

Combining QED with Electron Correlation

General

We shall now illustrate how the procedure of covariant-evolution operators presented here can be applied to the energy-dependent QED perturbations in combination with electron correlation. We consider first the case, where we have an iterated wave function of Coulomb interaction, as in Fig. 2, and add to that one energy-dependent perturbation. This is illustrated in Fig. 10.

In third order we have from (74)

$$\Omega^{(3)} = \Gamma_Q V \Omega_{\mathrm{I}}^{(2)} + \frac{\delta(\Gamma_Q V)}{\delta \mathcal{E}} \Gamma_Q V_{\mathrm{C}} W_0^{(1)} + \frac{\delta(\Gamma_Q V)}{\delta \mathcal{E}} W^{(2)} + \frac{\delta^2(\Gamma_Q V)}{\delta \mathcal{E}^2} \left(W_0^{(1)}\right)^2. \tag{77}$$

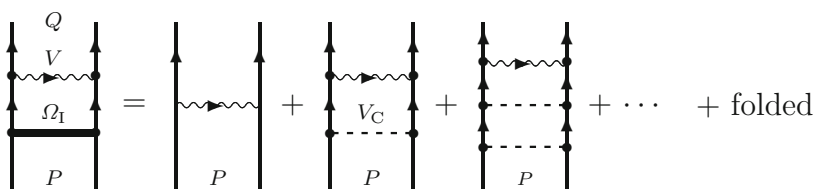


Fig. 10 A single energy-dependent interaction is added to the pair function of Coulomb interactions in Fig. 2

The last three terms represent the folded part due to the last energy-dependent perturbation, while the folds due to the energy-independent part are contained in the wave function $\Omega_{\rm I}$. This can be generalized to a Bloch-like equation:

$$\boxed{\Omega = 1 + \Gamma_Q V\, \Omega_{\rm I} + \sum_{m=1} \frac{\delta^m (\Gamma_Q V)}{\delta \mathcal{E}^m}\, \Omega_{\rm I}\, W_{\rm I}^m.} \tag{78}$$

Here,

$$\Omega_{\rm I} = 1 + \Gamma_Q V_{\rm C} + \Gamma_Q V_{\rm C} \Gamma_Q V_{\rm C} + \cdots + \text{folded} \tag{79}$$

is the Coulomb pair function in Fig. 2 and $W_{\rm I}$ is the corresponding effective interaction (23):

$$W_{\rm I} = P V_{{\rm C}\Omega_{\rm I}} P, \tag{80}$$

$V_{\rm C}$ being the Coulomb interaction. This makes it possible to add an energy-dependent perturbation to an iterated wave function of energy-independent interactions, such as a Coulomb pair function.

The corresponding relation for the effective interaction, with the final state in the model space, is

$$\boxed{W = P V \Omega_{\rm I} P + P \sum \frac{\delta^m (V)}{\delta \mathcal{E}^m}\, \Omega_{\rm I}\, W_{\rm I}^m.} \tag{81}$$

The energy-dependent interaction, discussed here, can, in principle, contain all kinds of QED effects, such as retardation, electron self-energy, vacuum polarization, etc., as illustrated in Fig. 11.

For the time being, it is only feasible to include *single-photon* effects in a QED potential. But the procedure can be iterated, as illustrated in the second diagram of Fig. 12. That will yield so-called *reducible* diagrams, i.e., diagrams that can be separated into simpler diagrams by a horizontal cut. A reducible diagram is generally defined as a diagram that can be separated into two allowed diagrams by a horizontal cut. If only energy diagrams are considered, then a diagram is reducible only if the intermediate state lies in the model space. If, on the other hand, also diagrams of the wave function or wave operator are considered, as we do here, then a diagram can be reducible also if the intermediate state lies in the complementary Q space and contains no photons. Examples of reducible diagrams are shown in Fig. 13.

After applying the energy-dependent perturbation to the Coulomb pair function, additional Coulomb perturbations can be added, as illustrated in the first diagram of Fig. 12.

Iterating the general potential in Fig. 11 and including all irreducible effects leads ultimately to the two-electron *Bethe-Salpeter equation* [12, 15].

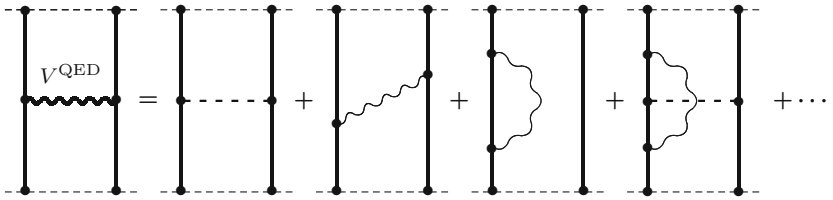


Fig. 11 QED potential, containing first-order QED effects

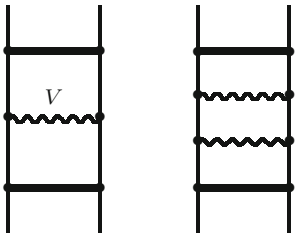


Fig. 12 Single and iterated QED potential with electron correlation

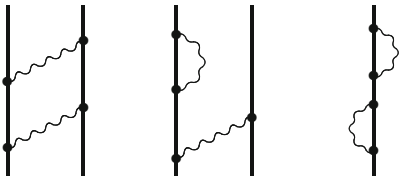


Fig. 13 Examples of *reducible* two-photon diagrams that can be separated into legitime diagrams by a horizontal cut

The QED pair functions, illustrated in Fig. 11, can, of course, also be used for systems with more than two electrons, for instance, in a coupled-cluster approach (CCA) in the same way as standard pair functions, discussed in section "All-Order Methods: Coupled-Cluster Approach" (see Ref. [40]). Due to the complete compability between the standard MBPT and CEO approaches, the QED and non-QED pair functions can be mixed and the former need only be inserted, when the QED effect is expected to be significant. For instance, in the coupled-pair equation (22), it might be sufficient to insert the QED-pair function in the linear term but use the standard pair functions in the non-linear, coupled term.

An additional important advantage of using the Coulomb gauge is that the instantaneous Breit interaction (Eq. 32) can be used in the same fashion as the Coulomb interaction, which will yield important effects that otherwise would require irreducible multi-photon effects.

Retardation and Virtual Pairs

A potential for the retarded electron-electron interaction, involving also negative-energy states can be constructed, allowing also for crossing Coulomb interactions

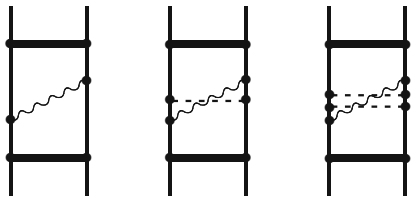


Fig. 14 Retarded transverse interaction with electron correlation and with one and several Coulomb crossings

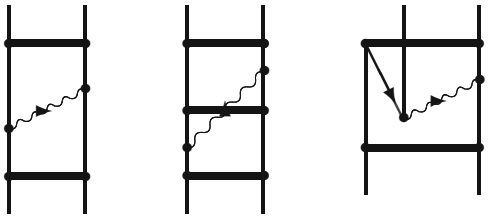


Fig. 15 Diagrams representing the combination of non-radiative QED perturbation (retardation, virtual pair) and electron correlation, evaluated by Hedendahl [41]

[41], [15, Chap. 8]. This can be combined with electron correlation, as discussed in the previous section and illustrated in Figs. 14 and 15.

Self-Energy and Vertex Correction

More recently, also *radiative* QED effects (electron self-energy and vertex correction) have been included in the procedure [14]. So far, all self-energy calculations on bound-state systems have been performed using a covariant gauge, like the Feynman gauge. When radiative QED effects, mainly, self-energy and vertex corrections, it turns out that there are large cancelations between the various contributions, which make the calculations more unreliable. In the Coulomb gauge, this is much less the case. This has the consequence that in going beyond second order, it is only possible to obtain reliable numerical results in the Coulomb gauge as recently demonstrated by our group [14]. This gauge is more complicated to use in QED calculations, but these problems here have recently been solved by Hedendahl and Holmberg [35,36] in the Gothenburg group, using dimensional regularization, based upon the works of Adkins [37,38] (see also [15, Chap. 12]).

For a two-electron system, we have in second order the screened self-energy and vertex correction, illustrated in Fig. 16. We assume that the self-energy is mass renormalized as mentioned above. Then the first diagram (A) is regular, while the second one is divergent, due to model-space contribution. This singularity is canceled by the last vertex-correction diagram (C), due to the Ward identity [15,42].

With an incoming pair function, we have the situation in Fig. 17 with a Q or a P state immediately before the QED perturbation. We can now apply the formula (81), with $V = \Sigma$ for the self-energy and $V = \Lambda^0 V_{\mathrm{C}}$ for the vertex correction (Λ^0 being vertex-correction parameter).

From the formula (81) we find for diagram AA

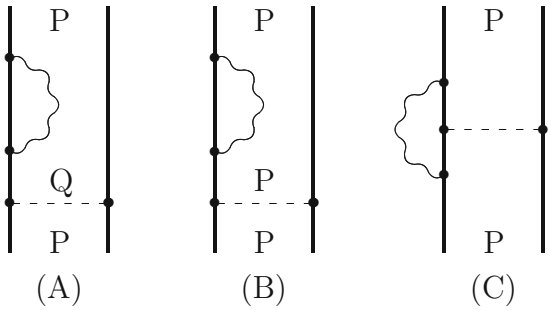


Fig. 16 First-order Coulomb screened self-energy and vertex correction for He-like systems

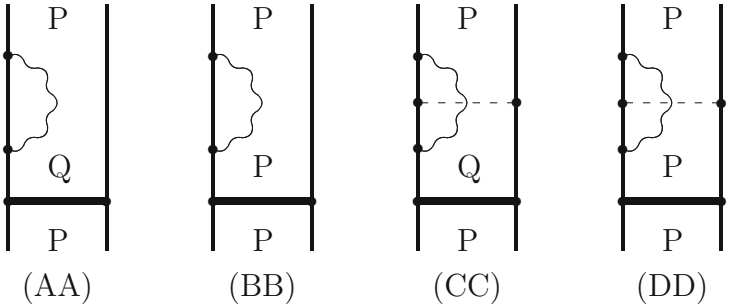


Fig. 17 All-order screened self-energy and vertex correction for He-like systems

$$W_A = P\,\Sigma\,Q\,\Omega_{\rm I}\,P + P\sum\frac{\delta\Sigma}{\delta\mathcal{E}}\,Q\,\Omega_{\rm I}\,W_{\rm I} + \cdots \tag{82}$$

diagram BB

$$W_B = P\sum\frac{\delta\Sigma}{\delta\mathcal{E}}\,W_{\rm I} + \cdots \tag{83}$$

diagram CC

$$W_B = P\,\Lambda^0 V_{\rm C}\,Q\,\Omega_{\rm I}\,P + P\frac{\delta\Lambda^0 V_{\rm C}}{\delta\mathcal{E}}\,Q\,\Omega_{\rm I}\,W_{\rm I} + \cdots \tag{84}$$

and diagram DD

$$W_B = P\frac{\delta\Lambda^0 V_{\rm C}}{\delta\mathcal{E}}\,W_{\rm I}^2 + \cdots \tag{85}$$

The diagrams BB and CC are divergent with the singularities canceling due to the Ward identity, while the diagrams AA and DD are regular.

Including additional Coulomb iterations, it leads to the situation indicated in Fig. 18.

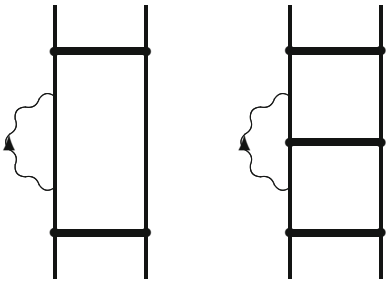


Fig. 18 Diagrams representing the combination of radiative QED perturbation (self-energy, vertex correction, vacuum polarization) and electron correlation, recently evaluated by Johan Holmberg et al. [50]

Summary

The X-ray transition energies for helium-like ions can in some cases be measured with extreme accuracy, in some cases up to the ppm level [43]. Such experimental results can therefore serve as reference for tests of computational methods and eventually for a test of the QED theory itself. Calculations on the two-photon level have been carried out by Artemyev *et al.* by means of the two-time Green's function procedure [44]. This takes the single- and two-photon exchange into account in a complete way but leaves out all third- and higher-order effects. Related calculations have been performed by Plante *et al.*, using the relativistic MBPT with first-order QED energy corrections to the energy [11]. Here, higher-order effects are included but only in a limited way. The agreement with experimental results is normally quite good for these calculations, but Chantler *et al.* claim that there are significant discrepancies [43, 45].

Third- and higher-order effects that are missing from the calculations of Artemyev, Plante, and coworkers might be relevant for an accurate comparison with the experiments. With the Green's operator procedure, discussed here, such high-order effects can be evaluated, and such calculations have been performed for some time by the Gothenburg group. The calculated effects contain a *single* transverse photon, radiative and non-radiative, combined with Coulomb interactions to essentially all orders. Such calculations, which have to be performed in the Coulomb gauge, are quite complicated and time-consuming but still manageable. The accumulated effect of these contributions is of the order of 5 meV for the ground states of the medium-heavy helium-like ions [14], which is at least comparable to many experimental uncertainties. Corresponding effects with *multiple* transverse photons are generally beyond reach, but they should be considerably smaller and expected to be negligible for the comparison between theory and experiment considered here.

The calculated higher-order QED effects (beyond second order) are much too small to be able to explain the discrepancies claimed by Chantler. Therefore, if these discrepancies are real, they must have other causes. Furthermore, several experimental groups have recently obtained results that are in excellent agreement with the theoretical results, thereby strongly challenging the findings of Chantler [46–48].

The Green's operator procedure was primarily developed for investigating *static* problems, but it has recently been demonstrated that also *dynamical* processes, like scattering processes or transitions between atomic states, are governed by the Green's operator and can be handled by a similar procedure [49].

Acknowledgements The author wants to express his gratitude to his coworkers Sten Salomonson and Johan Holmberg for their valuable collaboration and the latter also for making unpublished result available to him. The author also wants to acknowledge stimulating discussions with Professor Bogumil Jeziorski, Professor Werner Kutzelnigg, Professor Wenjian Liu, and Professor Debashis Mukherjee.

References

1. Čársky P, Paldus J, Pittner J (eds) (2009) Recent progress in coupled cluster methods: theory and applications. Springer, New York
2. Mohr PJ, Plunien G, Soff G (1998) CODATA recommended values of the fundamental physical constants:2010. Phys Rep 293:227
3. Shabaev VM (2002) Two-time Green's function method in quantum electrodynamics of high-Z few-electron atoms. Phys Rep 356:119
4. Lindgren I, Salomonson S, Åsén B (2004) The covariant-evolution-operator method in bound-state QED. Phys Rep 389:161
5. Drake GWF (1988) Theoretical energies for the $n = 1$ and 2 states of the helium isoelectronic sequence up to Z=100. Can J Phys 66:586
6. Drake GWF (2002) Progress in helium fine-structure calculations and the fine-structure constant. Can J Phys 80:1195
7. Pachucki K (1999) Quantum electrodynamics effects on helium fine structure. J Phys B 32:137
8. Pachucki K, Sapirstein J (2002) Determination of the fine structure constant from helium spectroscopy. J Phys B 35:1783
9. Pachucki K (2006) Improved theory of Helium fine structure. Phys Rev Lett 97:013002
10. Pachucki K, Yerokhin VA (2010) Reexamination of the helium fine structure. Phys Rev A 81:39903
11. Plante DR, Johnson WR, Sapirstein J (1994) Relativistic all-order many-body calculations of the n=1 and n=2 states of heliumlike ions. Phys Rev A 49:3519
12. Lindgren I, Salomonson S, Hedendahl D (2005) Many-body-QED perturbation theory: Connection to the two-electron Bethe-Salpeter equation. Einstein centennial review paper. Can J Phys 83:183
13. Lindgren I, Salomonson S, Hedendahl D (2006) Many-body procedure for energy-dependent perturbation: merging many-body perturbation theory with QED. Phys Rev A 73:062502
14. Holmberg J, Salomonson S, Lindgren I (2015) Coulomb-gauge calculation of the combined effect of the correlation and QED for heliumlike highly charged ions. Phys Rev A 92:012509
15. Lindgren I (2011) Relativistic many-body theory: a new field-theoretical approach. Springer, New York; Second revised edition (in production)
16. Lindgren I, Morrison J (1986) Atomic Many-Body Theory, 2nd edn. Springer, Berlin. Reprinted 2009
17. Lindgren I (1974) The Rayleigh-Schrödinger perturbation and the linked-diagram theorem for a multi-configurational model space. J Phys B 7:2441
18. Bloch C (1958) Sur la théorie des perurbations des etats liés. Nucl Phys 6/7:329/451
19. Brueckner KA (1955) Many-Body Problems for Strongly Interacting Particles. II. Linked Cluster Expansion. Phys Rev 100:36

20. Goldstone J (1957) Derivation of the Brueckner many-body theory. Proc R Soc Lond Ser A 239:267
21. Mukherjee D (1986) Linked-cluster theorem in the open-shell coupled-cluster theory for incomplete model spaces. Chem Phys Lett 125:207
22. Lindgren I (1978) A coupled-cluster approach to the many-body perturbation theory for open-shell systems. Int J Quantum Chem S12:33
23. Lindgren I, Mukherjee D (1987) On the connectivity criteria in the open-shell coupled-cluster theory for the general model spaces. Phys Rep 151:93
24. Bartlett RJ, Purvis GD (1978) Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem. Int J Quantum Chem 14:561
25. Pople JA, Krishnan R, Schlegel HB, Binkley JS (1978) Electron correlation theories and their application to the study of simple reaction potential surfaces. Int J Quantum Chem 14:545
26. Lindgren I, Salomonson S (1980) A numerical coupled-cluster procedure applied to the closed-shell atoms Be and Ne. Phys Scr 21:335
27. Purvis GD, Bartlett RJ (1982) A full coupled-cluster singles and doubles model: the inclusion of disconnected triples. J Chem Phys 76:1910
28. Löwdin P-O (1962) Studies in perturbation theory. V. Some aspects on the exact self-consistent field theory. J Math Phys 3:1171
29. Lindgren I (1985) Accurate many-body calculations on the lowest ^2S and ^2P states of the lithium atom. Phys Rev A 31:1273
30. Grant IP (2007) Relativistic quantum theory of atoms and molecules. Springer, Heidelberg
31. Sucher J (1980) Foundations of the relativistic theory of many electron atoms. Phys Rev A 22:348
32. Kutzelnigg W (2012) Solved and unsolved problems in quantum chemistry. Chem Phys 395:16
33. Liu W (2012) Perspectives of relativistic quantum chemistry: the negative energy cat smiles. Phys Chem Chem Phys 14:35
34. Liu W, Lindgren I (2013) Going beyond 'no-pair relativistic quantum chemistry'. J Chem Phys 139:014108
35. Holmberg J (2011) Scalar vertex operator for bound-state QED in Coulomb gauge. Phys Rev A 84:062504
36. Hedendahl D, Holmberg J (2012) Coulomb-gauge self-energy calculation for high-Z hydrogenic ions. Phys Rev A 85:012514
37. Adkins G (1983) One-loop renormalization of Coulomb-gauge QED. Phys Rev D 27:1814
38. Adkins G (1986) One-loop vertex function in Coulomb-gauge QED. Phys Rev D 34:2489
39. Itzykson C, Zuber JB (1980) Quantum Field Theory. McGraw-Hill, New York
40. Lindgren I, Salomonson S, Hedendahl D (2010) In: Čársky P, Paldus J, Pittner J (eds) Recent progress in coupled cluster methods: theory and applications. Springer, New York, pp 357–374
41. Hedendahl D (2010) Ph.D. thesis, University of Gothenburg, Gothenburg
42. Peskin ME, Schroeder DV (1995) An introduction to quantun field theory. Addison-Wesley Publ. Co., Reading
43. Chantler CT et al (2012) New X-ray measurements in Helium-like Atoms increase discrepancy between experiment and theoretical QED. arXiv-ph 0988193
44. Artemyev AN, Shabaev VM, Yerokhin VA, Plunien G, Soff G (2005) QED calculations of the n=1 and n=2 energy levels in He-like ions. Phys Rev A 71:062104
45. Chantler CT (2012) Testing Three-Body Quantum-Electrodynamics with Trapped T^{20+} Ions: Evidence for a Z-dependent Divergence Between Experimental and Calculation. Phys Rev Lett 109:153001
46. Kubiček K, Mokler PH, Mäckel V, Ullrich J, López-Urrutia JC (2014) Transition energy measurements in hydrogenlike and heliumlike ions strongly supporting bound-state QED calculations. PRA 90:032508
47. Beiersdorfer P, Brown GV (2015) Experimental study of the X-ray transition in the helium like isoelectronic sequence: updated results. Phys Rev A 91:032514

48. Epp SW et al (2015) Single-photon excitation of Kα in helium like Kr^{34+}: Results supporting quantum electrodynamics predictions. Phys Rev A 92:020502(R)
49. Lindgren I, Salomonson S, Holmberg J (2014) QED effects in scattering involving atomic bound states: radiative recombination. Phys Rev A 89:062504
50. Holmberg J, Salomonson S, Lindgren I (2015) Coulomb-gauge calculation of the combined effect of the correlation and QED for heliumlike highly charged ions. PRA 92:012509

Part III

Relativistic Hamiltonians

Wenjian Liu

With-Pair Relativistic Hamiltonians

10

Wenjian Liu

Contents

Introduction 346
Derivation of the eQED Hamiltonian 348
Charge Conjugation 348
Diagrammatic Derivation of the eQED Hamiltonian 353
Algebraic Derivation of the eQED Hamiltonian 358
The Model Q Operator 363
No-Pair Relativistic Hamiltonians 365
Application of the eQED Hamiltonian 366
Second-Order Energy of an N-Electron System 366
Mean-Field Theory of Positrons 370
Conclusion 371
References 372

Abstract

While the no-pair approximation widely used in relativistic quantum chemical calculations is good enough for most purposes, there is still a great need to go beyond it, not only for better accuracies but also for better understandings of relativistic quantum mechanics. It is shown here that, at variance with the usual top-down procedures for deriving relativistic Hamiltonians as approximations to quantum electrodynamics (QED), a with-pair relativistic many-electron Hamiltonian can be constructed in a bottom-up fashion without recourse to QED itself. It describes all virtual pair effects due to the instantaneous Coulomb/Gaunt/Breit

W. Liu (✉)
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: liuwjbdf@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_1

interaction and is compatible with all wave function or density-functional-based correlation methods. As such, it serves as the basis for "with-pair relativistic quantum chemistry," an extension of the traditional "no-pair relativistic quantum chemistry." Due to the short range nature, the effective potential Q describing the electron vacuum polarization (EVP) and self-energy (ESE) can be fitted into a model operator and included in variational mean-field calculations. The major QED effects, including EVP and ESE, as well as the Coulomb/Gaunt/Breit screenings of them, can then readily be accounted for in subsequent treatments of electron correlation and properties.

Keywords
Relativistic Hamiltonians • Virtual pair effects • Charge conjugation • Quantum electrodynamics

Introduction

Because of the very large gap ($2mc^2 \approx 1\,\mathrm{MeV}$) between the positive (PES) and negative (NES) energy states of the Dirac spectrum, freezing the NES in post-SCF (self-consistent field) relativistic quantum chemical calculations of low-energy chemical/physical processes/properties is usually a very good approximation. This is just the so-called no-virtual-pair or simply no-pair approximation (NPA). Not only so, under this approximation, all determinant-based wave function or density-functional methods developed in the nonrelativistic regime for electron correlation/excitation can directly be transplanted to the relativistic domain, although additional technical difficulties do arise from the reduction in symmetry and the use of complex algebra. The last decade has witnessed remarkable advances in this "no-pair relativistic quantum chemistry," symbolized primarily by the advent and development of the so-called exact two-component (X2C) Hamiltonians [1–4]. Nevertheless, one has to be aware that the NPA has several fundamental defects, both numerically and conceptually [5, 6]. First of all, the NPA results in an intrinsic error of $\mathscr{O}(Z\alpha)^3$ in the energy [7], which is not only significant for core or semi-core electrons but also relevant for valence electrons when the correlation treatment has reached an accuracy of a few hundredths of an electron volt. Second, unlike nonrelativistic FCI (full configuration interaction), the no-pair relativistic correlation energy calculated at whatever level of methodology is always dependent on the effective potential employed for generating the orbitals [8]. Third, there are even situations where the NPA completely breaks down. For instance, in the sum-over-state formulation of nuclear magnetic resonance shielding in terms of the field-free PES and NES, the relativistic diamagnetism would be missed entirely under the NPA [9]. Fourth, the no-pair projected Hamiltonians, whether four or two component, are incompatible with explicitly correlated wave function methods due to two reasons [7]: (a) Integrals like $\langle pq|\hat{h}_{\mathrm{rel}} f_{12}|rs\rangle$ and $\langle pq| f_{12}\hat{h}_{\mathrm{rel}} f_{12}|rs\rangle$ must be evaluated analytically to achieve fast convergence [10], but the required (first quantized) relativistic operator $\hat{h}_{\mathrm{rel}}$ is not available in such Hamiltonians. (b) More

seriously, such Hamiltonians have only a finite spectrum limited by the given basis, such that the effect of the correlation factor f_{12} is just null: The r_{12}-dependent two-electron basis functions try to simulate a correlation space that is complementary to the determinantal space but which is not part of the Hamiltonians. Fifth, since the correlation of NES is of the same order of $\mathcal{O}(Z\alpha)^3$ as the leading radiative QED effect (Lamb shift), the usual Ansatz "NPA + Lamb shift" adopted in calculations of many-electron systems is not fully justified. For instance, the remaining discrepancy (ca. 0.05 eV) [11] between the theoretical "NPA + Lamb shift" and experimental ionization potentials of the gold atom is likely due to the missing correlation of NES. Therefore, to be internally consistent, the electron vacuum polarization (EVP), electron self-energy (ESE), and the correlation of NES should be treated on an equal footing.

The question is how to handle the NES properly. There have been many attempts [12–15] to solve directly the non-projected, first-quantized Dirac-Coulomb/Gaunt/Breit (DC/DCG/DCB) equation with standard techniques such as CI and MBPT (many-body perturbation theory). Here, the number of electrons is conserved, and the NES are treated as if they were particles, just like the virtual PES. This manifestly violates charge conjugation symmetry, a key ingredient to distinguish relativistic from nonrelativistic quantum mechanics. As such, the so-obtained results are plainly wrong (vide post), regardless of numerical instabilities [16]. In short, the first-quantized DC/DCG/DCB equation cannot be solved as it stands, even if it has bound states. Instead, the NPA is a conceptual must for such configuration space formulation, rather than merely an effective means [17] for avoiding the (in)famous Brown-Ravenhall disease [16]. The proper treatment of the NES can only be done with a Fock space formulation, where it is the charge instead of the particle number that is conserved and the NES and PES are treated symmetrically. The essential ingredients of such a Fock space formulation include the filled Dirac picture and second quantization [18, 19] along with the charge-conjugated normal ordering of fermion operators [20], with which a with-pair relativistic many-electron Hamiltonian, or simply effective QED (eQED) Hamiltonian in the Coulomb gauge, can be obtained naturally without recourse to QED itself. It can be reduced [20, 21] to that [22] derived from QED in a top-down fashion. Further approximations based purely on physical arguments lead to a series of relativistic Hamiltonians that span a complete and continuous "Hamiltonian Ladder" [21, 23].

The present account aims to present a thorough discussion on charge conjugation symmetry (section "Charge Conjugation") as well as its incorporation in describing electron-electron interactions (sections "Diagrammatic Derivation of the eQED Hamiltonian" and "Algebraic Derivation of the eQED Hamiltonian"). As an illustration, the second-order energy of an N-electron system is derived explicitly in section "Second-Order Energy of an N-Electron System" and compared with that of the configuration space formulation. A mean-field theory for real positrons is further formulated in section "Mean-Field Theory of Positrons." The account is closed with concluding remarks in section "Conclusion." Apart from the Einstein summation convention over repeated indices, the following convention is to be

adopted: The occupied PES and NES are denoted by $\{i, j, \ldots\}$ and $\{\tilde{i}, \tilde{j}, \ldots\}$, respectively, whereas the unoccupied PES by $\{a, b, \ldots\}$. Unspecified orbitals are denoted as $\{p, q, r, s\}$. When necessary, the NES are designated explicitly by $\{\tilde{p}, \tilde{q}, \tilde{r}, \tilde{s}\}$.

Derivation of the eQED Hamiltonian

Charge conjugation plays a key role in relativistic quantum mechanics and is therefore first discussed in section "Charge Conjugation" before the eQED Hamiltonian is formulated diagrammatically in section "Diagrammatic Derivation of the eQED Hamiltonian" and algebraically in section "Algebraic Derivation of the eQED Hamiltonian."

Charge Conjugation

The charge conjugation transformation of a Dirac spinor $\varphi_p(x)$ (which can be understood as a classical field) is effected by

$$C : \varphi_p(x) \mapsto \varphi_p^C(x) = \mathbf{C}\beta\varphi_p^*(x), \quad x = \mathbf{r}t, \tag{1}$$

where the 4-by-4 matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = -\mathbb{i}\alpha_y, \tag{2}$$

with the following properties

$$\mathbf{C}^\dagger = \mathbf{C}^T = \mathbf{C}^{-1} = -\mathbf{C}, \tag{3}$$

$$\mathbf{C}\beta = -\beta\mathbf{C}, \quad \mathbf{C}\alpha_i = -\alpha_i^*\mathbf{C} = -\alpha_i^T\mathbf{C}, \quad i = x, y, z. \tag{4}$$

By taking the complex conjugate of the Dirac equation,

$$\mathbb{i}\frac{\partial}{\partial t}\varphi_p(x) = [c\boldsymbol{\alpha}\cdot(\mathbf{p} - q\mathbf{A}_{\text{ext}}) + \beta mc^2 + q\phi_{\text{ext}}]\varphi_p(x), \tag{5}$$

followed by multiplying from the left by $\mathbf{C}\beta$, we obtain

$$-\mathbb{i}\frac{\partial}{\partial t}\varphi_p^C(x) = \mathbf{C}\beta[-c\boldsymbol{\alpha}^*\cdot(\mathbf{p} + q\mathbf{A}_{\text{ext}}) + \beta mc^2 + q\phi_{\text{ext}}]\varphi_p^*(x), \tag{6}$$

$$= [-c\boldsymbol{\alpha}\cdot(\mathbf{p} + q\mathbf{A}_{\text{ext}}) - \beta mc^2 + q\phi_{\text{ext}}]\varphi_p^C(x), \tag{7}$$

and hence

$$\mathbb{i}\frac{\partial}{\partial t}\varphi_p^C(x) = [c\boldsymbol{\alpha}\cdot(\mathbf{p} + q\mathbf{A}_{\text{ext}}) + \beta mc^2 - q\phi_{\text{ext}}]\varphi_p^C(x). \tag{8}$$

It follows that, for the same time-independent external potential $(\phi_{\rm ext}(\mathbf{r}), \mathbf{A}_{\rm ext}(\mathbf{r}))$, if $\varphi_p(x) = \varphi_p(\mathbf{r})e^{-i\epsilon_p t}$ is a stationary solution of the Dirac equation for an electron $(q = -1)$ of energy ϵ_p, $\varphi_p^C(x) = \varphi_p^C(\mathbf{r})e^{i\epsilon_p t}$ will then be a stationary solution of the Dirac equation for a positron $(q = +1)$ of energy $-\epsilon_p$. Note in particular that the probability density of a negative energy/frequency $(-|\epsilon_{\tilde{p}}|)$ electron $\varphi_{\tilde{p}}(\mathbf{r})e^{-i|\epsilon_{\tilde{p}}|(-t)}$ is indistinguishable from that of a positive energy $(|\epsilon_{\tilde{p}}|)$ positron $\varphi_{\tilde{p}}^C(\mathbf{r})e^{-i|\epsilon_{\tilde{p}}|t}$, i.e., $|\varphi_{\tilde{p}}^C(\mathbf{r})|^2 = |\varphi_{\tilde{p}}(\mathbf{r})|^2$. Because of this feature, a negative energy electron can be regarded as the mirror image of a positive energy positron propagating in opposite direction and time.

Since the matrix elements of a quantum operator should have the same transformation property as the corresponding classical quantity, the classical field transformation (1) implies the following relation for the quantum field operator $\phi(x)$,

$$\langle\beta'|\phi(x')|\alpha'\rangle = \langle\beta|\mathbf{C}\beta\phi^*(x)|\alpha\rangle, \tag{9}$$

$$|\beta'\rangle = U_c|\beta\rangle, \quad |\alpha'\rangle = U_c|\alpha\rangle, \quad U_c^\dagger U_c = 1 \tag{10}$$

It is understood that the left- and right-hand sides of Eq. (9) refer to the new and old quantum systems characterized, respectively, by the primed and unprimed state vectors in Hilbert space. Since the transformation (1) does not alter the space-time, we have here $x' = x$. Moreover, the relation holds for arbitrary state vectors, such that the following charge conjugation transformation law for the field operator $\phi(x)$

$$C : \phi(x) \mapsto U_c\phi(x)U_c^\dagger = C\beta\phi^*(x) = C\beta\phi^{\dagger T}(x) \tag{11}$$

can be deduced immediately. Its complex conjugate reads

$$C : \phi^\dagger(x) \mapsto U_c\phi^\dagger(x)U_c^\dagger = \phi^T(x)\beta C^\dagger. \tag{12}$$

For a better understanding of the unitary operator U_c, we consider the expansion of the field operator $\phi(x)$ in terms of the solutions of the free-particle Dirac equation, viz.,

$$\phi(x)=\sum_{n,s}\left[b_{n,s}\varphi_{n,s}^{(+)}(x) + d^{n,s}\varphi_{-n,-s}^{(-)}(x)\right], \quad d^{n,s} = d_{n,s}^\dagger,\ b^{n,s} = b_{n,s}^\dagger, \tag{13}$$

where

$$\varphi_{n,s}^{(+)}(x) = N_n u(n,s)e^{-\mathbb{i}p_n\cdot x}, \quad N_n = \frac{1}{\sqrt{2E_nL^3}}, \quad E_n = \sqrt{c^2\mathbf{p}_n^2 + m^2c^4}, \tag{14}$$

$$\varphi_{-n,-s}^{(-)}(x) = N_n v(n,s)e^{\mathbb{i}p_n\cdot x}, \quad p_n\cdot x = E_nt - \mathbf{p}_n\cdot\mathbf{r}, \quad \mathbf{p}_n = \frac{2\pi}{L}(n_x, n_y, n_z), \tag{15}$$

$$u(n,s) = \sqrt{E_n + mc^2} \begin{pmatrix} 1 \\ \frac{c\boldsymbol{\sigma}\cdot\mathbf{p}_n}{E_n + mc^2} \end{pmatrix} \chi^{(s)}, \quad \chi^{(\frac{1}{2})} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi^{(-\frac{1}{2})} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{16}$$

$$v(n,s) = \sqrt{E_n + mc^2} \begin{pmatrix} \frac{c\boldsymbol{\sigma}\cdot\mathbf{p}_n}{E_n + mc^2} \\ 1 \end{pmatrix} [-\mathbb{i}\sigma_y \chi^{(s)}]. \tag{17}$$

The particular form (13) of $\phi(x)$ arises from the following considerations: (a) Only particles of positive energy are allowed in quantum field theory – $b_{n,s}$ annihilates a positive energy electron of momentum $\mathbf{p}_n$ and spin s, while $d^{n,s}$ creates a positive energy positron of momentum $\mathbf{p}_n$ and spin s. (b) That $d^{n,s}$ must go with $\varphi^{(-)}_{-n,-s}(x)$ is because the so-created positive energy positron propagates in opposite direction and time from the negative energy electron of momentum $-\mathbf{p}_n$ and spin $-s$. (c) That the operator $d^{n,s}$ (instead of $d_{n,s}$) must accompany $b_{n,s}$ is required by charge conservation. Both $b_{n,s}$ and $d^{n,s}$ increase the charge of a state by one unit; $b_{n,s}$ does this by destroying an electron, whereas $d^{n,s}$ does this by creating a positron. Thus, the field operator $\phi(x)$ always increases one unit of charge. Similarly, the field operator $\phi^\dagger(x)$ always decreases one unit of charge. Therefore, the operator $\phi^\dagger(x)\phi(x)$ conserves the charge. Had $d_{n,s}$ been chosen to accompany $b_{n,s}$, the operator $\phi^\dagger(x)\phi(x)$ would not conserve the charge: it would include terms like $b^{n,s}d_{n,s}$ and $d^{n,s}b_{n,s}$ which decrease and increase two units of charge, respectively.

Since the U_c operator acts on the state vectors (e.g., $|\alpha\rangle = \Pi_{n,s} b^{n,s}|0;\tilde{0}\rangle$) in Hilbert space, it is a function only of the fermion creation and annihilation operators and hence does not act on the Dirac matrices or spinors. Inserting the expression (13) into Eq. (11) leads to

$$U_c b_{n,s} U_c^\dagger \varphi^{(+)}_{n,s} + U_c d^{n,s} U_c^\dagger b^p \varphi^{(-)}_{-n,-s} = b^{n,s}\mathbf{C}\beta\varphi^{(+)*}_{n,s} + d_{n,s}\mathbf{C}\beta\varphi^{(-)*}_{-n,-s}. \tag{18}$$

It is straightforward to show that

$$\mathbf{C}\beta v^*(n,s) = u(n,s), \tag{19}$$

$$\mathbf{C}\beta u^*(n,s) = v(n,s), \tag{20}$$

and hence

$$\mathbf{C}\beta\varphi^{(-)*}_{-n,-s} = \varphi^{(+)}_{n,s}, \tag{21}$$

$$\mathbf{C}\beta\varphi^{(+)*}_{n,s} = \varphi^{(-)}_{-n,-s}. \tag{22}$$

Therefore, Eq. (18) implies the following relations

$$U_c b_{n,s} U_c^\dagger = d_{n,s}, \quad U_c d_{n,s} U_c^\dagger = b_{n,s}, \tag{23}$$

$$U_c b^{n,s} U_c^\dagger = d^{n,s}, \quad U_c d^{n,s} U_c^\dagger = b^{n,s}. \tag{24}$$

Relation (23) suggests the following first try [24] for U_c

$$U_1 = e^{\mathbb{i}\theta C_1}, \tag{25}$$

$$C_1 = \sum_{n,s} (b^{n,s} d_{n,s} + d^{n,s} b_{n,s}) = C_1^\dagger. \tag{26}$$

By virtue of the following commutation relations,

$$\begin{aligned} [C_1, b_{n,s}] = -d_{n,s}, \quad [C_1, d_{n,s}] = -b_{n,s}, \\ [C_1, [C_1, b_{n,s}]] = b_{n,s}, \quad [C_1, [C_1, d_{n,s}]] = d_{n,s}, \ldots, \end{aligned} \tag{27}$$

as well as the Baker-Campbell-Hausdorff identity

$$e^A B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \cdots, \tag{28}$$

we can readily find

$$\begin{aligned} U_1 b_{n,s} U_1^\dagger &= b_{n,s} \left[1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 + \cdots \right] \\ &\quad -\mathbb{i}\, d_{n,s} \left[\theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 + \cdots \right] \\ &= b_{n,s} \cos\theta - \mathbb{i} d_{n,s} \sin\theta. \end{aligned} \tag{29}$$

The choice of $\theta = \frac{\pi}{2}$ leads to

$$U_1 b_{n,s} U_1^\dagger = -\mathbb{i} d_{n,s}. \tag{30}$$

Similarly, we have

$$U_1 d_{n,s} U_1^\dagger = -\mathbb{i} b_{n,s}. \tag{31}$$

A second unitary transformation U_2 is needed to just cancel the phase factor yet without changing the nature of the fermion operators, i.e.,

$$U_2(-\mathbb{i} d_{n,s}) U_2^\dagger = d_{n,s}, \quad U_2(-\mathbb{i} b_{n,s}) U_2^\dagger = b_{n,s}. \tag{32}$$

A rather obvious choice is

$$U_2 = e^{\mathbb{i}\delta C_2}, \tag{33}$$

$$C_2 = \sum_{n,s} (b^{n,s} b_{n,s} + d^{n,s} d_{n,s}) = C_2^\dagger. \tag{34}$$

In view of the commutation relations

$$\begin{aligned} [C_2, b_{n,s}] &= -b_{n,s}, \quad [C_1, d_{n,s}] = -d_{n,s}, \\ [C_2, [C_2, b_{n,s}]] &= b_{n,s}, \quad [C_2, [C_2, d_{n,s}]] = d_{n,s}, \ldots, \end{aligned} \tag{35}$$

we have

$$U_2(-\mathbb{i}d_{n,s})U_2^\dagger = -\mathbb{i}d_{n,s}e^{-\mathbb{i}\delta}, \quad U_2(-\mathbb{i}b_{n,s})U_2^\dagger = -\mathbb{i}b_{n,s}e^{-\mathbb{i}\delta}. \tag{36}$$

The requirements (32) are hence fulfilled by choosing $\delta = -\frac{\pi}{2}$. The overall U_c operator can then be written as [24]

$$U_c = U_2U_1 = e^{-\mathbb{i}\frac{\pi}{2}C_2}e^{\mathbb{i}\frac{\pi}{2}C_1} = e^{\mathbb{i}\frac{\pi}{2}(C_1-C_2)}. \tag{37}$$

Therefore, an explicit construction of the quantum field charge conjugation U_c is indeed possible. However, it should be kept in mind that this particular form (37) of U_c holds only in the free-particle representation, at variance with the transformation law (11) which is independent of any representation. The free-particle solutions are of course hardly useful for describing bound states. Instead, the PES and NES of the Dirac equation

$$(D + U)\varphi_p = \epsilon_p\varphi_p, \tag{38}$$

$$D = D_0 - e\phi_{\text{ext}}, \quad D_0 = c\boldsymbol{\alpha}\cdot\mathbf{p} + \beta mc^2 \tag{39}$$

should actually be used. Here, $\phi_{\text{ext}}(\mathbf{r})$ is usually the field of the clamped nuclei and $U(\mathbf{r})$ a screening potential. In this case, relations (23) and (24) no longer hold. However, since any operator must be expanded in a complete basis spanned by the PES and NES of the same Dirac equation, we still ought to write the field operator $\phi(x)$ in the interaction (and particle-hole) picture as

$$\begin{aligned} \phi_\rho(x) &= b_p\varphi_{p,\rho}(\mathbf{r})e^{-i\epsilon_p t} + b^{\tilde{p}}\varphi_{\tilde{p},\rho}(\mathbf{r})e^{-i\epsilon_{\tilde{p}}t}, \quad \rho = 1, \ldots, 4, \\ b_p|0;\tilde{0}\rangle &= b_{\tilde{p}}|0;\tilde{0}\rangle = 0, \quad b^p = b_p^\dagger, \quad b^{\tilde{p}} = b_{\tilde{p}}^\dagger, \quad \epsilon_p > 0, \quad \epsilon_{\tilde{p}} < 0, \end{aligned} \tag{40}$$

with the usual interpretation, i.e., b_p (b^p) annihilates (creates) an electron of positive energy ϵ_p, whereas $b^{\tilde{p}}$ ($b_{\tilde{p}}$) creates (annihilates) a positron of positive energy $|\epsilon_{\tilde{p}}| = -\epsilon_{\tilde{p}}$. The electron (positron) here is a quasiparticle when referring to free particles: it is a composite of a free electron (positron) and a series of free electron-positron pairs, rather than being a bare particle. In general, a particle moving in a potential is composed of a series of pairs moving in another potential.

Diagrammatic Derivation of the eQED Hamiltonian

The point of departure is the symmetrized [25] 4-current J_s^μ

$$J_s^\mu(x) = -\frac{1}{2}ec\alpha^\mu_{\rho\sigma}\left[\phi^\dagger_\rho(x), \phi_\sigma(x)\right], \quad e = +1, \quad \alpha^\mu = \left(c^{-1}, \boldsymbol{\alpha}\right), \tag{41}$$

$$= J^\mu(x) - e\langle 0;\tilde{0}|\frac{1}{2}[\phi^\dagger(x), c\alpha^\mu\phi(x)]|0;\tilde{0}\rangle, \tag{42}$$

where $J^\mu(x)$ is the usual definition of the 4-current in the QED literature,

$$J^\mu(x) = -e\left\{\phi^\dagger(x)c\alpha^\mu\phi(x)\right\}. \tag{43}$$

Here, the curly brackets indicate normal ordering with respect to the vacuum $|0;\tilde{0}\rangle$ of no particles or holes. Note that the two forms of the 4-current become identical only for free particles; for the second, vacuum term of Eq. (42) then vanishes pointwise. For bound states, it is the symmetrized form (41) of the 4-current that should be used. Using relations (11) and (12) and the fact that the unitary U_c operator therein does not act on the Dirac matrices or spinors, it is straightforward to show that

$$\begin{aligned} U_c J_s^\mu(x) U_c^\dagger &= -\frac{1}{2}ec\alpha^\mu_{\rho\sigma}\left[\phi^T_\gamma(x)\left(\beta\mathbf{C}^\dagger\right)_{\gamma\rho}, (\mathbf{C}\beta)_{\sigma\tau}\phi^{\dagger T}_\tau(x)\right] \\ &= -\frac{1}{2}ec\left(\beta\mathbf{C}^\dagger\alpha^\mu\mathbf{C}\beta\right)_{\gamma\tau}\left[\phi^T_\gamma(x), \phi^{\dagger T}_\tau(x)\right] \\ &= -\frac{1}{2}ec\alpha^\mu_{\tau\gamma}\left[\phi_\gamma(x), \phi^\dagger_\tau(x)\right] \\ &= -J_s^\mu(x). \end{aligned} \tag{44}$$

It is hence clear that the definition (41) of the 4-current is actually dictated by charge conjugation, which interchanges particles and antiparticles, so as to reverse the charge current.

By virtue of the identity for time ordering of fermion operators

$$T\left[AB\frac{1}{2}(CD - DC)EF\cdots\right] = T[ABCDEF\cdots], \tag{45}$$

we can rewrite the interaction Hamiltonian density

$$\mathscr{H}(x) = T\left[J_s^\mu(x)A_\mu(x)\right], \quad A_\mu = (\phi_{\text{ext}}, -\mathbf{A}_{\text{ext}}) \tag{46}$$

in terms of the unnormal-ordered 4-current operator J_a^μ, viz.,

$$\mathscr{H}(x) = T\left[J_a^\mu A_\mu(x)\right], \quad J_a^\mu = -e\phi^\dagger(x)c\alpha^\mu\phi(x) \tag{47}$$

$$= T\left[\phi^\dagger(x)\left(V_{\rm ext}(x) + ec\boldsymbol{\alpha}\cdot\mathbf{A}_{\rm ext}(x)\right)\phi(x)\right], \quad V_{\rm ext} = -e\phi_{\rm ext}. \tag{48}$$

At this point, it is instructive to take a look at the so-called equal-time contraction (ETC) [25] of fermion operators,

$$\overbrace{A(t)B}(t) = \langle 0;\tilde{0}|T[A(t)B(t)|0;\tilde{0}\rangle \tag{49}$$

$$\triangleq \frac{1}{2}\langle 0;\tilde{0}|T[A(t)B(t')]|0;\tilde{0}\rangle|_{t'-t\to 0^\pm} \tag{50}$$

$$= \langle 0;\tilde{0}|\frac{1}{2}[A(t),B(t)]|0;\tilde{0}\rangle, \tag{51}$$

which is symmetric in time. That is, the two expressions $A(t)B(t)$ and $-B(t)A(t)$ obtained by letting t' approach t from the past and future are both considered and averaged here. Note that the ETC (51) is taken into account automatically by the Feynman electron propagator

$$S_F(x_2,x_1) = -\mathbb{i}\langle 0;\tilde{0}|T[\phi(x_2)\phi^\dagger(x_1)]|0;\tilde{0}\rangle \tag{52}$$

$$= -\mathbb{i}\langle 0;\tilde{0}|\Theta(t_2-t_1)\phi(x_2)\phi^\dagger(x_1) - \Theta(t_1-t_2)\phi^\dagger(x_1)\phi(x_2)|0;\tilde{0}\rangle \tag{53}$$

$$= \int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}S_F(\omega;\mathbf{r}_2,\mathbf{r}_1)e^{-i\omega(t_2-t_1)}, \tag{54}$$

$$S_F(\omega;\mathbf{r}_2,\mathbf{r}_1) = \frac{\varphi_r(\mathbf{r}_2)\varphi_r^\dagger(\mathbf{r}_1)}{\omega-\epsilon_r(1-i\eta)}, \quad r\in \text{PES, NES}. \tag{55}$$

To see this, we can calculate the equal-time propagator $S_F(\mathbf{r}_2t,\mathbf{r}_1t)$ either by means of Eq. (54),

$$S_F(\mathbf{r}_2t,\mathbf{r}_1t) = \varphi_r(\mathbf{r}_2)\varphi_r^\dagger(\mathbf{r}_1)\int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}\frac{1}{\omega-\epsilon_r(1-i\eta)}$$

$$= \frac{1}{2\mathbb{i}}\varphi_r(\mathbf{r}_2)\varphi_r^\dagger(\mathbf{r}_1)\text{sgn}(\epsilon_r), \tag{56}$$

or by means of the definition (52) and the ETC (51),

$$S_F(\mathbf{r}_2t,\mathbf{r}_1t) = -\mathbb{i}\langle 0;\tilde{0}|\frac{1}{2}\left[\phi(\mathbf{r}_2t),\phi^\dagger(\mathbf{r}_1t)\right]|0;\tilde{0}\rangle \tag{57}$$

$$= \frac{1}{2\mathbb{i}}\langle 0;\tilde{0}|\varphi_p(\mathbf{r}_2)\varphi_q^\dagger(\mathbf{r}_1)e^{-\mathbb{i}(\epsilon_p-\epsilon_q)t}b_pb^q - \varphi_{\tilde{p}}(\mathbf{r}_2)\varphi_{\tilde{q}}^\dagger(\mathbf{r}_1)e^{-\mathbb{i}(\epsilon_{\tilde{p}}-\epsilon_{\tilde{q}})t}b_{\tilde{p}}b^{\tilde{q}}|0;\tilde{0}\rangle \tag{58}$$

$$= \frac{1}{2\mathbb{i}}\left[\varphi_p(\mathbf{r}_2)\varphi_p^\dagger(\mathbf{r}_1) - \varphi_{\tilde{p}}(\mathbf{r}_2)\varphi_{\tilde{p}}^\dagger(\mathbf{r}_1)\right]. \tag{59}$$

The two expressions (56) and (59) are obviously identical. The vacuum polarization density as a result of the ETC (51) can be calculated as

$$\rho_{vp}(\mathbf{r}) = e\mathbb{i}\mathrm{Tr}[S_F(x,x)] = \frac{e}{2}\varphi_r^\dagger(\mathbf{r})\varphi_r(\mathbf{r})\mathrm{sgn}(\epsilon_r) = -\frac{e}{2}[n_-(\mathbf{r}) - n_+(\mathbf{r})], \tag{60}$$

$$n_+(\mathbf{r}) = \sum_p \varphi_p^\dagger(\mathbf{r})\varphi_p(\mathbf{r}), \quad n_-(\mathbf{r}) = \sum_{\tilde{p}} \varphi_{\tilde{p}}^\dagger(\mathbf{r})\varphi_{\tilde{p}}(\mathbf{r}), \tag{61}$$

where $n_+(\mathbf{r})$ and $n_-(\mathbf{r})$ are the number densities of the PES and NES, respectively. By virtue of the identity $n_+ + n_- = \bar{n}_+ + \bar{n}_- = 2\bar{n}_-$, with $\bar{n}_+$ and $\bar{n}_-$ $(= \bar{n}_+)$ being the free-particle number densities, we have

$$\rho_{vp}(\mathbf{r}) = -e[n_-(\mathbf{r}) - \bar{n}_-(\mathbf{r})], \tag{62}$$

which is clearly the charge polarization of the vacuum. Equation (60) implies that the NES are occupied by electrons e^- with charge -1, whereas the PES by positrons e^+ with charge $+1$. Since the two pictures are equivalent thanks to charge conjugation symmetry, they should be averaged. Otherwise, we would obtain twice the polarization density (62). Formally, the charge density $\rho_{vp}(\mathbf{r})$ (60) corresponds [6] to an ensemble Ψ_{vp} of noninteracting N_{e^-} electrons and N_{e^+} positrons

$$\Psi_{vp} = \frac{1}{\sqrt{2}}(|0; N_{e^-}\rangle + |N_{e^+}; 0\rangle), \tag{63}$$

the first term of which corresponds to a filled negative energy sea, whereas the second to a filled positive energy sea as a result of charge conjugation. The picture remains the same when the electrons and positrons interact with each other, for charge conjugation symmetry still holds in the absence of an external field.

After the above background, we are ready to derive diagrammatically the desired "with-pair, no-retardation" relativistic many-body Hamiltonian. By this we mean that only systems of electrons subject to the instantaneous Coulomb/Gaunt/Breit interaction are under consideration. Inserting the field (40) into the Hamiltonian density (47) gives rise to four terms, which can be represented by the diagrams shown in Fig. 1. Here, the up- and down-going lines represent electrons (particles) and positrons (holes), respectively, whereas the horizontal wavy lines represent photons. Figure 1a, b describe scattering of electrons and positrons, respectively, while Fig. 1c, d describe virtual pair creation and annihilation, respectively. Since we are not interested in real photons, the photon lines of the diagrams must always be contracted. The photon-only contractions lead to 16 two-body terms shown in Fig. 2. When both photon and fermion lines are contracted, we get 16 one-body diagrams shown in Fig. 3. Had the normal-ordered expression (43) been used for the 4-current, only fermion lines of different vertices can be contracted, thereby excluding the first 8 diagrams in Fig. 3 stemming from the fermion contractions at the same vertex. As shown before, this is only correct for free Dirac particles. Note also that the zero body terms, i.e., those closed bubble diagrams obtained by connecting all the photon and fermion lines, are unphysical and will be

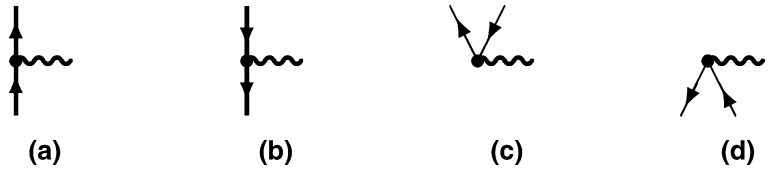


Fig. 1 The Hamiltonian density

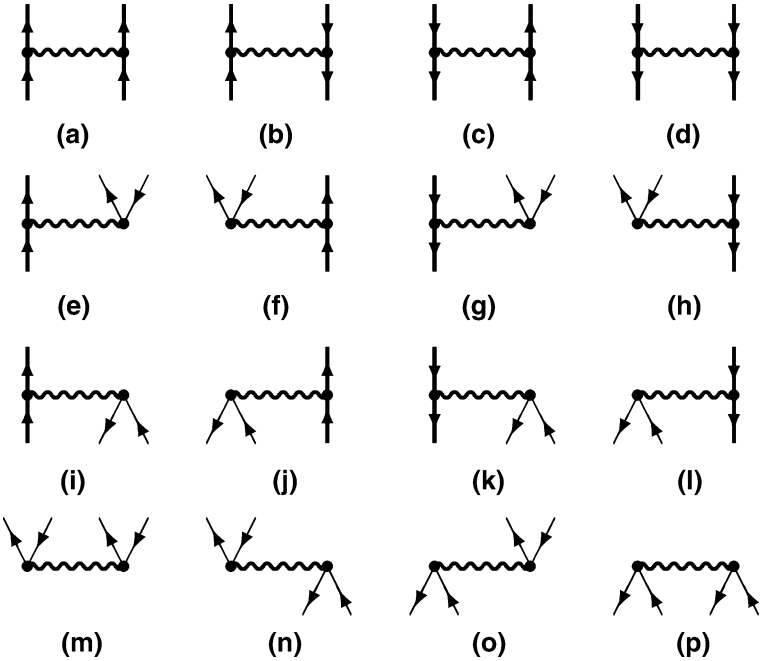


Fig. 2 The two-body terms from photon-only contractions (Reprinted with permission from Ref. [6]. ©2014, Wiley Periodicals, Inc.)

"renormalized away." The two-body "time-ordered" Goldstone-like diagrams in Fig. 2 can be represented by a single "time-unordered" Feynman diagram in Fig. 4a. Likewise, the first and second 8 diagrams in Fig. 3 can be represented by the EVP diagram Fig. 4b and the ESE diagram Fig. 4c, respectively. It is obvious that only diagram Fig. 4a could be obtained if the fermion contractions are not allowed at all [18, 19].

The question now is how to evaluate such diagrams before knowing the algebraic expressions. This is actually very simple for instantaneous interactions. Consider first the first two diagrams in Fig. 3. Following the standard rules for evaluating Goldstone diagrams [26], we have $-g_{ps}^{qs}$ ($\epsilon_s > 0$) for Fig. 3a (one loop, zero hole) and $+g_{ps}^{qs}$ ($\epsilon_s < 0$) for Fig. 3b (one loop, one hole). Further in view of the "averaging picture" of Eq. (60) or (63), these two terms should be averaged, thereby leading to

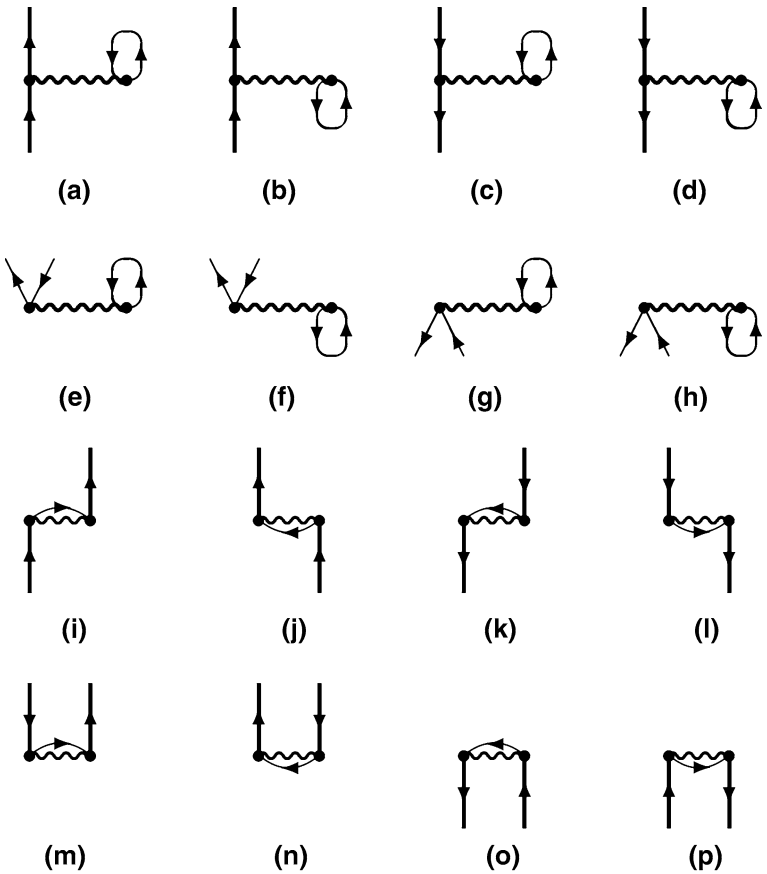


Fig. 3 The one-body terms from both photon and fermion contractions (Reprinted with permission from Ref. [6]. ©2014, Wiley Periodicals, Inc.)

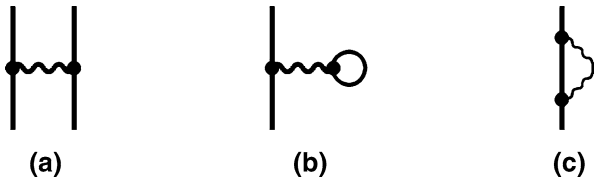


Fig. 4 The (time unordered) first-order Feynman diagrams

$$\tilde{Q}[3(a,b)] = -\frac{1}{2} g_{ps}^{qs} \mathrm{sgn}(\epsilon_s) \left\{ b^p b_q \right\}, \quad p,q \in \mathrm{PES}, \quad g_{pq}^{rs} = \langle pq|g(1,2)|rs\rangle, \tag{64}$$

for the first two diagrams of Fig. 3 (NB: it is the probability $(1/\sqrt{2})^2$ of the components of Ψ_{vp} (63) that enters the energy and hence the potential operator). The 9th and 10th diagrams of Fig. 3 are just the exchange of the first and second, respectively. Therefore, their sum read

$$\bar{Q}[3(i,j)] = \frac{1}{2} g_{ps}^{s\,q} \mathrm{sgn}(\epsilon_s) \left\{ b^p b_q \right\}, \quad p,q \in \mathrm{PES}. \tag{65}$$

The remaining diagrams in Fig. 3 differ from the first two only in the open fermion lines and are therefore of the same algebraic forms. Summing up all the 16 terms leads to

$$\tilde{Q} = -\frac{1}{2} g_{ps}^{qs} \mathrm{sgn}(\epsilon_s) \left\{ a_q^p \right\}, \quad p,q \in \mathrm{PES, NES}, \quad a_q^p = a^p a_q = a_p^\dagger a_q, \tag{66}$$

$$\bar{Q} = \frac{1}{2} g_{ps}^{s\,q} \mathrm{sgn}(\epsilon_s) \left\{ a_q^p \right\}, \quad p,q \in \mathrm{PES, NES}, \tag{67}$$

$$Q = \tilde{Q} + \bar{Q} = -\frac{1}{2} \bar{g}_{ps}^{qs} \mathrm{sgn}(\epsilon_s) \left\{ a^p a_q \right\}, \quad p,q \in \mathrm{PES}, \quad \bar{g}_{pq}^{rs} = g_{pq}^{rs} - g_{pq}^{sr}. \tag{68}$$

The effective Q potential (68) agrees with that in Ref. [20] derived algebraically (see also section "Algebraic Derivation of the eQED Hamiltonian"). However, the weight factor of $1/2$ has been introduced here in an ad hoc manner, for the weight factor of these (asymmetric) diagrams should be 1 instead of $1/2$ according to the standard rules. The algebraic expression for Fig. 4a reads

$$G = \frac{1}{2} g_{pq}^{rs} \left\{ a_{rs}^{pq} \right\}, \quad a_{rs}^{pq} = a^p a^q a_s a_r, \tag{69}$$

where the prefactor arises from the reflection symmetry of the diagram. Further collecting the one-electron Dirac operator D (39), the total Hamiltonian normal ordered with respect to $|0;\tilde{0}\rangle$ reads

$$H_b = H_b^{\mathrm{FS}} + Q_p^q \left\{ a_q^p \right\}, \tag{70}$$

$$H_b^{\mathrm{FS}} = D_p^q \left\{ a_q^p \right\} + \frac{1}{2} g_{pq}^{rs} \left\{ a_{rs}^{pq} \right\}, \tag{71}$$

where H_b^{FS} (71) is the so-called Fock space Hamiltonian [18, 19]. Note that Eqs. (66)–(71) assume implicitly the particle-hole picture, viz.,

$$a_p = b_p, \quad a^p = b^p, \quad \forall \epsilon_p > 0; \quad a_p = b^p, \quad a^p = b_p, \quad \forall \epsilon_p < 0. \tag{72}$$

Algebraic Derivation of the eQED Hamiltonian

The point of departure here is the second-quantized DC/DCG/DCB Hamiltonian

$$\mathscr{H} = D_p^q a_q^p + \frac{1}{2} g_{pq}^{rs} a_{rs}^{pq}, \tag{73}$$

which is already normal ordered with respect to the genuine vacuum $|0;\tilde{0}\rangle$, viz., $a_p|0;\tilde{0}\rangle = 0$, $p \in$ PES, NES. Note that the particle-hole picture (72) is not used hereafter. As it stands, the Hamiltonian $\mathscr{H}$ (73) does not distinguish the empty from the filled Dirac picture. While the former is unphysical (vide post), the latter can formally be imposed in a finite basis representation by setting the Fermi level below the energetically lowest of the $\tilde{N}$ occupied NES. The physical energy of an N-electron state is then defined [7] as the difference between those of states $\Psi(N;\tilde{N})$ and $\Psi(0;\tilde{N})$, viz.,

$$E = \langle\Psi(N;\tilde{N})|\mathscr{H}|\Psi(N;\tilde{N})\rangle - \langle\Psi(0;\tilde{N})|\mathscr{H}|\Psi(0;\tilde{N})\rangle. \tag{74}$$

To do so, we can shift up the Fermi level to zero. This amounts to normal ordering the Hamiltonian $\mathscr{H}$ (73) with respect to the noninteracting, floating vacuum $|0;\tilde{N}\rangle$ of zero positive energy electrons and $\tilde{N}$ negative energy electrons (The term "floating" here means that the same NES are used to construct the noninteracting $|0;\tilde{N}\rangle$ and the interacting $\Psi(0;\tilde{N})$ and $\Psi(N;\tilde{N})$ states. For comparison, a frozen vacuum $|0;\bar{N}\rangle$ would employ, e.g., the free-particle NES that are different from those for $\Psi(0;\tilde{N})$ and $\Psi(N;\tilde{N})$. This is just a different setting of the zero-point energy. For more discussions, see Ref. [21].). Here, the charge-conjugated contraction (CCC) [20] of fermion operators must be invoked, viz.,

$$\overbrace{a^p a_q} = \langle 0;\tilde{N}|\frac{1}{2}[a^p, a_q]|0;\tilde{N}\rangle, \quad p,q \in \text{PES, NES} \tag{75}$$

$$= \frac{1}{2}\langle 0;\tilde{N}|a^{\tilde{p}}a_{\tilde{q}}|0;\tilde{N}\rangle|_{\epsilon_{\tilde{p}}<0,\epsilon_{\tilde{q}}<0} - \frac{1}{2}\langle 0;\tilde{N}|a_q a^p|0;\tilde{N}\rangle|_{\epsilon_p>0,\epsilon_q>0} \tag{76}$$

$$= -\frac{1}{2}\delta_q^p \text{sgn}(\epsilon_q), \quad p,q \in \text{PES, NES}, \tag{77}$$

in terms of which we have

$$a_q^p = \{a^p a_q\}_n - \frac{1}{2}\delta_q^p \text{sgn}(\epsilon_q), \quad p,q \in \text{PES, NES}, \tag{78}$$

$$D_p^q a_q^p = D_p^q \{a^p a_q\}_n + C_{1n}, \quad C_{1n} = -\frac{1}{2}D_p^p \text{sgn}(\epsilon_p), \tag{79}$$

where the subscript n of the curly brackets emphasizes that the normal ordering is taken with respect to the reference $|0;\tilde{N}\rangle$. More specifically,

$$\{a^p a_q\}_n = \begin{cases} a^p a_q, & \epsilon_p > 0, \epsilon_q > 0, \\ a^p a_q, & \epsilon_p > 0, \epsilon_q < 0, \\ a^p a_q, & \epsilon_p < 0, \epsilon_q > 0, \\ -a_q a^p, & \epsilon_p < 0, \epsilon_q < 0. \end{cases} \tag{80}$$

By applying the relation (77) repeatedly, we obtain

$$a_{rs}^{pq} = \left\{a_{rs}^{pq}\right\}_n - \frac{1}{2}\left\{\delta_r^p a_s^q \mathrm{sgn}(\epsilon_r) + \delta_s^q a_r^p \mathrm{sgn}(\epsilon_s) - \delta_r^q a_s^p \mathrm{sgn}(\epsilon_r) - \delta_s^p a_r^q \mathrm{sgn}(\epsilon_s)\right\}_n + \frac{1}{4}\left(\delta_r^p \delta_s^q - \delta_r^q \delta_s^p\right) \mathrm{sgn}(\epsilon_r)\mathrm{sgn}(\epsilon_s), \tag{81}$$

and hence

$$\frac{1}{2} g_{pq}^{rs} a_{rs}^{pq} = \frac{1}{2} g_{pq}^{rs} \left\{a_{rs}^{pq}\right\}_n + Q_p^q \left\{a_q^p\right\}_n + C_{2n}, \tag{82}$$

$$Q_p^q = \tilde{Q}_p^q + \bar{Q}_p^q = -\frac{1}{2} \bar{g}_{ps}^{qs} \mathrm{sgn}(\epsilon_s), \tag{83}$$

$$\tilde{Q}_p^q = -\frac{1}{2} g_{ps}^{qs} \mathrm{sgn}(\epsilon_s), \tag{84}$$

$$\bar{Q}_p^q = \frac{1}{2} g_{ps}^{sq} \mathrm{sgn}(\epsilon_s), \tag{85}$$

$$C_{2n} = \frac{1}{8} \bar{g}_{pq}^{pq} \mathrm{sgn}(\epsilon_p)\mathrm{sgn}(\epsilon_q) = -\frac{1}{4} Q_p^p \mathrm{sgn}(\epsilon_p). \tag{86}$$

Note that the implicit summations in C_{1n} (79), $\tilde{Q}$ (84), $\bar{Q}$ (85), and C_{2n} (86) include all the PES and NES, whether occupied or not. The Hamiltonian $\mathscr{H}$ (73) can then be rewritten as

$$\mathscr{H} = H_a + C_n, \tag{87}$$

$$H_a = H_a^{\mathrm{FS}} + Q_p^q \left\{a_q^p\right\}_n, \tag{88}$$

$$H_a^{\mathrm{FS}} = D_p^q \left\{a_q^p\right\}_n + \frac{1}{2} g_{pq}^{rs} \left\{a_{rs}^{pq}\right\}_n, \tag{89}$$

$$C_n = C_{1n} + C_{2n} = \langle 0; \tilde{N} | H | 0; \tilde{N} \rangle = -\frac{1}{2} D_p^p \mathrm{sgn}(\epsilon_p) - \frac{1}{4} Q_p^p \mathrm{sgn}(\epsilon_p). \tag{90}$$

By introducing the particle-hole picture (72), viz.,

$$\left\{a^p a_q\right\}_n \to \begin{cases} b^p b_q, & \epsilon_p > 0, \epsilon_q > 0, \\ b^p b^q, & \epsilon_p > 0, \epsilon_q < 0, \\ b_p b_q, & \epsilon_p < 0, \epsilon_q > 0, \\ -b^q b_p, & \epsilon_p < 0, \epsilon_q < 0, \end{cases} \tag{91}$$

the Hamiltonian H_a (88), normal ordered with respect to the floating vacuum $|0; \tilde{N}\rangle$, will be converted to the Hamiltonian H_b (70) normal ordered with respect to the genuine vacuum $|0; \tilde{0}\rangle$. Therefore, we have arrived at the same eQED Hamiltonian

through both diagrammatic and algebraic procedures. It is clear that the particle-hole formalism (91) merely converts the reference $|0;\tilde{N}\rangle$ to the mathematically equivalent $|0;\tilde{0}\rangle$ and does not introduce new physics. It is useful only when going to the diagrammatic presentation, as shown in section "Diagrammatic Derivation of the eQED Hamiltonian."

In short, the filled Dirac picture, second quantization, and normal ordering in conjunction with the CCC [20] are the essential ingredients for constructing the eQED Hamiltonian in a bottom-up fashion. Had the standard contraction of fermion operators, viz.,

$$\underbrace{a^p a_q} = \langle 0;\tilde{N}|a^p a_q|0;\tilde{N}\rangle = \delta^{\tilde{p}}_{\tilde{q}} n_{\tilde{q}}, \tag{92}$$

been taken, we would obtain the following Hamiltonian

$$H_a^{\rm CS} = H_a^{\rm FS} + \bar{g}^{q\tilde{j}}_{p\tilde{j}} \left\{a^p_q\right\}_n, \tag{93}$$

where the mean-field potential $\bar{g}^{q\tilde{j}}_{p\tilde{j}}$ arises from the occupied NES alone, which obviously breaks the charge conjugation symmetry. Actually, $H_a^{\rm CS}$ (93) is more akin to the second-quantized Schrödinger-Coulomb Hamiltonian if the NES $\tilde{j}$ were regarded as the occupied Hartree-Fock (HF) orbitals. It can readily be shown that even the first-order energy of $H_a^{\rm CS}$ (93), i.e.,

$$\tilde{E}^{[1]} = \langle N;\tilde{N}|H_a^{\rm CS}|N;\tilde{N}\rangle = E^{[1]}_{\rm np} + \bar{g}^{i\tilde{j}}_{i\tilde{j}}, \tag{94}$$

$$E^{[1]}_{\rm np} = \langle N;\tilde{N}|H_a^{\rm FS}|N;\tilde{N}\rangle = \left(D + \frac{1}{2}V_{\rm HF}\right)^i_i, \quad (V_{\rm HF})^q_p = \bar{g}^{qj}_{pj}, \tag{95}$$

is already different from that of H_a (88),

$$E^{[1]} = \langle N;\tilde{N}|H_a|N;\tilde{N}\rangle = E^{[1]}_{\rm np} - \frac{1}{2}\bar{g}^{is}_{is}{\rm sgn}(\epsilon_s) = E^{[1]}_{\rm np} + Q^i_i. \tag{96}$$

In essence, the CCC (75) is the time-independent equivalent of the ETC (51), whereas the standard contraction (92) corresponds to the following ETC

$$\underbrace{A(t)B}(t) = \lim_{\eta\to 0^+}\langle 0;\tilde{0}|T[A(t)B(t+\eta)]|0;\tilde{0}\rangle, \tag{97}$$

which is obviously asymmetric in time and holds only in the nonrelativistic limit. It follows that the standard contraction (92) even of relativistic operators will result in wrong, nonrelativistic type of expressions.

Likewise, we can also set the Fermi level (denoted as ϵ_F) just above the highest occupied PES, i.e., normal order the Hamiltonian $\mathscr{H}$ (73) with respect to the noninteracting reference $|N;\tilde{N}\rangle$, which is just the zero order of $\Psi(N;\tilde{N})$.

Following the same procedure, we obtain [21]

$$\mathscr{H} = H_F + C_F, \tag{98}$$

$$H_F = D_p^q \left\{a_q^p\right\}_F + Y_p^q \left\{a_q^p\right\}_F + \frac{1}{2} g_{pq}^{rs} \left\{a_{rs}^{pq}\right\}_F, \tag{99}$$

$$Y_p^q = \tilde{Y}_p^q + \bar{Y}_p^q = -\frac{1}{2} \bar{g}_{p\omega}^{q\omega} \mathrm{sgn}(\epsilon_\omega - \epsilon_F), \tag{100}$$

$$\tilde{Y}_p^q = -\frac{1}{2} g_{p\omega}^{q\omega} \mathrm{sgn}(\epsilon_\omega - \epsilon_F), \tag{101}$$

$$\bar{Y}_p^q = \frac{1}{2} g_{\omega p}^{q\omega} \mathrm{sgn}(\epsilon_\omega - \epsilon_F), \tag{102}$$

$$\begin{aligned} C_F &= \langle N; \tilde{N} | H | N; \tilde{N} \rangle = -\frac{1}{2} D_p^p \mathrm{sgn}(\epsilon_p - \epsilon_F) - \frac{1}{4} Y_p^p \mathrm{sgn}(\epsilon_p - \epsilon_F) \\ &= C_n + E^{[1]}, \end{aligned} \tag{103}$$

where C_n and $E^{[1]}$ are given in Eqs. (90) and (96), respectively. In view of the identity

$$Y_p^q = Q_p^q + (V_{\mathrm{HF}})_p^q, \tag{104}$$

the Hamiltonian H_F (99) can further be decomposed as

$$H_F = H_F^{\mathrm{FS}} + Q_p^q \left\{a_q^p\right\}_F, \tag{105}$$

$$H_F^{\mathrm{FS}} = (F_e)_p^q \left\{a_q^p\right\}_F + \frac{1}{2} g_{pq}^{rs} \left\{a_{\mathrm{rs}}^{pq}\right\}_F, \tag{106}$$

$$(F_e)_p^q = D_p^q + (V_{\mathrm{HF}})_p^q. \tag{107}$$

It is interesting to note that the standard contraction of the Fermion operators, i.e.,

$$\underbrace{a^p a_q} = \langle N; \tilde{N} | \left\{a^p a_q\right\}_n | N; \tilde{N} \rangle = \langle N; 0 | a^p a_q | N; 0 \rangle = \delta_q^p n_q, \quad \epsilon_q > 0, \tag{108}$$

should be invoked when going from H_a (88) to H_F (99). More specifically,

$$D_p^q \left\{a_q^p\right\}_n = D_p^q \left\{a_q^p\right\}_F + D_i^i, \tag{109}$$

$$Q_p^q \left\{a_q^p\right\}_n = Q_p^q \left\{a_q^p\right\}_F + Q_i^i, \tag{110}$$

$$\frac{1}{2} g_{pq}^{rs} \left\{a_{rs}^{pq}\right\}_n = \frac{1}{2} g_{pq}^{rs} \left\{a_{rs}^{pq}\right\}_F + (V_{\mathrm{HF}})_p^q \left\{a_q^p\right\}_F + \frac{1}{2} (V_{\mathrm{HF}})_i^i. \tag{111}$$

Consequently, we have

$$H_a = H_F + E^{[1]}, \tag{112}$$

$$H_a^{\mathrm{FS}} = H_F^{\mathrm{FS}} + E_{\mathrm{np}}^{[1]}. \tag{113}$$

That is, H_a (88) differs from H_F (99) by a constant $E^{[1]}$ (96), which is just the first-order QED energy (see Fig. 4). Likewise, H_a^{FS} (89) differs from H_F^{FS} (106) also by a constant $E_{\mathrm{np}}^{[1]}$ (95), which is nothing but the no-pair first-order energy (see Fig. 4a).

The Model *Q* Operator

The Coulomb-only $\tilde{Q}$ term (84) is the full EVP and can itself be charge renormalized. However, the $\bar{Q}$ term (85) represents the Coulomb-only ESE and remains logarithmically divergent even after mass renormalization. While such type of divergence is hardly a serious issue in a finite Gaussian representation furnishing a natural regularization, it is strongly recommended to combine the Coulomb- and transverse-photon contributions to the ESE that can together be renormalized. In the Coulomb gauge adopted here, the transverse part of the ESE reads [27]

$$(\bar{Q}_T)_p^q = \frac{1}{2}\langle p|\Sigma_T^C(\epsilon_p) + \Sigma_T^C(\epsilon_q)|q\rangle, \tag{114}$$

$$\langle p|\Sigma_T^C(\epsilon_p)|q\rangle = \langle ps|\int_0^\infty \frac{c\,dk\,f_T^C(k,r_1,r_2)}{\epsilon_p - \epsilon_s - (ck - i\gamma)\mathrm{sgn}(\epsilon_s)}|sq\rangle, \tag{115}$$

$$f_T^C(k,r_1,r_2) = \frac{\sin(kr_{12})}{\pi r_{12}}\left[\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2 - \frac{(\boldsymbol{\alpha}_1\cdot\boldsymbol{\nabla}_1)(\boldsymbol{\alpha}_2\cdot\boldsymbol{\nabla}_2)}{k^2}\right]. \tag{116}$$

Therefore, the total ESE (still denoted as $\bar{Q}$) can be written in a symmetric form

$$\bar{Q}_p^q = \frac{1}{2}\langle p|\Sigma_C^C + \Sigma_T^C(\epsilon_p) + \Sigma_T^C(\epsilon_q)|q\rangle, \tag{117}$$

$$\langle p|\Sigma_C^C|q\rangle = g_{ps}^{sq}\mathrm{sgn}(\epsilon_s). \tag{118}$$

Both $\tilde{Q}$ (84) and $\bar{Q}$ (117) are extremely short ranged and die off beyond a distance of ca. 0.001 a.u. from the position of the nucleus. Therefore, the kernels of $\tilde{Q}$ (84) and $\bar{Q}$ (117) can be fitted into a model operator for each atom [21, 28]:

$$Q = \tilde{Q}_{\mathrm{loc}} + \bar{Q}_{\mathrm{loc}} + \bar{Q}_{n\mathrm{loc}}, \tag{119}$$

$$\tilde{Q}_{\mathrm{loc}} = \sum_\kappa (\tilde{Q}_{\mathrm{loc}})_\kappa P_\kappa, \quad (\tilde{Q}_{\mathrm{loc}})_\kappa(r) = \tilde{A}_\kappa f(r), \tag{120}$$

$$\bar{Q}_{\mathrm{loc}} = \sum_\kappa (\bar{Q}_{\mathrm{loc}})_\kappa P_\kappa, \quad (\bar{Q}_{\mathrm{loc}})_\kappa(r) = \bar{A}_\kappa f(r), \tag{121}$$

$$\bar{Q}_{n\mathrm{loc}} = \sum_{\kappa} (\bar{Q}_{n\mathrm{loc}})_\kappa P_\kappa, \tag{122}$$

$$(\bar{Q}_{n\mathrm{loc}})_\kappa = \sum_{m,n} \begin{pmatrix} |R^L_{m\kappa}\rangle \bar{B}_{mn,\kappa} \langle R^L_{n\kappa}| & 0 \\ 0 & |R^S_{m\kappa}\rangle \bar{B}_{mn,\kappa} \langle R^S_{n\kappa}| \end{pmatrix} P_\kappa, \tag{123}$$

$$P_\kappa = \sum_{m_j} \begin{pmatrix} |\Omega_{\kappa m_j}\rangle\langle\Omega_{\kappa m_j}| & 0 \\ 0 & |\Omega_{-\kappa m_j}\rangle\langle\Omega_{-\kappa m_j}| \end{pmatrix}, \tag{124}$$

where $R^L_{m\kappa}$ ($\Omega_{\kappa m_j}$) and $R^S_{m\kappa}$ ($\Omega_{-\kappa m_j}$) are the radial (angular) parts of an atomic spinor $\phi_{m\kappa m_j}$ of principle quantum number m and angular quantum number $k = (-1)^{j+l+1/2}(j+1/2)$. The parameters $\tilde{A}_\kappa$ for the semi-local EVP $\tilde{Q}_{\mathrm{loc}}$, as well as $\bar{A}_\kappa$ and $\bar{B}_{mn,\kappa}$ for the semi-local and nonlocal parts of the ESE, are to be determined by least-square fittings of the known atomic data. The so-obtained model operator Q (119) can then be included in the SCF calculations. In this way, screening effects on the EVP and ESE are obtained automatically.

If wanted, the transverse part of the one-photon exchange, Fig. 4a, can also be included, so as to replace the integral g^{rs}_{pq} in Eq. (89) with

$$g^{rs}_{pq} = \langle pq|V_C + \frac{1}{2}V_T(\epsilon_r - \epsilon_p) + \frac{1}{2}V_T(\epsilon_s - \epsilon_q)|rs\rangle, \tag{125}$$

$$V_C = \frac{1}{r_{12}}, \tag{126}$$

$$V_T(\omega) = -\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 \frac{e^{\mathrm{i}|q|r_{12}}}{r_{12}} + (\boldsymbol{\alpha}_1 \cdot \boldsymbol{\nabla}_1)(\boldsymbol{\alpha}_2 \cdot \boldsymbol{\nabla}_2)\frac{e^{\mathrm{i}|q|r_{12}} - 1}{q^2 r_{12}}, \quad \omega = qc. \tag{127}$$

As can be seen from the retarded no-pair first-order energy

$$E^{[1]}_{\mathrm{np,r}} = D^i_i + \frac{1}{2}\langle ij|V_C + V_T(0)|ij\rangle - \frac{1}{2}\langle ij|V_C + V_T(\epsilon_i - \epsilon_j)|ji\rangle, \tag{128}$$

retardation [$\omega \neq 0$ in $V_T(\omega)$] occurs only to the exchange (and correlation) but not to the direct interaction. In the limit $\omega \to 0$, $V_T(\omega)$ (127) will reduce to the well-known instantaneous Breit term V_B (which can also be obtained in a semiclassical manner),

$$V_B = V_G + V_g, \tag{129}$$

$$V_G = -\frac{\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{r_{12}}, \tag{130}$$

$$V_g = \frac{\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{2r_{12}} - \frac{(\boldsymbol{\alpha}_1 \cdot \mathbf{r}_{12})(\boldsymbol{\alpha}_2 \cdot \mathbf{r}_{12})}{2r^3_{12}}, \tag{131}$$

where V_G and V_g are the Gaunt and (scalar) gauge terms, respectively.

No-Pair Relativistic Hamiltonians

In the present context, the NPA is just to confine the orbital indices of H_a (88) only to PES, thereby leading to

$$H_+^{\mathrm{QED}} = D_p^q a_q^p + Q_p^q a_q^p + \frac{1}{2} g_{pq}^{rs} a_{rs}^{pq}, \quad p,q,r,s \in \text{ PES} \tag{132}$$

$$= (F_e + Q)_p^q \left\{ a_q^p \right\}_N + \frac{1}{2} g_{pq}^{rs} \left\{ a_{rs}^{pq} \right\}_N + E^{[1]}, \quad p,q,r,s \in \text{ PES} \tag{133}$$

with $E^{[1]}$ and $(F_e)_p^q$ defined in Eqs. (96) and (107), respectively. Here, the subscript N indicates that the normal ordering is taken with respect to $|N;\tilde{0}\rangle$. The H_+^{QED} Hamiltonian (132), along with $\tilde{Q}$ (84), $\bar{Q}$ (117) and g_{pq}^{rs} (125), was already obtained by Shabaev [22] but in a top-down fashion. Further neglecting the Q term leads to the standard no-pair Hamiltonian H_+

$$H_+ = (F_e)_p^q \left\{ a_q^p \right\}_N + \frac{1}{2} g_{pq}^{rs} \left\{ a_{rs}^{pq} \right\}_N + E_{\mathrm{np}}^{[1]}, \quad p,q,r,s \in \text{ PES}, \tag{134}$$

which has been the basis of "no-pair relativistic quantum chemistry." Both H_+^{QED} (133) and H_+ (134) have an intrinsic error of $\mathscr{O}(Z\alpha)^3$. More seriously, the calculated energy is always dependent on the mean-field potential for generating the orbitals. However, as shown numerically by Sapirstein [8], such potential dependence can largely be removed by introducing the following correction

$$E_{\mathrm{PC}}^{(2)} = \frac{(V_{\mathrm{HF}})_{\tilde{i}}^i U_i^{\tilde{i}} + U_{\tilde{i}}^i (V_{\mathrm{HF}})_i^{\tilde{i}} - U_{\tilde{i}}^i U_i^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_i}, \tag{135}$$

where U is the potential in Eq. (38). One then has a "potential-independent no-pair QED" (PI-QED) approach [20]. The corresponding Hamiltonian reads

$$H_+^{\mathrm{PI-QED}} = (F_e + Q - U)_p^q \left\{ a_q^p \right\}_N + \frac{1}{2} g_{pq}^{rs} \left\{ a_{rs}^{pq} \right\}_N + E^{[1]} + E_{\mathrm{PC}}^{(2)}, \quad p,q,r,s \in \text{ PES}. \tag{136}$$

Further neglecting the Q term in $H_+^{\mathrm{PI-QED}}$ leads to

$$H_+^{\mathrm{PI}} = (F_e - U)_p^q \left\{ a_q^p \right\}_N + \frac{1}{2} g_{pq}^{rs} \left\{ a_{rs}^{pq} \right\}_N + E_{\mathrm{NP}}^{[1]} + E_{\mathrm{PC}}^{(2)}, \quad p,q,r,s \in \text{ PES}. \tag{137}$$

The various Hamiltonians discussed so far, together with the quasi-four-component as well as exact and approximate two-component ones [2], can be collected to span a complete and continuous "Hamiltonian Ladder" [21, 23].

Application of the eQED Hamiltonian

Second-Order Energy of an *N*-Electron System

The Hamiltonian H_a (88) (or H_b (70)) can be employed in the Bloch equation for determining the wave operators order by order. The resulting energy expressions are in full agreement with those obtained by the S-matrix formulation of QED [20]. However, the procedure treating all the PES as particles is rather involved. It is more expedite [7] to calculate the physical energy according to Eq. (74) by treating the occupied PES also as holes. Formally, the second term on the right-hand side of Eq. (74) is the analog of nonrelativistic HF reference, whereas the left-hand side is the analog of electron correlation. In reality, the second term corresponds to a polarizable vacuum

$$\Psi(0;\tilde{N}) = \Psi^{(0)}(0;\tilde{N}) + \Psi^{(1)}(0;\tilde{N}) + \cdots, \tag{138}$$

with $\Psi^{(0)}(0;\tilde{N}) = |0,\tilde{N}\rangle$ being the reference for the normal ordering (75). To facilitate the application of MBPT to the first term of Eq. (74), the Hamiltonian (98) normal ordered with respect to $|N;\tilde{N}\rangle$ can be rewritten as

$$\mathscr{H} = H_{0A} + V_{0A} + V_{1A} + V_{2A}, \tag{139}$$

$$H_{0A} = \epsilon_p \left\{a_p^p\right\}_F - \frac{1}{2}\epsilon_p \mathrm{sgn}(\epsilon_p - \epsilon_F), \tag{140}$$

$$= \epsilon_p \left\{a_p^p\right\}_F + \sum_i^N \epsilon_i - \frac{1}{2}\epsilon_p \mathrm{sgn}(\epsilon_p), \tag{141}$$

$$V_{0A} = \left(\frac{1}{2}U - \frac{1}{4}Y\right)_p^p \mathrm{sgn}(\epsilon_p - \epsilon_F), \tag{142}$$

$$= \left(Q - U + \frac{1}{2}V_{\mathrm{HF}}\right)_i^i + \left(\frac{1}{2}U - \frac{1}{4}Q\right)_p^p \mathrm{sgn}(\epsilon_p), \tag{143}$$

$$V_{1A} = (V_{1A})_p^q \left\{a_q^p\right\}_F, \quad V_{1A} = Y - U, \tag{144}$$

$$V_{2A} = \frac{1}{2} g_{pq}^{rs} \left\{a_{rs}^{pq}\right\}_F \tag{145}$$

while the Hamiltonian in the second term of Eq. (74) can be normal ordered with respect to $|0;\tilde{N}\rangle$, i.e.,

$$\mathscr{H} = H_{0B} + V_{0B} + V_{1B} + V_{2B}, \tag{146}$$

$$H_{0B} = \epsilon_p \left\{a_p^p\right\}_n - \frac{1}{2}\epsilon_p \mathrm{sgn}(\epsilon_p), \tag{147}$$

$$V_{0B} = \left(\frac{1}{2}U - \frac{1}{4}Q\right)_p^p \mathrm{sgn}(\epsilon_p), \tag{148}$$

$$V_{1B} = (V_{1B})_p^q \left\{a_q^p\right\}_n, \quad V_{1B} = Q - U, \tag{149}$$

$$V_{2B} = \frac{1}{2} g_{pq}^{rs} \left\{a_{rs}^{pq}\right\}_n. \tag{150}$$

Here, both the PES and NES are solutions of the mean-field equation (38), where the potential U can be chosen in various ways. For instance, $U = V_{\mathrm{HF}}$ corresponds to the standard Dirac HF (DHF) (107) with the empty Dirac picture, whereas $U = V_{\mathrm{HF}} + Q$ to the DHF with the filled Dirac picture [21]. Following the standard MBPT, we obtain immediately

$$E^{(0)} = -\frac{1}{2}\epsilon_p \mathrm{sgn}(\epsilon_p - \epsilon_F) + \frac{1}{2}\epsilon_p \mathrm{sgn}(\epsilon_p) = \sum_i^N \epsilon_i, \tag{151}$$

$$E^{(1)} = V_{0A} - V_{0B} = \left(\frac{1}{2}V_{\mathrm{HF}} - U + Q\right)_i^i, \tag{152}$$

$$E^{(2)} = E_1^{(2)} + E_2^{(2)}, \tag{153}$$

$$\begin{aligned} E_1^{(2)} &= \left[\frac{(V_{1A})_i^a (V_{1A})_a^i}{\epsilon_i - \epsilon_a} + \frac{(V_{1A})_{\tilde{i}}^a (V_{1A})_a^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_a}\right] \\ &\quad - \left[\frac{(V_{1B})_{\tilde{i}}^i (V_{1B})_i^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_i} + \frac{(V_{1B})_{\tilde{i}}^a (V_{1B})_a^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_a}\right] \end{aligned} \tag{154}$$

$$= E_{\mathrm{FS},1}^{(2)} + E_{\mathrm{Q},1}^{(2)}, \tag{155}$$

$$\begin{aligned} E_{\mathrm{FS},1}^{(2)} &= \frac{(V_{\mathrm{HF}} - U)_i^a (V_{\mathrm{HF}} - U)_a^i}{\epsilon_i - \epsilon_a} + \frac{(V_{\mathrm{HF}} - U)_{\tilde{i}}^a (V_{\mathrm{HF}} - U)_a^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_a} \\ &\quad - \frac{U_{\tilde{i}}^a U_a^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_a} - \frac{U_{\tilde{i}}^i U_i^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_i}, \end{aligned} \tag{156}$$

$$\begin{aligned} E_{\mathrm{Q},1}^{(2)} &= \frac{(V_{\mathrm{HF}} - U)_i^a Q_a^i + Q_i^a (V_{\mathrm{HF}} - U)_a^i + Q_i^a Q_a^i}{\epsilon_i - \epsilon_a} \\ &\quad + \frac{(V_{\mathrm{HF}})_{\tilde{i}}^a Q_a^{\tilde{i}} + Q_{\tilde{i}}^a (V_{\mathrm{HF}})_a^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_a} - \frac{Q_{\tilde{i}}^i Q_i^{\tilde{i}} - U_{\tilde{i}}^i Q_i^{\tilde{i}} - Q_{\tilde{i}}^i U_i^{\tilde{i}}}{\epsilon_{\tilde{i}} - \epsilon_i}, \end{aligned} \tag{157}$$

$$E_2^{(2)} = \frac{1}{4}\left.\frac{\bar{g}_{mn}^{ab}\bar{g}_{ab}^{mn}}{\epsilon_m+\epsilon_n-\epsilon_a-\epsilon_b}\right|_{m,n=i,j,\tilde{i},\tilde{j}} - \frac{1}{4}\left.\frac{\bar{g}_{\tilde{i}\tilde{j}}^{pq}\bar{g}_{pq}^{\tilde{i}\tilde{j}}}{\epsilon_{\tilde{i}}+\epsilon_{\tilde{j}}-\epsilon_p-\epsilon_q}\right|_{p,q=i,j,a,b} \tag{158}$$

$$= \left[\frac{1}{4}\frac{\bar{g}_{ij}^{ab}\bar{g}_{ab}^{ij}}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b} + \frac{1}{2}\frac{\bar{g}_{i\tilde{j}}^{ab}\bar{g}_{ab}^{i\tilde{j}}}{\epsilon_i+\epsilon_{\tilde{j}}-\epsilon_a-\epsilon_b}\right] - \left[\frac{1}{4}\frac{\bar{g}_{\tilde{i}\tilde{j}}^{ij}\bar{g}_{ij}^{\tilde{i}\tilde{j}}}{\epsilon_{\tilde{i}}+\epsilon_{\tilde{j}}-\epsilon_i-\epsilon_j} + \frac{1}{2}\frac{\bar{g}_{\tilde{i}\tilde{j}}^{ia}\bar{g}_{ia}^{\tilde{i}\tilde{j}}}{\epsilon_{\tilde{i}}+\epsilon_{\tilde{j}}-\epsilon_i-\epsilon_a}\right]. \tag{159}$$

As shown before [20], $E^{(2)}$ (153) is identical with that obtained by the S-matrix approach of QED. However, the latter involves 28 Feynman diagrams (see Fig. 5 in Ref. [20]) and requires careful cancelation of divergent integrals. The first and second terms of $E_1^{(2)}$ (154) and $E_2^{(2)}$ (158) arise from the $\Psi(N;\tilde{N})$ and $\Psi(0;\tilde{N})$ states, respectively. The one-body $E_1^{(2)}$ (154) can further be decomposed into two terms, $E_{\mathrm{FS},1}^{(2)}$ (156) and $E_{\mathrm{Q},1}^{(2)}$ (157). Both $E_{\mathrm{FS},1}^{(2)}$ (156) and $E_2^{(2)}$ (159) arise from the Fock space Hamiltonian [18, 19] H_a^{FS} (89), while $E_{\mathrm{Q},1}^{(2)}$ (157) is due exclusively to the EVP and ESE. The two terms of $E^{(2)}$ (153) can nicely be represented by the same Goldstone-like diagrams shown in Fig. 5. It is just that the particles and holes, as well as the one-body potential, are interpreted differently. Due to the large gap between the NES and PES, the up-to-second-order treatment of the NES should be sufficient. Therefore, the major challenge still resides in the no-pair correlation within the manifold of PES. Yet, this can be facilitated by using relativistic explicitly correlated methods, which are completely parallel [29] to the nonrelativistic counterparts under the concept of "extended no-pair projection" [7].

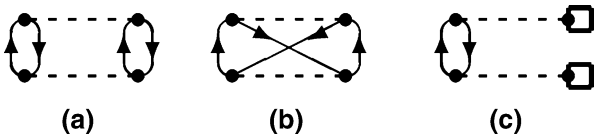


Fig. 5 Diagrammatical representation of the second-order energy. (**a**) Two-body direct; (**b**) two-body exchange; (**c**) one body. For the $\Psi(N;\tilde{N})$ state, the particles (up-going lines) and holes (down-going lines) are $\{a,b\}$ and $\{i,j,\tilde{i},\tilde{j}\}$, respectively. The one-body potential represented by the square is V_{1A}. For the $\Psi(0;\tilde{N})$ state, the particles and holes are $\{a,b,i,j\}$ and $\{\tilde{i},\tilde{j}\}$, respectively. The one-body potential is V_{1B}. A global negative sign should be inserted to the terms of $\Psi(0;\tilde{N})$ (Reprinted with permission from Ref. [20]. ©2013, American Institute of Physics)

Finally, a brief comparison with the configuration space approach is in order. It gives rise to the following second-order energy [7]

$$E_{\rm CS}^{(2)} = E_{\rm CS,1}^{(2)} + E_{\rm CS,2}^{(2)}, \tag{160}$$

$$E_{\rm CS,1}^{(2)} = \frac{(V_{\rm HF}-U)_i^a (V_{\rm HF}-U)_a^i}{\epsilon_i - \epsilon_a} + \frac{(V_{\rm HF}-U)_{\tilde{i}}^{\tilde{i}} (V_{\rm HF}-U)_{\tilde{i}}^{i}}{\epsilon_i - \epsilon_{\tilde{i}}}, \tag{161}$$

$$E_{\rm CS,2}^{(2)} = \frac{1}{4}\frac{\bar{g}_{ij}^{ab}\bar{g}_{ab}^{ij}}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_b} + \frac{1}{4}\frac{\bar{g}_{ij}^{\tilde{i}\tilde{j}}\bar{g}_{\tilde{i}\tilde{j}}^{ij}}{\epsilon_i+\epsilon_j-\epsilon_{\tilde{i}}-\epsilon_{\tilde{j}}} + \frac{1}{2}\frac{\bar{g}_{ij}^{a\tilde{j}}\bar{g}_{a\tilde{j}}^{ij}}{\epsilon_i+\epsilon_j-\epsilon_a-\epsilon_{\tilde{j}}}. \tag{162}$$

It is seen that $E_{\rm CS,1}^{(2)}$ and $E_{\rm CS,2}^{(2)}$ agree, respectively, with $E_{\rm FS,1}^{(2)}$ (156) and $E_2^{(2)}$ (159) only in the first terms involving solely the PES, but are very different from the latter in the terms involving the NES. Therefore, the configuration space approach is plainly wrong. This can also be seen from the "recovery" [7] of the configuration space approach from QED by replacing the Feynman propagator $S_F(x_2,x_1)$ (54) with the "configuration space propagator"

$$S_C(x_2,x_1) = -i\langle 0|T[\phi(x_2)\phi^\dagger(x_1)]|0\rangle, \tag{163}$$

$$= -i\Theta(t_2-t_1)\langle 0|\phi(x_2)\phi^\dagger(x_1)+\phi^\dagger(x_1)\phi(x_2)|0\rangle \tag{164}$$

$$= \int_{-\infty}^{+\infty}\frac{d\omega}{2\pi}\frac{\varphi_r(\mathbf{r}_2)\varphi_r^\dagger(\mathbf{r}_1)}{\omega-\epsilon_r+i\eta|\epsilon_r|}e^{-i\omega(t_2-t_1)}, \quad r\in \text{PES, NES}, \tag{165}$$

where the field $\phi(x)$ is defined as

$$\phi(x) = a_p\varphi_p(\mathbf{r})e^{-i\epsilon_p t} + a_{\tilde{p}}^\dagger\varphi_{\tilde{p}}(\mathbf{r})e^{-i\epsilon_{\tilde{p}}t} \tag{166}$$

in accordance with the empty Dirac picture. The distinction between $S_F(x_2,x_1)$ (54) and $S_C(x_2,x_1)$ (165) lies in that the former propagates the PES forward in time (through the Heaviside function $\Theta(t_2-t_1)$ or equivalently the lower-half-plane poles $\epsilon_p - i\eta$) and the NES backward in time (through the Heaviside function $\Theta(t_1-t_2)$ or equivalently the upper-half-plane poles $\epsilon_{\tilde{p}}+i\eta$), whereas the latter propagates both the PES and NES forward in time. Since the replacement of $S_F(x_2,x_1)$ (54) with $S_C(x_2,x_1)$ (165) changes the nature of NES completely, it is clear that the configuration space approach is even *not* an approximation of QED, the highest level of theory for electromagnetic interactions between charged particles.

Mean-Field Theory of Positrons

The term "electron-positron pair" discussed so far refers to virtual positrons. However, the present relativistic many-body Hamiltonian in the form of H_b (70) or H_a (88) can also be used to describe real positrons. For instance, up to first order, the energy of a system of N electrons and $\tilde{M}$ real positrons can be calculated as

$$E_{\mathrm{ep}}^{[1]} = \langle N;\tilde{0}|B^\dagger H_b B|N;\tilde{0}\rangle, \quad B = \Pi_{\tilde{i}}^{\tilde{M}} b^{\tilde{i}} \tag{167}$$

$$= \langle N;\tilde{N}|A^\dagger H_a A|N;\tilde{N}\rangle, \quad A = \Pi_{\tilde{i}}^{\tilde{M}} a_{\tilde{i}} \tag{168}$$

$$= \left[\sum_{i=1}^{N}(D+Q)_i^i + \frac{1}{2}\sum_{i,j=1}^{N}\bar{g}_{ij}^{ij}\right] + \left[-\sum_{\tilde{i}=\tilde{1}}^{\tilde{M}}(D+Q)_{\tilde{i}}^{\tilde{i}} + \frac{1}{2}\sum_{\tilde{i},\tilde{j}=\tilde{1}}^{\tilde{M}}\bar{g}_{\tilde{i}\tilde{j}}^{\tilde{i}\tilde{j}}\right]$$
$$-\sum_{i=1}^{N}\sum_{\tilde{j}=\tilde{1}}^{\tilde{M}}\bar{g}_{i\tilde{j}}^{i\tilde{j}}, \tag{169}$$

where the first and second terms are the energies of the N electrons and of the $\tilde{M}$ positrons, respectively, whereas the third, cross term represents their mutual interactions. The expression (169) can further be written in a more symmetric form

$$E_{\mathrm{ep}}^{[1]} = \sum_k n_k (D+Q)_k^k + \frac{1}{2}\sum_{k,l} n_k n_l \bar{g}_{kl}^{kl}, \quad k,l \in \mathrm{PES,NES} \tag{170}$$

by assigning an occupation number n_k to each orbital φ_k: n_k is zero for the unoccupied PES and NES, $+1$ for the occupied PES, and -1 for the occupied NES. This occupation pattern has a nice interpretation [30]: A negative energy electron has both a negative charge and a negative mass, such that its depletion leaves behind a hole (positron) of positive charge and positive mass, with the occupation number being -1. At first glance, the notation $|N;\tilde{N}\rangle$ in Eq. (168) implies that all the NES are occupied such that there are no unoccupied NES. However, the unoccupied NES do exist due to the underlying normal ordering of H_a (99) over $|0;\tilde{N}\rangle$. Moreover, one must work with a basis in which the functions for the PES and NES are strongly orthogonal to each other. Otherwise, the distinction between the PES and NES cannot be made in the energy expression. Therefore, the Lagrangian to be minimized should be written as

$$L = E_{\mathrm{ep}}^{[1]} - \sum_{i,j=1}^{N}(\langle i|j\rangle - \delta_{ij})\epsilon_{ij} + \sum_{\tilde{i},\tilde{j}=1}^{\tilde{M}}(\langle\tilde{i}|\tilde{j}\rangle - \delta_{\tilde{i}\tilde{j}})\epsilon_{\tilde{i}\tilde{j}}, \tag{171}$$

which has no multipliers between the PES and NES. The conditions $\frac{\delta L}{\delta \varphi_i^\dagger} = 0$ and $\frac{\delta L}{\delta \varphi_{\tilde{i}}^\dagger} = 0$ then give rise to the stationarity conditions for the variation of $E_{\rm ep}^{[1]}$ (170)

$$f\varphi_i = \sum_j^N \varphi_j \epsilon_{ji}, \quad f\varphi_{\tilde{i}} = \sum_{\tilde{j}}^{\tilde{M}} \varphi_{\tilde{j}} \epsilon_{\tilde{j}\tilde{i}}, \tag{172}$$

where

$$f = D + Q + \sum_k n_k \bar{g}_{k\cdot}^{k\cdot}, \quad k \in \text{PES, NES}, \tag{173}$$

$$f_p^q = (D+Q)_p^q + \sum_k n_k \bar{g}_{kp}^{kq}, \quad k \in \text{PES, NES}. \tag{174}$$

Since the Fock operator f (173)/(174) is manifestly Hermitian and invariant with respect to separate unitary transformations of the occupied PES and NES, one can go to a canonical representation, viz.,

$$f\varphi_k = \varphi_k \epsilon_k, \quad k \in \text{PES, NES}. \tag{175}$$

The so-called dual kinetic balance [31] is most recommended for discretizing Eq. (175), for it provides variational safety in conjunction with Gaussian type of functions [32]. Moreover, the iterations should proceed by occupying the energetically lowest PES and highest NES, which corresponds to an implicit but updated projection for avoiding variational collapse.

A few remarks can further be made here: (a) The energy (170) does vary with respect to rotations between the occupied PES and NES, since they belong to distinct subspaces. (b) It is easy to show that the cross exchange term $-\sum_{i\tilde{j}} g_{i\tilde{j}}^{\tilde{j}i}$ vanishes in the nonrelativistic limit, meaning that electron and positron are distinguishable particles in the nonrelativistic world, such that their mutual anti-symmetrization is no longer required. In other words, only the QED formulation provides an equal-footing treatment of electrons and positrons. (c) If the EVP-ESE term Q is neglected, the present mean-field theory of electrons and positrons will reduce to that formulated by Dyall [30] in a different way. A second-order MBPT can also be formulated following the procedure in section "Second-Order Energy of an N-Electron System," but this goes beyond the scope of the work.

Conclusion

Charge conjugation is the key for distinguishing relativistic from nonrelativistic quantum mechanics. Its incorporation via the charge-conjugated contractions of fermion operators leads directly to a relativistic many-body Hamiltonian (70)/(88)

that is not only fully in line with the principles of QED but also fully compatible with any wave-function-/density-functional-based correlation method. The Hamiltonian along with accurate correlation methods can be used for ultrahigh precision calculations of molecular spectroscopies and potential energy surfaces. It can therefore be envisaged that "molecular QED" will soon become an active and exciting field of research.

Acknowledgements This work was supported by the NSFC (Project Nos. 21033001, 21273011, and 21290192).

References

1. The acronym 'X2C' (pronounced as ecstacy) for exact two-component Hamiltonians was proposed by W. Liu after intensive discussions with H. J. Aa. Jensen, W. Kutzelnigg, T. Saue and L. Visscher during the Twelfth International Conference on the Applications of Density Functional Theory (DFT-2007), Amsterdam, August 26–30, 2007. Note that the 'exact' here emphasizes that all the solutions of the matrix Dirac equation can be reproduced up to machine accuracy. It is particularly meaningful when compared with the approximate two-component (A2C) Hamiltonians.
2. Liu W (2010) Ideas of relativistic quantum chemistry. Mol Phys 108:1679
3. Saue T (2011) Relativistic Hamiltonians for chemistry: a primer. Chem Phys Chem 12:3077
4. Peng D, Reiher M (2012) Exact decoupling of the Fock operator. Theor Chem Acc 131:1081
5. Liu W (2012) The big picture of relativistic molecular quantum mechanics. AIP Conf Proc 1456:62
6. Liu W (2015) Effective quantum electrodynamics Hamiltonians: a tutorial review. Int J Quantum Chem 115:631
7. Liu W (2012) Perspectives of relativistic quantum chemistry: the negative energy cat smiles. Phys Chem Chem Phys 14:35
8. Sapirstein J, Cheng KT, Chen MH (1999) Potential independence of the solution to the relativistic many-body problem and the role of negative-energy states in helium-like ions. Phys Rev A 59:259
9. Kutzelnigg W (2003) Diamagnetism in relativistic theory. Phys Rev A 67:032109
10. Kutzelnigg W (2008) Relativistic corrections to the partial wave expansion of two-electron atoms. Int J Quantum Chem 108:2280
11. Thierfelder C, Schwerdtfeger P (2010) Quantum electrodynamic corrections for the valence shell in heavy many-electron atoms. Phys Rev A 82:062503
12. Júregui R, Bunge C, Ley-Koo E (1997) Upper bounds to the eigenvalues of the no-pair Hamiltonian. Phys Rev A 55:1781
13. Nakatsuji H, Nakashima H (2005) Analytically solving the relativistic Dirac-Coulomb equation for atoms and molecules. Phys Rev Lett 95:050407
14. Watanabe Y, Nakano H, Tatewaki H (2007) Effect of removing the no-virtual-pair approximation on the correlation energy of the He isoelectronic sequence. J Chem Phys 126:174105
15. Pestka G, Bylicki M, Karwowski J (2006) Application of the complex-coordinate rotation to the relativistic Hylleraas-CI method: a case study. J Phys B At Mol Opt Phys 39:2979
16. Brown GE, Ravenhall DG (1951) On the Interaction of two electrons. Proc R Soc Lond A 208:552
17. Sucher J (1984) Foundations of the relativistic theory of many-electron bound states. Int J Quantum Chem 25:3
18. Dyall KG, Fægri K Jr (2007) Introduction to relativistic quantum chemistry. Oxford University Press, New York

19. Kutzelnigg W (2012) Solved and unsolved problems in relativistic quantum chemistry. Chem Phys 395:16
20. Liu W, Lindgren I (2013) Going beyond "no-pair relativistic quantum chemistry". J Chem Phys 139:014108
21. Liu W (2104) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59
22. Shabaev VM (1993) Schrödinger-like equation for the relativistic few-electron atom. J Phys B At Mol Opt Phys 26:4703
23. Liu W (2014) Perspective: relativistic Hamiltonians. Int J Quantum Chem 114:983
24. Greiner W, Reinhart J (1996) Field quantization. Springer, Berlin
25. Schwinger J (1951) On Gauge invariance and vacuum polarization. Phys Rev 82:664
26. Lindgren I, Morrison J (1986) Atomic many-body theory, 2nd edn. Springer, Berlin
27. Lindgren I (2011) Relativistic many-body theory: a new field-theoretical approach. Springer, New York
28. Shabaev VM, Tupitsyn II, Yerokhin VA (2013) Model operator approach to the Lamb shift calculations in relativistic many-electron atoms. Phys Rev A 88:012513
29. Li Z, Shao S, Liu W (2012) Relativistic explicit correlation: Coalescence conditions and practical suggestions. J Chem Phys 136:144117
30. Dyall KG (2012) A question of balance: kinetic balance for electrons and positrons. Chem Phys 395:35
31. Shabaev VM, Tupitsyn II, Yerokhin VA, Plunien G, Soff G (2004) Dual kinetic balance approach to basis-set expansions for the Dirac equation. Phys Rev Lett 93:130405
32. Sun Q, Liu W, Kutzelnigg W (2011) Comparison of restricted, unrestricted, inverse, and dual kinetic balances for four-component relativistic calculations. Theor Chem Acc 129:423

No-Pair Relativistic Hamiltonians: Q4C and X2C

11

Wenjian Liu

Contents

Introduction 375
Q4C 376
X2C 383
Summary 392
References 392

Abstract

Under the no-pair approximation, four- and two-component relativistic quantum chemical calculations can be made identical in all the aspects of simplicity, accuracy, and efficiency through the quasi-four-component (Q4C) and exact two-component (X2C) approaches.

Keywords

No-pair Hamiltonian • Quasi-four-component • Exact two-component

Introduction

Given some fundamental defects [1, 2], the no-pair approximation, which neglects electron-positron pair creation and annihilation effects, is usually sufficiently accurate for most chemical and physical properties. Under this approximation, the machinery of relativistic quantum chemistry consists of two sets of methods, four-component and two-component, and both sets have been developed with con-

W. Liu (✉)
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: liuwjbdf@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_2

siderable success in the last two and a half decades [3–7]. Yet, as of now, statements like "four-component good, two-component bad" or "two-component good, four-component bad" can still be seen from the literature or heard from conference lectures. In essence, the former emphasizes the accuracy and simplicity of the Dirac operator and other four-component property operators against the approximate and overly complicated structure of the two-component counterparts, whereas the latter emphasizes the computational efficiency of two-component approaches against the additional overhead of four-component approaches due to the existence of negative energy states (NES). However, such judgments were based on personal preferences instead of impersonal criteria in terms of all the aspects of simplicity, accuracy, and efficiency. While those in favor of four-component approaches are extremely reluctant in introducing any approximations (other than truncations in the 1- and n-particle basis sets), those in favor of two-component approaches are willing to accept even drastic approximations in the development of two-component Hamiltonians and property operators. That is, the accuracy and simplicity of four-component approaches have to pay for the computational cost, whereas the efficiency and complexity of two-component approaches come along with the loss of accuracy. In reality, none of these is true. If one is willing to accept certain approximations that are orders of magnitude smaller than, e.g., basis set incompleteness errors as well as uncertainties of experimental measurements (NB: such approximations are hardly "approximate"!), four- and two-component approaches can be made identical in all the aspects of simplicity, accuracy, and efficiency, at both mean-field and correlated levels, for both electronic structure and static/response properties. This has already been achieved via the quasi-four-component (Q4C) [8] and exact two-component (X2C) [5–7] approaches (NB: The acronym 'X2C' (pronounced as 'ecstacy') for exact two-component Hamiltonians was proposed by W. Liu after intensive discussions with H. J. Aa. Jensen, W. Kutzelnigg, T. Saue and L. Visscher during the Twelfth International Conference on the Applications of Density Functional Theory (DFT-2007), Amsterdam, August 26–30, 2007. Note that the 'exact' here emphasizes that all the solutions of the matrix Dirac equation can be reproduced up to machine accuracy. It is particularly meaningful when compared with the approximate two-component (A2C) Hamiltonians). As such, one should really speak of "four- and two-component equally good!" [9]. The essential ideas underlying these approaches are to be summarized here. Plain and boldface letters are used to denote operators and matrices, respectively. The Einstein summation convention over repeated indices is always employed.

Q4C

The Dirac equation for an electron moving in a local effective potential V reads

$$\begin{pmatrix} V & c\boldsymbol{\sigma}\cdot\mathbf{p} \\ c\boldsymbol{\sigma}\cdot\mathbf{p} & V-2mc^2 \end{pmatrix} \begin{pmatrix} \psi_i^L \\ \psi_i^S \end{pmatrix} = \begin{pmatrix} \psi_i^L \\ \psi_i^S \end{pmatrix} \epsilon_i, \tag{1}$$

where the rest-mass energy mc^2 has been subtracted. For a positive energy state (PES) ψ_i of energy ϵ_i, we get from the second row of Eq. (1) the relation between the lower and upper components

$$c\psi_i^S = \frac{1}{2m} R_i \boldsymbol{\sigma} \cdot \mathbf{p} \psi_i^L \overset{c\to\infty}{\to} \frac{1}{2m} \boldsymbol{\sigma} \cdot \mathbf{p} \psi_i^L, \tag{2}$$

$$R_i(\mathbf{r}) = \left[1 + \frac{\epsilon_i - V(\mathbf{r})}{2mc^2}\right]^{-1} \overset{c\to\infty}{\to} 1, \tag{3}$$

while for a NES $\psi_{\bar{i}}$ of energy $\epsilon_{\bar{i}} = \epsilon_i + 2mc^2$, we will get from the first row of Eq. (1)

$$c\psi_{\bar{i}}^L = \frac{1}{2m} R_{\bar{i}} \boldsymbol{\sigma} \cdot \mathbf{p} \psi_{\bar{i}}^S \overset{c\to\infty}{\to} -\frac{1}{2m} \boldsymbol{\sigma} \cdot \mathbf{p} \psi_{\bar{i}}^S, \tag{4}$$

$$R_{\bar{i}}(\mathbf{r}) = -\left[1 - \frac{\epsilon_{\bar{i}} - V(\mathbf{r})}{2mc^2}\right]^{-1} \overset{c\to\infty}{\to} -1. \tag{5}$$

It follows that, for the PES ψ_i, the lower component ψ_i^S is smaller (in the mean sense) than the upper component ψ_i^L by a factor of c^{-1}. They are hence called small and large components, respectively. In particular, the former vanishes in the nonrelativistic limit. Obviously, the opposite is true for the NES $\psi_{\bar{i}}$. It is also obvious that the lower and upper components of both PES and NES are of opposite parity for the $\boldsymbol{\sigma} \cdot \mathbf{p}$ operator changes the parity, whereas the $R_i(\mathbf{r})$ $[R_{\bar{i}}(\mathbf{r})]$ operator is totally symmetric. A direct deduction is that the two components must be expanded in separate but properly related basis sets. Among the various prescriptions [10] for constructing suitable basis sets, the so-called restricted kinetic balance (RKB) [11] is the simplest. It provides the least adequate representation of the small components. Specifically, a RKB basis consists of a 2-spinor basis $\{g_\mu, \mu = 1, \cdots, 2N^L\}$ for the large components, from which the small component 2-spinor basis is generated in a one-to-one fashion according to the following prescription:

$$f_\mu = \frac{1}{2mc} \boldsymbol{\sigma} \cdot \mathbf{p} g_\mu, \quad \mu = 1, \cdots, 2N^L. \tag{6}$$

The molecular 4-spinors (M4S) ψ_i are then expanded as

$$\psi_i = \begin{pmatrix} g_\mu \\ 0 \end{pmatrix} A_{\mu i} + \begin{pmatrix} 0 \\ f_\mu \end{pmatrix} B_{\mu i}, \quad i = 1, \cdots, 4N^L. \tag{7}$$

The corresponding matrix Dirac equation reads

$$\begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & \frac{1}{4m^2c^2}\mathbf{W} - \mathbf{T} \end{pmatrix} \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{S} & 0 \\ 0 & \frac{1}{2mc^2}\mathbf{T} \end{pmatrix} \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \boldsymbol{\epsilon}, \tag{8}$$

where the individual matrices are all of dimension $2N^L \times 2N^L$, with the elements being

$$\mathbf{V}_{\mu\nu} = \langle g_\mu | V | g_\nu \rangle, \quad \mathbf{T}_{\mu\nu} = \frac{1}{2m} \langle \boldsymbol{\sigma} \cdot \mathbf{p} g_\mu | \boldsymbol{\sigma} \cdot \mathbf{p} g_\nu \rangle = \langle g_\mu | T | g_\nu \rangle,$$

$$\mathbf{W}_{\mu\nu} = \langle \boldsymbol{\sigma} \cdot \mathbf{p} g_\mu | V | \boldsymbol{\sigma} \cdot \mathbf{p} g_\nu \rangle = \langle g_\mu | \boldsymbol{\sigma} \cdot \mathbf{p} V \boldsymbol{\sigma} \cdot \mathbf{p} | g_\nu \rangle, \quad \mathbf{S}_{\mu\nu} = \langle g_\mu | g_\nu \rangle. \tag{9}$$

Equation (8) is therefore of dimension $4N^L \times 4N^L$ with $2N^L$ PES and $2N^L$ NES, which are separated by ca. $2mc^2 \approx 1\,\text{MeV}$. When solving Eq. (8) iteratively, the energetically lowest PES are chosen to be occupied in each iteration cycle, so as to avoid variational collapse. While the rotations between the occupied and unoccupied PES lower the total energy, those between the occupied PES and unoccupied NES raise the total energy, to a much lesser extent though.

It is clear that it is the appearance of the NES that makes four-component calculations very expensive. There exist two paradigms to reduce the computational cost. One is to retain the aesthetically simple four-component structure but *freeze* these degrees of freedom, while the other is to *remove* these degrees of freedom so as to obtain a two-component relativistic theory describing only electrons. Both paradigms try to decouple the solutions of positive and negative energy. While the former employs the untransformed Hamiltonian and introduces approximations from the outset, the latter invokes an effective Hamiltonian and has to introduce suitable approximations at a later stage, without which any such electron-only theory, whether finite or infinite order, cannot really compete with the original four-component approach. Anyhow, the two paradigms stem from the same physics and should hence be fully equivalent.

To realize the first paradigm, we first take a close look at the $R_i(\mathbf{r})$ operator (3), which is plotted in Fig. 1 as a function of the distance from the position of the Rn atom. It is seen that the effect of $R_i(\mathbf{r})$ is extremely short ranged. That is, each $R_i(\mathbf{r})$ becomes just a constant factor beyond a small radius r_c (ca. 0.05 a.u., roughly the radii of $2s$ and $2p$). Imagine we have first solved the (radial) Dirac equation for each isolated (spherical and unpolarized) atom and thus obtained the corresponding atomic 4-spinors (A4S) $\{\varphi_p\}$. Then, the atoms are brought together to synthesize the molecule. While both the large and small components of φ_p will change, the S/L ratio will *not*! [9, 12]. The mathematical realization of such a physical picture is to expand the M4S ψ_i as a linear combination of such A4S (LCA4S) [13]:

$$\psi_i = \varphi_p C_{pi} = \begin{pmatrix} \varphi_p^L \\ \varphi_p^S \end{pmatrix} C_{pi} = \begin{pmatrix} g_\mu a_{\mu p} \\ f_\mu b_{\mu p} \end{pmatrix} C_{pi}. \tag{10}$$

Note that $\{\varphi_p\}$ can either be non-expanded numerical (grid) A4S or generally contacted atomic natural spinors (ANSs) expanded in a RKB basis, as indicated by the last equality. Such an expansion amounts to projecting the molecular mean-field

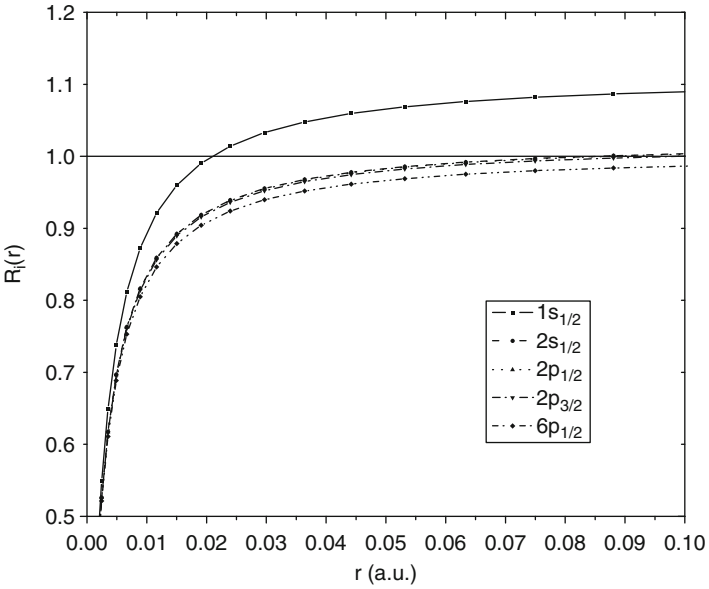


Fig. 1 The $R_i(r)$ operator (3) with $V = V_N + V_H + V_{LDA}$ as a function of the distance from the position of Rn. The radial expectation values of $1s_{1/2}$, $2s_{1/2}$, $2p_{1/2}$, $2p_{3/2}$, and $3s_{1/2}$ are 0.015, 0.063, 0.051, 0.060, and 0.163 a.u., respectively (Reprinted with permission from J. Chem. Phys. 127, 104106 (2007). © 2007 American Institute of Physics)

Hamiltonian F onto the PES of the isolated constituent atoms,

$$F_+^{\mathrm{Q4C}} = P_+ F P_+ \approx \left(\sum_A^{\oplus} P_+^A\right) F \left(\sum_B^{\oplus} P_+^B\right), \tag{11}$$

by which molecular NES are all excluded. Physically, this amounts to neglecting rotations between the PES and NES of the isolated atoms, a kind of polarization on the atomic vacua induced by the molecular field. As molecular formation is a very low-energy process, its perturbation on the vacuum is of $O(c^{-4})$, smaller than, e.g., the basis set incompleteness errors by orders of magnitude [8,9]. At this stage, the computational efficiency is gained only in the matrix diagonalization step (by a factor of 8), which is very little for a moderate basis. The real gain in efficiency can only be achieved by further invoking a "model small component approximation" (MSCA) [8]. The result is of four-component structure but looks very much like an effective two-component approach and is therefore termed "quasi-4-component" (Q4C) [8].

To be general enough, we write the mean-field operator F as [3]

$$F = \begin{pmatrix} F^{LL} & F^{LS} \\ F^{SL} & F^{SS} \end{pmatrix}, \tag{12}$$

where

$$F^{LL} = V^{LL} + c_2(J^{LL} + J^{SS}) - c_3 K^{LL} + c_4 \tilde{K}^{SS}, \tag{13}$$

$$F^{LS} = c\boldsymbol{\sigma} \cdot \boldsymbol{\pi} - c_3 K^{LS} - c_4(\tilde{J}^{LS} - \tilde{K}^{LS}), \tag{14}$$

$$F^{SL} = c\boldsymbol{\sigma} \cdot \boldsymbol{\pi} - c_3 K^{SL} - c_4(\tilde{J}^{SL} - \tilde{K}^{SL}), \tag{15}$$

$$F^{SS} = V^{SS} + c_2(J^{LL} + J^{SS}) - c_3 K^{SS} + c_4 \tilde{K}^{LL}, \tag{16}$$

$$V^{LL} = V_N + c_0 Q^{LL} + c_1 V_{xc}, \tag{17}$$

$$V^{SS} = V_N + c_0 Q^{SS} + c_1 V_{xc} - 2mc^2, \tag{18}$$

$$J^{XX}_{\mu\nu} = \sum_i (\mu^Y \nu^Y | i^X i^X), \quad X, Y = L, S, \tag{19}$$

$$K^{XY}_{\mu\nu} = \sum_i (\mu^X i^X | i^Y \nu^Y), \quad X, Y = L, S, \tag{20}$$

$$\tilde{J}^{XY}_{\mu\nu} = \sum_i (\mu^X \boldsymbol{\sigma} \nu^Y | i \boldsymbol{\alpha} i), \quad X \neq Y, \tag{21}$$

$$\tilde{K}^{XX}_{\mu\nu} = \sum_i (\mu^Y \boldsymbol{\sigma} i^X | i^X \boldsymbol{\sigma} \nu^Y), \quad X \neq Y, \tag{22}$$

$$\tilde{K}^{XY}_{\mu\nu} = \sum_i (\mu^X \boldsymbol{\sigma} i^Y | i^X \boldsymbol{\sigma} \nu^Y), \quad X \neq Y. \tag{23}$$

A specific mean-field operator F (12) is obtained by setting the constants c_i to the corresponding values. For instance, the standard hybrid Dirac-Kohn-Sham (DKS) scheme is obtained by setting $c_2 = 1$, $c_0 = c_4 = 0$, and c_1 and c_3 to some empirical parameters. The model electron self-energy (ESE) and vacuum polarization (EVP) operator Q (see Refs. [3, 14]) can be treated just like the nuclear attraction V_N. It is also interesting to see that the direct contributions (J^{XX}) from the Coulomb interaction appear in the diagonal blocks, whereas the counterparts ($\tilde{J}^{XY}$) of the Gaunt interaction only in the off-diagonal blocks. In contrast, the exchange contributions from both the Coulomb (K^{XY}) and Gaunt ($\tilde{K}^{XY}$) interactions are distributed over all the four blocks. If wanted, the instantaneous Breit term can further be added. The matrix representation of F_+^{Q4C} (11) reads [5]

$$\mathbf{F}^{\mathrm{Q4C}}\mathbf{C} = \tilde{\mathbf{S}}\mathbf{C}\boldsymbol{\epsilon}, \tag{24}$$

$$(\mathbf{F}^{\mathrm{Q4C}})_{pq} = (\mathbf{h}_D)_{pq} + \mathbf{G}_{pq}, \tag{25}$$

where

$$(\mathbf{h}_D)_{pq} = \langle \varphi_p^L | V^{LL} | \varphi_q^L \rangle + \langle \varphi_p^S | c\boldsymbol{\sigma} \cdot \mathbf{p} | \varphi_q^L \rangle + \langle \varphi_p^L | c\boldsymbol{\sigma} \cdot \mathbf{p} | \varphi_q^S \rangle + \langle \varphi_p^S | V^{SS} | \varphi_q^S \rangle, \tag{26}$$

$$\begin{aligned}
\mathbf{G}_{pq} &= [c_2(\varphi_p\varphi_q|\varphi_r\varphi_s) - c_3(\varphi_p\varphi_s|\varphi_r\varphi_q) \\
&\quad - c_4(\varphi_p\boldsymbol{\alpha}\varphi_q|\varphi_r\boldsymbol{\alpha}\varphi_s) + c_4(\varphi_p\boldsymbol{\alpha}\varphi_s|\varphi_r\boldsymbol{\alpha}\varphi_q)]\mathbf{P}_{sr}, && (27)\\
&= \mathbf{G}^{LL}_{pq} + \mathbf{G}^{SS}_{pq}, && (28)\\
\mathbf{G}^{LL}_{pq} &= [c_2(\varphi^L_p\varphi^L_q|\varphi^L_r\varphi^L_s) - c_3(\varphi^L_p\varphi^L_s|\varphi^L_r\psi^L_q)]\mathbf{P}_{sr}, && (29)\\
\mathbf{G}^{SS}_{pq} &= [c_2(\varphi^L_p\varphi^L_q|\varphi^S_r\varphi^S_s) + c_2(\varphi^S_p\varphi^S_q|\varphi^L_r\varphi^L_s) + c_2(\varphi^S_p\varphi^S_q|\varphi^S_r\psi^S_s) \\
&\quad - c_3(\varphi^L_p\varphi^L_s|\varphi^S_r\varphi^S_q) - c_3(\varphi^S_p\varphi^S_s|\varphi^L_r\varphi^L_q) - c_3(\varphi^S_p\varphi^S_s|\varphi^S_r\psi^S_q) \\
&\quad - c_4(\varphi^L_p\boldsymbol{\sigma}\varphi^S_q + \varphi^S_p\boldsymbol{\sigma}\varphi^L_q|\varphi^L_r\boldsymbol{\sigma}\varphi^S_s + \varphi^S_r\boldsymbol{\sigma}\varphi^L_s) \\
&\quad + c_4(\varphi^L_p\boldsymbol{\sigma}\varphi^S_s + \varphi^S_p\boldsymbol{\sigma}\varphi^L_s|\varphi^L_r\boldsymbol{\sigma}\varphi^S_q + \varphi^S_r\boldsymbol{\sigma}\varphi^L_q)]\mathbf{P}_{sr} && (30)\\
&\approx [c_2(\varphi^L_p\varphi^L_q|\varphi^S_{i_A}\varphi^S_{i_A}) + c_2(\varphi^S_p\varphi^S_q|\varphi^L_{i_A}\varphi^L_{i_A}) + c_2(\varphi^S_p\varphi^S_q|\varphi^S_{i_A}\varphi^S_{i_A}) \\
&\quad - c_3(\varphi^L_p\varphi^L_{i_A}|\varphi^S_{i_A}\varphi^S_q) - c_3(\varphi^S_p\varphi^S_{i_A}|\varphi^L_{i_A}\varphi^L_q) - c_3(\varphi^S_p\varphi^S_{i_A}|\varphi^S_{i_A}\varphi^S_q) \\
&\quad - c_4(\varphi^L_p\boldsymbol{\sigma}\varphi^S_q + \varphi^S_p\boldsymbol{\sigma}\varphi^L_q|\varphi^L_{i_A}\boldsymbol{\sigma}\varphi^S_{i_A} + \varphi^S_{i_A}\boldsymbol{\sigma}\varphi^L_{i_A}) \\
&\quad + c_4(\varphi^L_p\boldsymbol{\sigma}\varphi^S_{i_A} + \varphi^S_p\boldsymbol{\sigma}\varphi^L_{i_A}|\varphi^L_{i_A}\boldsymbol{\sigma}\varphi^S_q + \varphi^S_{i_A}\boldsymbol{\sigma}\varphi^L_q)]n_{i_A}, && (31)\\
\tilde{\mathbf{S}}_{pq} &= \langle\varphi^L_p|\varphi^L_q\rangle + \langle\varphi^S_p|\varphi^S_q\rangle, && (32)
\end{aligned}$$

$$\mathbf{P} = \mathbf{C}\mathbf{n}\mathbf{C}^\dagger, \quad \mathbf{P}^0 = \bigoplus_A \mathbf{n}_A, \quad Tr(\mathbf{P}\tilde{\mathbf{S}}) = Tr\mathbf{P}^0 = N. \tag{33}$$

Here, the Mulliken notation for the two-electron integrals has been employed. The two-body term $\mathbf{G}$ (27) is decomposed deliberately into a term $\mathbf{G}^{LL}$ (29) that involves only the large components and a term $\mathbf{G}^{SS}$ (30) that involves the small components. The latter is further approximated by replacing the molecular density matrix $\mathbf{P}$ with $\mathbf{P}^0$ (33), the direct sum of the diagonal atomic density matrices. With this MSCA, only the G^{LL} (29) term, the analog of nonrelativistic Coulomb and exchange interactions, needs to be updated in the SCF iterations. Usually only the occupied and low-lying virtual ANSs are needed to form the backbone of the basis, which is to be augmented with some flat functions $\{\chi_\mu\}$ for describing the deformation and polarization of the atoms when forming the molecule. As $R_\mu(\mathbf{r}) \approx 1$ in the valence region (see Fig. 1), such flat functions χ_μ can be taken as

$$\chi_\mu = \begin{pmatrix} g_\mu \\ f_\mu \end{pmatrix}, \quad f_\mu = \frac{1}{2mc}\boldsymbol{\sigma}\cdot\mathbf{p}g_\mu \approx 0. \tag{34}$$

Unlike Eq. (7), the large and small component basis functions are here combined together. Such an "ANS+P"-type basis, single-zeta for each core shell and multiple-zeta for valence shells, is very efficient [8, 9]. Without much loss of accuracy, the two-electron integrals in G^{SS} can further be approximated. For instance, all

two-center $|SS)$ and $|LS)$ types of differential overlaps can be neglected. The Fock matrix elements among the added flat functions can be treated nonrelativistically, i.e.,

$$(\mathbf{F}^{\mathrm{Q4C}})_{\mu\nu} = \mathbf{T}_{\mu\nu} + \mathbf{V}_{\mu\nu} + (g_\mu g_\nu|\psi_i^L \psi_i^L) - (g_\mu \psi_i^L|\psi_i^L g_\nu), \tag{35}$$

while the elements between φ_p and χ_ν can be approximated as

$$\begin{aligned}(\mathbf{F}^{\mathrm{Q4C}})_{p\nu} \approx\ & \langle\varphi_p^L|T|g_\nu\rangle + \langle\varphi_p^L|V|g_\nu\rangle + [(\varphi_p^L g_\nu|\psi_i^L\psi_i^L) - (\varphi_p^L\psi_i^L|\psi_i^L g_\nu)]n_i \\ & + [(\varphi_p^L g_\nu|\varphi_{i_A}^S\varphi_{i_A}^S) - (\varphi_p^S\varphi_{i_A}^S|\varphi_{i_A}^L g_\nu)]n_{i_A} = (\mathbf{F}^{\mathrm{Q4C}})_{\nu p}^*\end{aligned} \tag{36}$$

according to the RKB0 prescription [9]. Overall, Q4C has little overhead over two-component approaches in orbital optimizations: Only the initial step is four-component, whereas all subsequent SCF iterations are just two-component. The many-electron second-quantized, normal-ordered Hamiltonian and equation can then be written as

$$H^{\mathrm{Q4C}}\Psi^{\mathrm{Q4C}} = E^{\mathrm{Q4C}}\Psi^{\mathrm{Q4C}}, \quad |\Psi^{\mathrm{Q4C}}\rangle = W^{\mathrm{Q4C}}|0\rangle, \tag{37}$$

$$H^{\mathrm{Q4C}} = H_0^{\mathrm{Q4C}} + H_1^{\mathrm{Q4C}} + H_2^{\mathrm{Q4C}}, \tag{38}$$

$$H_0^{\mathrm{Q4C}} = \langle 0|H^{\mathrm{Q4C}}|0\rangle, \tag{39}$$

$$H_1^{\mathrm{Q4C}} = \mathbf{F}_{pq}^{\mathrm{Q4C}}\{a_p^\dagger a_q\}, \tag{40}$$

$$H_2^{\mathrm{Q4C}} = \frac{1}{2}(\psi_p\psi_q|\psi_r\psi_s)\{a_p^\dagger a_r^\dagger a_s a_q\}. \tag{41}$$

The Hamiltonian H^{Q4C} (38) can be combined with any method (ansatz) for parameterizing the electronic wavefunction Ψ^{Q4C} (the wave operator W^{Q4C}) which has no reference to NES. As the major effect of the Gaunt/Breit interaction is already accounted for at the self-consistent field (SCF) level, it need not be included in the fluctuation potential (41) for routine calculations. On the other hand, one should not attempt to approximate the Coulomb integrals therein only to the $(LL|LL)$ portion because then not all designated electrons are correlated. This arises from the fact that the large components are themselves not normalized; see Eq. (32). Such an approximation becomes increasingly deteriorated when more and more electrons are to be correlated. Nonetheless, the transformation of the integrals involving the small components can be confined only to those M4S of appreciable small components with $\langle\varphi_p^S|\varphi_p^S\rangle$ larger than a preset threshold. As a matter of fact, the integrals $(\varphi_p\varphi_q|\varphi_r\varphi_s)$ can be transformed as a whole since the large and small components share the same coefficients $\mathbf{C}$. So, both the integral transformation and the correlation treatment are precisely the same as in two-component calculations. In addition, because of the initial four-component step, Q4C does not suffer from picture change errors (PCE) [15] which otherwise plague all two-component approaches.

With this Q4C ansatz, four-component relativistic calculations have been made as efficient as two-component relativistic ones at both the mean-field and correlated levels, without sacrificing the accuracy and simplicity.

Additional remarks should be made here. The above LCA4S (10) has a close analogy with the nonrelativistic LCAO (linear combination of atomic orbitals), but there exist some important differences. The A4S in Eq. (10) must be eigen or pseudo-eigen functions of the atomic Dirac equation, but no particular requirement needs to be imposed on the LCAO. Moreover, the LCAO can in principle achieve arbitrary accuracy in the nonrelativistic energy, whereas the above LC4AS has an inherent error of $O(c^{-4})$ in the relativistic energy, very small though. To respect such subtleties as well as the MSCA, the above manipulations have been visualized as "from atoms to molecule" (FATM) [9, 16] to emphasize the fact that the atomic information is employed to synthesize the molecular relativistic Hamiltonian. In other words, the interatomic interaction strength is taken as a perturbation parameter to expand the projector defining the molecular no-pair relativistic Hamiltonian. The zeroth order of the expansion is just the atomic approximation (11), while higher orders would correspond to a fragmental approximation for the projector. What makes this idea work is the atomic nature of the small component density. It will be shown that the same idea can also be employed to simplify two-component relativistic theories.

X2C

The development of two-component relativistic theories can be done at both operator and matrix levels. The former aims to first construct an *analytic* two-component Hamiltonian by transforming the Dirac operator (1) in one way or another [17–19] (for a recent comprehensive analysis, see Ref. [5]) and then introduces the matrix representation in the end, whereas the latter starts from the outset with a matrix representation of a Dirac-based mean-field Hamiltonian [e.g., F (12)] and hence obtains an *algebraic* two-component Hamiltonian that can directly be used in the calculation. The main problem with the operator type of formulations lies in that the underlying transformations cannot be performed in closed form (except for the trivial free-particle case), such that suitable approximations have to be introduced from the outset, thereby leading to various approximate two-component (A2C) Hamiltonians [20, 21]. Since the underlying approximations all stem from the atomic cores, such analytic A2C Hamiltonians can best be viewed as "valence-only" Hamiltonians, in the sense that they can describe only valence properties to a good accuracy. In contrast, the algebraic two-component Hamiltonians can reproduce all the solutions of the parent four-component matrix equation up to machinery accuracy and can hence be christened "exact two-component" (X2C). It was realized [22] only recently that the operator and matrix formulations are actually equivalent. The key is to make a proper matrix representation for the *non-expanded* analytic two-component Hamiltonian through appropriate resolutions of the identity (RI) [5]. Because of this, we need to focus only on the much simpler

matrix formulation, which is immune to singularities on one hand and does not invoke additional integrals other than those enter the four-component formalism on the other.

The matrix formulation of two-component Hamiltonians can be done in various ways. The one-step decoupling of the matrix Dirac equation (8) was initiated by Dyall [23] but was formulated properly and generally only by Kutzelnigg and Liu [24]. Both two-step [25,26] and multiple-step [27,28] exact decoupling procedures have also been proposed, with the resulting Hamiltonians coined Barysz-Sadlej-Snijders (BSS) and Douglas-Kroll-Hess (DKH), respectively. Note in passing that, although BSS and DKH are usually formulated in terms of analytic operators, they are de facto matrix formulations due to the use of RI [5]. In other words, the same DKH and BSS Hamiltonians can also be obtained [29] by starting with the separation of the matrix Dirac equation (8) into a free-particle part and the rest. In essence, all the three types of algebraic Hamiltonians share the same decoupling condition (vide post) and differ only in the renormalization (and hence picture change) [5]. The X2C and BSS Hamiltonians can even be mapped to each other due to the existence of a closed relationship [5]. Hereafter, we focus only on the simplest one-step X2C scheme.

To make the presentation as general as possible, we consider a generic four-component mean-field equation

$$F\psi_i = \epsilon_i \psi_i \tag{42}$$

with F given in Eq. (12), and introduce the following expansion of the M4S ψ_i

$$\psi_i = Z\tilde{\psi}_i, \qquad \tilde{\psi}_i = \begin{pmatrix} g_\mu \mathbf{A}_{\mu i} \\ g_\mu \mathbf{B}_{\mu i} \end{pmatrix} \tag{43}$$

$$= \begin{pmatrix} Z_{11} g_\mu \\ Z_{21} g_\mu \end{pmatrix} \mathbf{A}_{\mu i} + \begin{pmatrix} Z_{12} g_\mu \\ Z_{22} g_\mu \end{pmatrix} \mathbf{B}_{\mu i}, \tag{44}$$

where

$$Z = Z_g(Z_k + Z_m) = \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix}. \tag{45}$$

Inserting the expansion (44) into Eq. (42) leads to the following matrix equation

$$\mathbf{FC} = \mathbf{MC}\epsilon, \tag{46}$$

where

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}, \tag{47}$$

$$(\mathbf{F}_{11})_{\mu\nu} = \langle g_\mu | Z_{11}^\dagger F^{LL} Z_{11} + Z_{11}^\dagger F^{LS} Z_{21} + Z_{21}^\dagger F^{SL} Z_{11} + Z_{21}^\dagger F^{SS} Z_{21} | g_\nu \rangle, \tag{48}$$

$$\begin{aligned}(\mathbf{F}_{12})_{\mu\nu} &= \langle g_\mu | Z_{11}^\dagger F^{LL} Z_{12} + Z_{11}^\dagger F^{LS} Z_{22} + Z_{21}^\dagger F^{SL} Z_{12} + Z_{21}^\dagger F^{SS} Z_{22} | g_\nu \rangle \\ &= (\mathbf{F}_{21})_{\nu\mu}^*,\end{aligned} \tag{49}$$

$$(\mathbf{F}_{22})_{\mu\nu} = \langle g_\mu | Z_{12}^\dagger F^{LL} Z_{12} + Z_{12}^\dagger F^{LS} Z_{22} + Z_{22}^\dagger F^{SL} Z_{12} + Z_{22}^\dagger F^{SS} Z_{22} | g_\nu \rangle, \tag{50}$$

$$(\mathbf{M}_{11})_{\mu\nu} = \langle g_\mu | Z_{11}^\dagger Z_{11} + Z_{21}^\dagger Z_{21} | g_\nu \rangle, \tag{51}$$

$$(\mathbf{M}_{12})_{\mu\nu} = \langle g_\mu | Z_{11}^\dagger Z_{12} + Z_{21}^\dagger Z_{22} | g_\nu \rangle = (\mathbf{M}_{21})_{\nu\mu}^*, \tag{52}$$

$$(\mathbf{M}_{22})_{\mu\nu} = \langle g_\mu | Z_{12}^\dagger Z_{12} + Z_{22}^\dagger Z_{22} | g_\nu \rangle. \tag{53}$$

To formulate the X2C counterpart of Eq. (46), we first introduce the following formal relations

$$\mathbf{B} = \mathbf{X}\mathbf{A}, \quad \tilde{\mathbf{A}} = \tilde{\mathbf{X}}\tilde{\mathbf{B}} \tag{54}$$

between the small and large coefficients for the PES and NES, respectively. We then seek a transformation $\mathbf{U}$ that can block-diagonalize both the $\mathbf{F}$ and $\mathbf{M}$ matrices in Eq. (47). It can generally be written as

$$\mathbf{U} = \mathbf{U}_N \mathbf{U}_D, \quad \mathbf{U}_N = \begin{pmatrix} \mathbf{R}_+^\dagger & 0 \\ 0 & \mathbf{R}_-^\dagger \end{pmatrix}, \quad \mathbf{U}_D = \begin{pmatrix} \mathbf{I} & \mathbf{X}^\dagger \\ \tilde{\mathbf{X}}^\dagger & \mathbf{I} \end{pmatrix}, \tag{55}$$

where $\mathbf{U}_D$ does the decoupling, whereas $\mathbf{U}_N$ reestablishes the normalization. By requiring that

$$\mathbf{U}\mathbf{F}\mathbf{U}^\dagger = \begin{pmatrix} \mathbf{F}_+^{\mathrm{X}} & 0 \\ 0 & \mathbf{F}_-^{\mathrm{X}} \end{pmatrix}, \tag{56}$$

we obtain

$$\mathbf{F}_+^{\mathrm{X}} = \mathbf{R}_+^\dagger \tilde{\mathbf{F}}_+^{\mathrm{X}} \mathbf{R}_+, \quad \mathrm{X} = \mathrm{NESC}, \mathrm{SESC}, \tag{57}$$

$$\mathbf{F}_-^{\mathrm{X}} = \mathbf{R}_-^\dagger \tilde{\mathbf{F}}_-^{\mathrm{X}} \mathbf{R}_-, \quad \mathrm{X} = \mathrm{NESC}, \mathrm{SESC}, \tag{58}$$

$$\bar{\mathbf{L}}_+ + \tilde{\mathbf{X}}^\dagger \mathbf{L}_+ = \bar{\mathbf{L}}_-^\dagger + \mathbf{L}_-^\dagger \mathbf{X} = 0, \tag{59}$$

where

$$\tilde{\mathbf{F}}_+^{\mathrm{NESC}} = \mathbf{L}_+ + \mathbf{X}^\dagger \bar{\mathbf{L}}_+ = \mathbf{F}_{11} + \mathbf{F}_{12}\mathbf{X} + \mathbf{X}^\dagger \mathbf{F}_{21} + \mathbf{X}^\dagger \mathbf{F}_{22} \mathbf{X}, \tag{60}$$

$$\mathbf{L}_+ = \mathbf{F}_{11} + \mathbf{F}_{12}\mathbf{X}, \quad \bar{\mathbf{L}}_+ = \mathbf{F}_{21} + \mathbf{F}_{22}\mathbf{X}, \tag{61}$$

$$\tilde{\mathbf{F}}_{-}^{\mathrm{NESC}} = \mathbf{L}_{-} + \tilde{\mathbf{X}}^{\dagger}\bar{\mathbf{L}}_{-} = \mathbf{F}_{22} + \mathbf{F}_{21}\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^{\dagger}\mathbf{F}_{12} + \tilde{\mathbf{X}}^{\dagger}\mathbf{F}_{11}\tilde{\mathbf{X}}, \tag{62}$$

$$\mathbf{L}_{-} = \mathbf{F}_{22} + \mathbf{F}_{21}\tilde{\mathbf{X}}, \quad \bar{\mathbf{L}}_{-} = \mathbf{F}_{12} + \mathbf{F}_{11}\tilde{\mathbf{X}}. \tag{63}$$

Equation (59) is the state-universal decoupling condition [24]. By further requiring that

$$\mathbf{U}\mathbf{M}\mathbf{U}^{\dagger} = \begin{pmatrix} \mathbf{S}_{+} & 0 \\ 0 & \mathbf{S}_{-} \end{pmatrix}, \tag{64}$$

we obtain the following relations:

$$\mathbf{R}_{+}^{\dagger}\tilde{\mathbf{S}}_{+}\mathbf{R}_{+} = \mathbf{S}_{+}, \tag{65}$$

$$\mathbf{R}_{-}^{\dagger}\tilde{\mathbf{S}}_{-}\mathbf{R}_{-} = \mathbf{S}_{-}, \tag{66}$$

$$\bar{\mathbf{M}}_{+} + \tilde{\mathbf{X}}^{\dagger}\mathbf{M}_{+} = \bar{\mathbf{M}}_{-}^{\dagger} + \mathbf{M}_{-}^{\dagger}\mathbf{X} = 0, \tag{67}$$

where

$$\tilde{\mathbf{S}}_{+} = \mathbf{M}_{+} + \mathbf{X}^{\dagger}\bar{\mathbf{M}}_{+}, \tag{68}$$

$$\mathbf{M}_{+} = \mathbf{M}_{11} + \mathbf{M}_{12}\mathbf{X}, \quad \bar{\mathbf{M}}_{+} = \mathbf{M}_{21} + \mathbf{M}_{22}\mathbf{X}, \tag{69}$$

$$\tilde{\mathbf{S}}_{-} = \mathbf{M}_{-} + \tilde{\mathbf{X}}^{\dagger}\bar{\mathbf{M}}_{-}, \tag{70}$$

$$\mathbf{M}_{-} = \mathbf{M}_{22} + \mathbf{M}_{21}\tilde{\mathbf{X}}, \quad \bar{\mathbf{M}}_{-} = \mathbf{M}_{12} + \mathbf{M}_{11}\tilde{\mathbf{X}}. \tag{71}$$

Equation (67) implies that $\tilde{\mathbf{X}}$ and $\mathbf{X}$ are not independent but are related as

$$\tilde{\mathbf{X}} = -(\mathbf{M}_{+}^{\dagger})^{-1}\bar{\mathbf{M}}_{+}^{\dagger}, \quad \mathbf{X} = (\mathbf{M}_{-}^{\dagger})^{-1}\bar{\mathbf{M}}_{-}^{\dagger}, \tag{72}$$

in terms of which the decoupling condition (59) can be rewritten as

$$\mathbf{M}_{+}^{-1}\mathbf{L}_{+} = \bar{\mathbf{M}}_{+}^{-1}\bar{\mathbf{L}}_{+} \text{ or } \mathbf{M}_{-}^{-1}\mathbf{L}_{-} = \bar{\mathbf{M}}_{-}^{-1}\bar{\mathbf{L}}_{-}. \tag{73}$$

Since the NESC (normalized elimination of the small component) [23] Hamiltonians $\tilde{\mathbf{F}}_{+}^{\mathrm{NESC}}$ (60) and $\tilde{\mathbf{F}}_{-}^{\mathrm{NESC}}$ (62) are manifestly Hermitian, the SESC (symmetrized elimination of the small component) [8, 9] counterparts can readily be obtained as

$$\tilde{\mathbf{F}}_{+}^{\mathrm{SESC}} = \frac{1}{2}[(\mathbf{L}_{+} + \mathbf{X}^{\dagger}\bar{\mathbf{L}}_{+}) + c.c.] = \frac{1}{2}[\tilde{\mathbf{S}}_{+}\mathbf{M}_{+}^{-1}\mathbf{L}_{+} + c.c.], \tag{74}$$

$$\tilde{\mathbf{F}}_{-}^{\mathrm{SESC}} = \frac{1}{2}[(\mathbf{L}_{-} + \tilde{\mathbf{X}}^{\dagger}\bar{\mathbf{L}}_{-}) + c.c.] = \frac{1}{2}[\tilde{\mathbf{S}}_{-}\mathbf{M}_{-}^{-1}\mathbf{L}_{-} + c.c.], \tag{75}$$

where use of the relation (73) has been made. Note that both NESC and SESC are still in the Dirac picture featured by the relativistic metric (68)/(70), viz.,

$$\tilde{\mathbf{F}}_{+}^{\mathrm{X}}\mathbf{A} = \tilde{\mathbf{S}}_{+}\mathbf{A}\boldsymbol{\epsilon}_{+}, \quad \mathrm{X} = \mathrm{NESC}, \mathrm{SESC}, \tag{76}$$

$$\tilde{\mathbf{F}}_{-}^{\mathrm{X}}\tilde{\mathbf{B}} = \tilde{\mathbf{S}}_{-}\tilde{\mathbf{B}}\boldsymbol{\epsilon}_{-}, \quad \mathrm{X} = \mathrm{NESC}, \mathrm{SESC}. \tag{77}$$

To go to the Schrödinger picture, the nonrelativistic metric, i.e.,

$$(\mathbf{S}_{+})_{\mu\nu} = \langle g_\mu | Z_{11}^{\dagger} Z_{11} | g_\nu \rangle, \quad (\mathbf{S}_{-})_{\mu\nu} = \langle g_\mu | Z_{22}^{\dagger} Z_{22} | g_\nu \rangle, \tag{78}$$

should be adopted. Note in passing that an "intermediate picture" is also possible by choosing a different definition for $\mathbf{S}_{+}/\mathbf{S}_{-}$. However, the nonrelativistic metric (78) is most preferred since it allows an easy interface with the nonrelativistic machinery. This is particularly so for correlated calculations, where the use of untransformed two-electron integrals is only valid for the nonrelativistic metric.

With the above quantities, the unitary transformation of Eq. (46), viz.,

$$\mathbf{U}\mathbf{F}\mathbf{U}^{\dagger}(\mathbf{U}^{\dagger})^{-1}\mathbf{C} = \mathbf{U}\mathbf{M}\mathbf{U}^{\dagger}(\mathbf{U}^{\dagger})^{-1}\mathbf{C}\boldsymbol{\epsilon}, \tag{79}$$

gives rise to

$$\begin{pmatrix} \mathbf{F}_{+}^{\mathrm{X}} & 0 \\ 0 & \mathbf{F}_{-}^{\mathrm{X}} \end{pmatrix} \begin{pmatrix} \mathbf{C}_{+} & 0 \\ 0 & \mathbf{C}_{-} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{+} & 0 \\ 0 & \mathbf{S}_{-} \end{pmatrix} \begin{pmatrix} \mathbf{C}_{+} & 0 \\ 0 & \mathbf{C}_{-} \end{pmatrix} \begin{pmatrix} \boldsymbol{\epsilon}_{+} & 0 \\ 0 & \boldsymbol{\epsilon}_{-} \end{pmatrix}, \tag{80}$$

or

$$\mathbf{F}_{+}^{\mathrm{X}}\mathbf{C}_{+} = \mathbf{S}_{+}\mathbf{C}_{+}\boldsymbol{\epsilon}_{+}, \quad \mathrm{X} = \mathrm{NESC}, \mathrm{SESC}, \tag{81}$$

$$\mathbf{F}_{-}^{\mathrm{X}}\mathbf{C}_{-} = \mathbf{S}_{-}\mathbf{C}_{-}\boldsymbol{\epsilon}_{-}, \quad \mathrm{X} = \mathrm{NESC}, \mathrm{SESC}, \tag{82}$$

which are the desired X2C equations for the PES and NES, respectively. In view of the identity

$$\mathbf{C} = \mathbf{U}^{\dagger} \begin{pmatrix} \mathbf{C}_{+} & 0 \\ 0 & \mathbf{C}_{-} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{+}\mathbf{C}_{+} & \tilde{\mathbf{X}}\mathbf{R}_{-}\mathbf{C}_{-} \\ \mathbf{X}\mathbf{R}_{+}\mathbf{C}_{+} & \mathbf{R}_{-}\mathbf{C}_{-} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \tilde{\mathbf{X}}\tilde{\mathbf{B}} \\ \mathbf{X}\mathbf{A} & \tilde{\mathbf{B}} \end{pmatrix}, \tag{83}$$

the relationships between $\mathbf{C}_{+}$ and A and between $\mathbf{C}_{-}$ and $\tilde{\mathbf{B}}$ read

$$\mathbf{C}_{+} = \mathbf{R}_{+}^{-1}\mathbf{A}, \quad \mathbf{C}_{-} = \mathbf{R}_{-}^{-1}\tilde{\mathbf{B}}. \tag{84}$$

At this stage, the renormalization matrices $\mathbf{R}_{+}$ and $\mathbf{R}_{-}$ still remain to be specified. Unfortunately, they cannot uniquely be determined by conditions (65) and (66), respectively. Yet, the $\mathbf{R}_{+}$ ($\mathbf{R}_{-}$) is just a transformation matrix between the $\{\varphi'_u = Z_{11}g'_\mu\}$ ($\{\tilde{\chi}'_u = Z_{22}g'_\mu\}$) and $\{\varphi_\mu = Z_{11}g_\mu\}$ ($\{\tilde{\chi}_u = Z_{22}g_\mu\}$) bases due to the action of some operator O ($\tilde{O}$), viz.,

$$\varphi' = O\varphi = \varphi\mathbf{R}_{+}, \quad \tilde{\chi}' = \tilde{O}\tilde{\chi} = \tilde{\chi}\mathbf{R}_{-}. \tag{85}$$

Requiring that O and $\tilde{O}$ be Hermitian leads to

$$\mathbf{S}_+\mathbf{R}_+ = \mathbf{O} = \mathbf{O}^\dagger = (\mathbf{S}_+\mathbf{R}_+)^\dagger = \mathbf{R}_+^\dagger\mathbf{S}_+, \tag{86}$$

$$\mathbf{S}_-\mathbf{R}_- = \tilde{\mathbf{O}} = \tilde{\mathbf{O}}^\dagger = (\mathbf{S}_-\mathbf{R}_-)^\dagger = \mathbf{R}_-^\dagger\mathbf{S}_-. \tag{87}$$

Conditions (65) and (86) then determine $\mathbf{R}_+$, while (66) and (87) determine $\mathbf{R}_-$ uniquely [5,22], viz.,

$$\mathbf{R}_+ = \mathbf{S}_+^{-1/2}[\mathbf{S}_+^{-1/2}\tilde{\mathbf{S}}_+\mathbf{S}_+^{-1/2}]^{-1/2}\mathbf{S}_+^{1/2} \equiv [\mathbf{S}_+^{-1}\tilde{\mathbf{S}}_+]^{-1/2}, \tag{88}$$

$$\mathbf{R}_- = \mathbf{S}_-^{-1/2}[\mathbf{S}_-^{-1/2}\tilde{\mathbf{S}}_-\mathbf{S}_-^{-1/2}]^{-1/2}\mathbf{S}_-^{1/2} \equiv [\mathbf{S}_-^{-1}\tilde{\mathbf{S}}_-]^{-1/2}. \tag{89}$$

Alternatively, conditions (65) and (87), as well as (66) and (87), can be unified into a single condition

$$\mathbf{R}_+\mathbf{S}_+^{-1}\tilde{\mathbf{S}}_+\mathbf{R}_+ = \mathbf{I}, \quad \mathbf{R}_-\mathbf{S}_-^{-1}\tilde{\mathbf{S}}_-\mathbf{R}_- = \mathbf{I}, \tag{90}$$

such that $\mathbf{R}_+$ and $\mathbf{R}_-$ can be calculated as

$$\tilde{\mathbf{S}}_+\mathbf{Z}_+ = \mathbf{S}_+\mathbf{Z}_+\Lambda_+, \quad \mathbf{R}_+ = \mathbf{Z}_+\Lambda_+^{-1/2}\mathbf{Z}_+^{-1} = \mathbf{Z}_+\Lambda_+^{-1/2}\mathbf{Z}_+^\dagger\mathbf{S}_+, \tag{91}$$

$$\tilde{\mathbf{S}}_-\mathbf{Z}_- = \mathbf{S}_-\mathbf{Z}_-\Lambda_-, \quad \mathbf{R}_- = \mathbf{Z}_-\Lambda_-^{-1/2}\mathbf{Z}_-^{-1} = \mathbf{Z}_-\Lambda_-^{-1/2}\mathbf{Z}_-^\dagger\mathbf{S}_-. \tag{92}$$

The so-normalized two-component eigenvectors $\psi_{p,+}^{\mathrm{FW}} = \sum_\mu g_\mu(\mathbf{C}_+)_{\mu p}$ are closest [30] to the large components $\psi_{p,+}^L$ of the PES in the least-squares sense,

$$\min \sum_p \|\psi_{p,+}^{\mathrm{FW}} - \psi_{p,+}^L\|^2 = \min \sum_p \langle \psi_{p,+}^{\mathrm{FW}} - \psi_{p,+}^L | \psi_{p,+}^{\mathrm{FW}} - \psi_{p,+}^L \rangle. \tag{93}$$

In contrast, the BSS [25,26] and DKH [27,28] approaches amount to using different renormalizations. However, they share the same decoupling condition (73) as X2C [5].

Some remarks are in order.

1. The above formulation of the X2C Hamiltonians is more general than the previous ones [22–24] that were based solely on the RKB condition (6) or $Z_{11} = 1$, $Z_{12} = Z_{21} = 0$, and $Z_{22} = \frac{1}{2mc}\boldsymbol{\sigma}\cdot\mathbf{p}$ in Eq. (45). Here, various kinetic and magnetic balances (see Table 1 in Ref. [31]) have been incorporated in a unified manner. In particular, an X2C quantum electrodynamics (QED) approach is also possible by treating the EVP-ESE Q self-consistently. Two special cases deserve to be mentioned here. One is the relativistic description of real positrons, which are accommodated by high-lying NES. In this case, the dual kinetic balance (DKB) [32] is most recommended since it describes the PES and NES on an equal footing and additionally offers variational safety in conjunction with Gaussian functions [10]. To simplify the integral evaluations, one can use DKB for the one-electron terms while RKB for the two-electron terms in the spirit of "different representations for different operators." The other is the relativistic

description of electrons in the presence of an ultra-strong magnetic field. In this case, the dual magnetic balance (DMB) [10] is most recommended. Likewise, one can use DMB only for the one-electron terms while the restricted magnetic balance (RMB) for the two-electron terms.

2. Equations (81) [or (76)] and (73) form a coupled pair of equations, each of which is of dimension $2N^L \times 2N^L$. In principle, they can be solved iteratively through a dual-level algorithm. In the macro-iterations, Eq. (81) [or (76)] is solved and the mean-field potential is updated. Meanwhile, for each macro-iteration, some micro-iterations are performed to determine $\mathbf{X}$ via Eq. (73), which require typically 3–4 cycles to converge [33]. At convergence, one obtains results which fully duplicate those of the original Dirac matrix (47). However, the computation is more expensive than directly solving the original matrix Dirac equation (even for one-electron systems) and becomes impractical for molecular applications. A more practical way is to find an accurate estimate of $\mathbf{X}$ from the outset. As shown by Eq. (54), $\mathbf{X}$ is nothing but the ratio between the small- and large-component coefficients. Therefore, the previous idea of FATM [9, 16] underlying the Q4C approach can also be used here. Anyhow, the interatomic interaction strength is much smaller and therefore more effective than the perturbation parameters e^2 (coupling strength) and c^{-1} employed in the DKH and BSS approaches, respectively. Various schemes have been designed following this philosophy; see Ref. [5]. A particular point to be made here is that, while the popular one-electron approximation [34] to $\mathbf{X}$ works very well for electronic structure calculations, it results in sizeable errors for, e.g., NMR (nuclear magnetic resonance) shieldings [31]. In contrast, the superposition of atomic/fragmental $\mathbf{X}$ matrices works uniformly well. It has two additional merits. First, it also allows for an easy cure [8] of two-electron PCE, which is essential for extended systems. Second, it allows for a natural "1/2C" hybrid treatment of the light (one-component; 1C) and heavy (two-component; 2C) atoms in the molecule, so as to interface seamlessly the Schrödinger and Dirac equations. There have been attempts [35, 36] to approximate the renormalization matrix $\mathbf{R}_+$ (88) as well. However, $\mathbf{R}_+$ is much less local than $\mathbf{X}$, such that the approximations are generally not safe, especially for spatially adjacent heavy atoms with diffuse functions [37]. Since both the construction of the $\mathbf{R}_+$ matrix itself and its use in transforming the integrals are computationally very cheap, any approximation to $\mathbf{R}_+$ is not recommended.
3. Once the $\mathbf{X}$ matrix is obtained, the SESC Hamiltonian $\tilde{\mathbf{F}}_+^{\mathrm{SESC}}$ /$\mathbf{F}_+^{\mathrm{SESC}}$ has a great advantage over $\tilde{\mathbf{F}}_+^{\mathrm{NESC}}$/$\mathbf{F}_+^{\mathrm{NESC}}$: It does not involve $(SS|SS)$-type of integrals.
4. If wanted, the NES can also be obtained by solving Eq. (82) or (77) without any iteration.
5. The second-quantized, normal-ordered many-electron X2C Hamiltonian can be defined [9] as

$$H^{\mathrm{X2C}} = \langle 0|H^{\mathrm{X2C}}|0\rangle + (\mathbf{F}_+^{\mathrm{X}})_{pq}\{a_p^\dagger a_q\} + \frac{1}{2}(\psi_p\psi_q|\psi_r\psi_s)\{a_p^\dagger a_r^\dagger a_s a_q\},$$
$$\psi_p = \psi_{p,+}^{\mathrm{FW}}, \quad \mathrm{X} = \mathrm{NESC}, \mathrm{SESC}, \tag{94}$$

which differs from the Q4C Hamiltonian H^{Q4C} (38) only in the mean-field term. For comparison, A2C approaches [20, 21] usually adopt an approximately transformed one-electron operator h_+^{A2C} and the bare Coulomb operator, viz.,

$$H^{\mathrm{A2C}} = \sum_i^N h_+^{\mathrm{A2C}}(i) + \frac{1}{2}\sum_{ij}^N \frac{1}{r_{ij}}, \tag{95}$$

$$= \langle 0|H^{\mathrm{A2C}}|0\rangle + (\mathbf{F}_+^{\mathrm{A2C}})_{pq}\{a_p^\dagger a_q\} + \frac{1}{2}(\psi_p\psi_q|\psi_r\psi_s)\{a_p^\dagger a_r^\dagger a_s a_q\}, \tag{96}$$

$$(\mathbf{F}_+^{\mathrm{A2C}})_{pq} = (h_+^{\mathrm{A2C}})_{pq} + [(pq|ii) - (pi|iq)]n_i. \tag{97}$$

Apart from different one-electron terms, the $H^{\mathrm{X2C}}/H^{\mathrm{Q4C}}$ Hamiltonian differs from H^{A2C} also in that the latter neglects all the picture change corrections, whereas the former neglects only the picture change corrections to the genuine two-electron terms. Since the two-electron PCE stem solely from the innermost shells of heavy atoms, they can entirely be corrected at the mean-field level. Therefore, the $H^{\mathrm{X2C}}/H^{\mathrm{Q4C}}$ Hamiltonian is more accurate than albeit computationally the same as H^{A2C}.

6. Although defined only algebraically, the spin separation of the X2C (and Q4C) Hamiltonians is still possible, leading to a series of new Hamiltonians sf-X2C+so-DKHn/so-BSSn/so-X2Cn that are infinite order in scalar relativity and finite order in spin-orbit coupling [29, 30]. In this regard, the sf-X2C+so-DKH$_n$ Hamiltonian as a combination of the spin-free (sf) part of X2C and the nth order DKH (DKHn) type of spin-dependent (sd) terms is particularly promising [38].
7. The no-pair Q4C, X2C, and sf-X2C+sd-DKHn Hamiltonians, together with the with-pair effective QED ones [39], form a continuous and complete "Hamiltonian ladder" [3, 4]. One can just pick up the right Hamiltonian for the target physics and accuracy. Therefore, the development of relativistic Hamiltonians can be considered as completed.
8. Unlike DKH and BSS, the X2C formalism allows easy formulations of nuclear derives as well as electric/magnetic response properties. For instance, the responses of $\mathbf{X}$, $\mathbf{R}_+$, and $\tilde{\mathbf{F}}_+^{\mathrm{NESC}}$ can readily be obtained by expanding the respective Eqs. (73), (90), and (60) in powers of the perturbations; see, e.g., Ref. [31] for the formulation of NMR shieldings.

Finally, to confirm the above arguments, we compare Q4C and X2C with the full DKS equation (8) for the energy levels of the heaviest diatomic molecule of $E117_2$. To simplify the calculations, only the local density approximation (LDA) of the exchange-correlation functional was employed, in conjunction with a modified QZ4P Slater-type basis set [9] and the point nuclear model. It is seen from Table 1 that Q4C and X2C[NESC] are indeed very accurate for all the energy levels. The only caveat is that, at variance with X2C[NESC], X2C[SESC] is a bit more sensitive

Table 1 Energy levels (in a.u.) of $E117_2$ calculated with LDA at 3.425 Å. Mod**X**, model potential (superposition of spherical atomic densities) approximation to **X**; $\mathbf{X}_A^{\oplus} = \sum_A^{\oplus} \mathbf{X}_A$, superposition of atomic **X** matrices; $\mathbf{X}_{1e}$, one-electron **X** matrix

			X2C[NESC][a]			X2C[SESC][a]		
	DKS	Q4C[a]	Mod**X**	$\mathbf{X}_A^{\oplus}$	$\mathbf{X}_{1e}$	Mod**X**	$\mathbf{X}_A^{\oplus}$	$\mathbf{X}_{1e}$
Core shells								
1s1/2	−8046.3342	−0.0073	−0.0067	−0.0070	−0.0385	−0.0003	−0.0485	16.2793
2s1/2	−1675.0098	−0.0016	−0.0014	−0.0015	−0.0119	−0.0001	−0.0255	2.5114
2p1/2	−1622.3629	−0.0015	−0.0014	−0.0014	−0.0220	−0.0001	−0.0034	−5.4746
2p3/2	−1101.8697	−0.0009	−0.0008	−0.0008	−0.0106	0.0000	−0.0031	3.4665
3s1/2	−453.2514	−0.0004	−0.0004	−0.0004	−0.0028	0.0000	−0.0190	0.5444
3p1/2	−430.5377	−0.0004	−0.0004	−0.0004	−0.0045	0.0000	−0.0030	−1.3213
3p3/2	−303.4697	−0.0003	−0.0002	−0.0002	−0.0024	0.0000	−0.0029	0.8667
3d3/2	−273.3387	−0.0002	−0.0002	−0.0002	−0.0025	0.0000	−0.0029	−1.1775
3d5/2	−253.2403	−0.0002	−0.0002	−0.0002	−0.0019	0.0000	−0.0029	0.8993
4s1/2	−132.6890	−0.0001	−0.0001	−0.0001	−0.0007	0.0000	−0.0183	0.1422
4p1/2	−122.0502	−0.0001	−0.0001	−0.0001	−0.0011	0.0000	−0.0029	−0.4069
4p3/2	−85.2800	−0.0001	−0.0001	−0.0001	−0.0006	0.0000	−0.0029	0.2461
4d3/2	−70.4765	−0.0001	−0.0001	−0.0001	−0.0005	0.0000	−0.0029	−0.3538
4d5/2	−64.9376	−0.0001	0.0000	0.0000	−0.0004	0.0000	−0.0029	0.2463
4f5/2	−45.5699	0.0000	0.0000	0.0000	−0.0002	0.0000	−0.0029	−0.3081
4f7/2	−43.9159	0.0000	0.0000	0.0000	−0.0001	0.0000	−0.0029	0.2187
5s1/2	−36.3363	0.0000	0.0000	0.0000	−0.0001	0.0000	−0.0183	0.0379
5p1/2	−31.5952	0.0000	0.0000	0.0000	−0.0003	0.0000	−0.0029	−0.1302
5p3/2	−20.6721	0.0000	0.0000	0.0000	−0.0001	0.0000	−0.0029	0.0691
5d3/2	−14.2983	0.0000	0.0000	0.0000	−0.0001	0.0000	−0.0029	−0.1014
5d5/2	−12.8304	0.0000	0.0000	0.0000	−0.0001	0.0000	−0.0028	0.0628
6s1/2	−7.5898	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0198	0.0080
6p1/2	−5.7986	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	−0.0349
5f5/2	−5.1971	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	−0.0706
5f7/2	−4.8612	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	0.0434
6p3/2	−3.2453	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	0.0141
Valence shells								
E1/2g	−1.2331	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	−0.0183
E1/2u	−1.2322	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0029	−0.0183
E3/2u	−1.2260	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	−0.0184
E3/2g	−1.2259	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	−0.0184
E1/2g	−1.0055	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0029	0.0088
E1/2u	−1.0023	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0029	0.0089
E3/2u	−1.0010	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0029	0.0088
E3/2g	−0.9997	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0029	0.0089
E5/2g	−0.9945	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	0.0088
E5/2u	−0.9944	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0028	0.0088
E1/2g	−0.8968	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0078	0.0007
E1/2u	−0.8864	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0161	0.0007
E1/2g	−0.4646	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0027	−0.0051
E1/2u	−0.4493	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0027	−0.0053
E1/2g	−0.2033	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0024	0.0007
E3/2u	−0.1874	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0022	0.0007
E3/2g	−0.1532	0.0000	0.0000	0.0000	0.0000	0.0000	−0.0023	0.0009

[a] Deviations from DKS

to the approximations to $\mathbf{X}$. This is because the NESC Hamiltonian $\tilde{\mathbf{F}}_+^{\mathrm{NESC}}$ (60) arises from the stationarity condition for the energy variation [24], whereas the SESC Hamiltonian $\tilde{\mathbf{F}}_+^{\mathrm{SESC}}$ (74) arises from $\tilde{\mathbf{F}}_+^{\mathrm{NESC}}$ (60) via the exact decoupling condition (73) which cannot be satisfied by an approximate $\mathbf{X}$. Still, however, the diatomic approximation to $\mathbf{X}$, i.e., $\mathbf{X} \approx \sum_{A<B}^{\oplus} \mathbf{X}_{AB}$, is sufficiently accurate for X2C[SESC] in case of polyatomic systems.

Summary

The essential ideas underlying the Q4C and X2C approaches have been highlighted. The take-home message is that four- and two-component relativistic approaches can be made fully equivalent under very mild conditions and are hence equally good for routine calculations of systems containing heavy elements. Yet, there may exist some situations where four-component is a somewhat better choice than two-component or vice versa. For instance, four-component is better suited for spectroscopies of core electrons, while spin-separated two-component Hamiltonians are better suited for two-step treatments of spin-free and spin-dependent relativistic effects in systems where spin-orbit couplings are not too strong.

Acknowledgements This work was supported by the NSFC (Project Nos. 21033001, 21273011, and 21290192).

References

1. Liu W (2012) The big picture of relativistic molecular quantum mechanics. AIP Conf Proc 1456:62
2. Liu W (2015) Effective quantum electrodynamics Hamiltonians: a tutorial review. Int J Quantum Chem 116:631
3. Liu W (2104) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59
4. Liu W (2014) Perspective: relativistic Hamiltonians. Int J Quantum Chem 114:983
5. Liu W (2010) Ideas of relativistic quantum chemistry. Mol Phys 108:1679
6. Saue T (2011) Relativistic Hamiltonians for chemistry: a primer. Chem Phys Chem 12:3077
7. Peng D, Reiher M (2012) Exact decoupling of the relativistic Fock operator. Theor Chem Acc 131:1081
8. Liu W, Peng D (2006) Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory. J Chem Phys 125:044102; (E) 125:149901 (2006)
9. Peng D, Liu W, Xiao Y, Cheng L (2007) Making four- and two-component relativistic density functional methods fully equivalent based on the idea of "from atoms to molecule". J Chem Phys 127:104106
10. Sun Q, Liu W, Kutzelnigg W (2011) Comparison of restricted, unrestricted, inverse, and dual kinetic balances for four-component relativistic calculations. Theor Chem Acc 129:423
11. Stanton RE, Havriliak S (1984) Kinetic balance: a partial solution to the problem of variational safety in Dirac calculations. J Chem Phys 81:1910
12. Dyall KG, Enevoldsen T (1999) Interfacing relativistic and nonrelativistic methods. III. Atomic 4-spinor expansions and integral approximations. J Chem Phys 111:10000
13. Liu W (1995) Ph. D Dissertation, Peking University; Liu W, Hong G, Dai D, Li L, Dolg M (1997) The Beijing four-component density functional program package (BDF) and its application to EuO, EuS, YbO and YbS. Theor Chem Acc 96:75

14. Shabaev VM, Tupitsyn II, Yerokhin VA (2013) Model operator approach to the Lamb shift calculations in relativistic many-electron atoms. Phys Rev A 88:012513
15. Kellö V, Sadlej AJ (1998) Picture change and calculations of expectation values in approximate relativistic theories. Int J Quantum Chem 68:159
16. Liu W (2007) New advances in relativistic quantum chemistry. Prog Chem 19:833
17. Foldy LL, Wouthuysen SA (1950) On the Dirac theory of spin 1/2 particles and its non-relativistic limit. Phys Rev 78:29
18. Douglas M, Kroll NM (1974) Quantum electrodynamical corrections to the fine structure of helium. Ann Phys 82:89
19. Kutzelnigg W (1997) Relativistic one-electron Hamiltonians 'for electrons only' and the variational treatment of the Dirac equation. Chem Phys 225:203
20. Hess BA (1986) Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. Phys Rev A 33:3742
21. van Lenthe E, Baerends EJ, Snijders JG (1993) Relativistic regular two-component Hamiltonians. J Chem Phys 99:4597
22. Liu W, Peng D (2009) Exact two-component Hamiltonians revisited. J Chem Phys 131:031104
23. Dyall KG (1997) Interfacing relativistic and nonrelativistic methods. I. Normalized elimination of the small component in the modified Dirac equation. J Chem Phys 106:9618
24. Kutzelnigg W, Liu W (2005) Quasirelativistic theory equivalent to fully relativistic theory. J Chem Phys 123:241102
25. Barysz M, Sadlej AJ, Snijders JG (1997) Nonsingular two/one-component relativistic Hamiltonians accurate through arbitrary high order in α^2. Int J Quant Chem 65:225
26. Barysz M, Sadlej AJ (2001) Two-component methods of relativistic quantum chemistry: from the Douglas-Kroll approximation to the exact two-component formalism. J Mol Struct (THEOCHEM) 573:181
27. Reiher M, Wolf A (2004) Exact decoupling of the Dirac Hamiltonian. I. General theory. J Chem Phys 121:2037
28. Reiher M, Wolf A (2004) Exact decoupling of the Dirac Hamiltonian. II. The generalized Douglas-Kroll-Hess transformation up to arbitrary order. J Chem Phys 121:10945
29. Li Z, Xiao Y, Liu W (2012) On the spin separation of algebraic two-component relativistic Hamiltonians. J Chem Phys 137:154114
30. Li Z, Xiao Y, Liu W (2014) On the spin separation of algebraic two-component relativistic Hamiltonians: Molecular properties. J Chem Phys 141:054111
31. Sun Q, Xiao Y, Liu W (2012) Exact two-component relativistic theory for NMR parameters: general formulation and pilot application. J Chem Phys 137:174105
32. Shabaev VM, Tupitsyn II, Yerokhin VA, Plunien G, Soff G (2004) Dual kinetic balance approach to basis-set expansions for the Dirac equation. Phys Rev Lett 93:130405
33. Liu W, Kutzelnigg W (2007) Quasirelativistic theory. II. Theory at matrix level. J Chem Phys 126:114107
34. Iliaš M, Saue T (2007) An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. J Chem Phys 126:064102
35. Seino J, Nakai H (2012) Local unitary transformation method for large-scale two-component relativistic calculations: case for a one-electron Dirac Hamiltonian. J Chem Phys 136:244102
36. Peng D, Reiher M (2012) Local relativistic exact decoupling. J Chem Phys 136:244108
37. Tamukong PK, Khait YG, Hoffmann MR, Li Z, Liu W (2014) Relativistic GVVPT2 multireference perturbation theory description of the electronic states of Y_2 and Tc_2. J Phys Chem A 118:1489
38. Li Z, Suo B, Zhang Y, Xiao Y, Liu W (2013) Combining spin-adapted open-shell TD-DFT with spin–orbit coupling. Mol Phys 111:3741
39. Liu W, Lindgren I (2013) Going beyond "no-pair relativistic quantum chemistry". J Chem Phys 139:014108

Sequential Decoupling of Negative-Energy States in Douglas–Kroll–Hess Theory

12

Markus Reiher

Contents

Introduction 396
Formal Exact-Decoupling 397
Foldy–Wouthuysen Transformations 398
Douglas–Kroll Transformations 399
Implementation of Douglas–Kroll Transformations 402
Relation to Other Exact-Decoupling Approaches 403
Conclusions and Outlook 405
References 406

Abstract

Here, we review the historical development, current status, and prospects of Douglas–Kroll–Hess theory as a quantum chemical relativistic electrons-only theory.

Keywords

Douglas-Kroll-Hess theory • Exact decoupling • Relativistic quantum chemistry • Dirac Hamiltonian • Unitary transformation • Block diagonalization • Exact two-component methods • Electrons-only Hamiltonian • Foldy-Wouthuysen transformation

M. Reiher (✉)
Laboratorium für Physikalische Chemie, ETH Zürich, Zürich, Switzerland
e-mail: markus.reiher@phys.chem.ethz.ch

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_3

Introduction

In this conceptual review, we describe the development of Douglas–Kroll–Hess (DKH) theory. While we discuss the essential concepts of this theory explicitly, we refer the reader for further technical details to recent reviews [1–4] (for a comprehensive background of relativistic quantum chemistry, see the monograph in Ref. [5]).

The symmetric occurrence of positive- and negative-energy continua in the spectrum of the field-free Dirac Hamiltonian is the basis for states describing electrons and antielectrons (positrons) in quantum electrodynamics (QED). The sophisticated framework of QED is neither feasible nor necessary for a theoretical description of matter at the molecular level. In fact, the quantization of the radiation field is hardly needed in molecular science. In a first-quantized theory, however, in which all electromagnetic interactions are described in terms of classical fields (electromagnetic scalar and vector potentials) rather than as being transmitted by photons as in second-quantized QED, the negative-energy continuum creates pathologies such as variational collapse and continuum dissolution because of the resulting boundlessness of the Dirac Hamiltonian [5].

These pathologies make Dirac's relativistic theory of the electron a difficult basis for standard numerical solution methods. Still, all technical issues can be solved in an orbital-based approach [6–8] – mostly by respecting the structure of the one-electron Hamiltonian (kinetic balance) and the proper exponentially decreasing long-range behavior of the electronic bound states when representing the one-electron positive-energy states on a grid or in terms of basis functions. Clearly, the negative-energy one-particle states, which are obtained whenever a one-electron operator containing the Dirac Hamiltonian – such as the four-dimensional Fock operator of four-component methods ("four-component" Fock operator) – is diagonalized, are to be omitted from the construction of the density matrix required for the calculation of the electron–electron potential and interaction energy.

In low-energy physics and therefore also in chemistry, electron–positron pair creation processes are energetically not accessible, and therefore, negative-energy states are usually not required for a sufficiently accurate quantum-mechanical description of molecular matter. For this reason, approaches have been searched for that produce relativistic (no-pair, electrons-only) Hamiltonians with an energy spectrum that resembles the positive-energy states only. We should note that these states feature positive energies even for bound electronic states as the zero-energy reference is not the state, in which all particles are found at rest and at infinitely large distance from each other. In this situation, which marks the nonrelativistic zero-energy reference, each particle still possesses a rest energy defined by the mass observed for the particle when it is at rest. The rest energy of an electron is given by the product of its rest mass and the square of the speed of light, mc^2, and it is so large that even bound electronic states will feature a positive energy when the rest energy is added to them (and if the attractive potential is not too strong as is the case for all atomic nuclei known).

An elegant option of removing the coupling to the charge-conjugated negative-energy states is the application of a unitary transformation U that block-diagonalizes the Dirac Hamiltonian h_D in such a way that two decoupled operator blocks, h_+ and h_-, emerge:

$$h_{bd} = U h_D U^\dagger = \begin{pmatrix} h_+ & \mathbf{0}_2 \\ \mathbf{0}_2 & h_- \end{pmatrix} \tag{1}$$

where $\mathbf{0}_2$ denotes a two-dimensional null matrix entering the off-diagonal blocks of h_{bd}. h_+ and h_- then account for the positive- and negative-energy states separately. Both h_+ and h_- are two-dimensional one-electron Hamiltonians. The 2×2 superstructure of h_D

$$h_D = k_{\mathscr{O}} + r_{\mathscr{E}} + v \tag{2}$$

with

$$k_{\mathscr{O}} = \begin{pmatrix} \mathbf{0}_2 & c\,\boldsymbol{\sigma}\cdot\mathbf{p} \\ c\,\boldsymbol{\sigma}\cdot\mathbf{p} & \mathbf{0}_2 \end{pmatrix} \tag{3}$$

and

$$r_{\mathscr{E}} = \begin{pmatrix} mc^2 & \mathbf{0}_2 \\ \mathbf{0}_2 & -mc^2 \end{pmatrix}, \tag{4}$$

where v is the potential energy operator, m the rest mass of the electron, c the speed of light, $\mathbf{p}$ the momentum operator, and $\boldsymbol{\sigma}$ the 3-vector of Pauli spin matrices, is preserved by the transformation in Eq. (1). In the standard representation, the kinetic energy operator $k_{\mathscr{O}}$ is off-diagonal (odd, "$\mathscr{O}$"), while an external electrostatic potential $v_{\mathscr{E}}$ in $v = v_{\mathscr{E}} + v_{\mathscr{O}}$ is block-diagonal (even, "$\mathscr{E}$") and vector-potential contributions $v_{\mathscr{O}}$ to v are odd. Also, the rest energy $r_{\mathscr{E}}$ is an even operator, hence the subscript.

Formal Exact-Decoupling

In the mid-1980s, a formal expression for the block-diagonalizing unitary transformation U was derived [9],

$$U = \begin{pmatrix} \left(1 + X^\dagger X\right)^{-1/2} & \left(1 + X^\dagger X\right)^{-1/2} X^\dagger \\ -\mathrm{e}^{\mathrm{i}\varphi}\left(1 + XX^\dagger\right)^{-1/2} X & \mathrm{e}^{\mathrm{i}\varphi}\left(1 + XX^\dagger\right)^{-1/2} \end{pmatrix} \tag{5}$$

as a function of the X-operator, which relates the large "L" and small "S" components,

$$\psi^S = X\psi^L, \tag{6}$$

of the 4-spinor, $\psi = (\psi^L, \psi^S)$. In their original work [9], Heully and coworkers chose $\varphi = \pi$ for the relative phase φ.

An expression for X depending on the energy eigenvalue ϵ can be easily derived from the Dirac equation,

$$X = \left(\epsilon - v + 2mc^2\right)^{-1} c\,\boldsymbol{\sigma}\cdot\mathbf{p}. \tag{7}$$

As the energy eigenvalue is the sought-for solution *after* applying the unitary transformation, the energy-dependent X-operator is not very useful. In fact, it was possible [9] to derive an equation for the determination of X that does not depend on the energy eigenvalue,

$$X = \frac{1}{2m_e c^2}\left\{c\,\boldsymbol{\sigma}\cdot\mathbf{p} - [X, V] - Xc\,\boldsymbol{\sigma}\cdot\mathbf{p}X\right\} \tag{8}$$

However, the solution of this equation for X was considered to be as complicated as the solution of the Dirac equation itself. However, it was not before the dawn of the new millennium that such an equation was solved by numerical means as a true option for exact decoupling [10]. It was this paper by Barysz and Sadlej that introduced the first infinite-order two-component (IOTC) method and that initiated the intense development of exact-decoupling methods in the first decade of the twenty-first century. We shall later discuss some of its ingredients in more detail (see section "Relation to Other Exact-Decoupling Approaches").

Foldy–Wouthuysen Transformations

Foldy and Wouthuysen were the first to find a block-diagonalizing unitary transformation in closed form [11], but only for the free-particle (field-free) Dirac Hamiltonian, for which $v = 0$. Unfortunately, such a closed-form solution is not known for the general many-electron case in an external field of atomic nuclei for two reasons: (i) already for a single electron in the presence of an external electrostatic potential $v_{\mathscr{E}}$, no closed-form expression for a unitary transformation can be constructed [12], and (ii) vector potentials, such as those emerging from external magnetic fields or the magnetic interaction of two electrons, as well as contributions from exchange integrals (in Hartree–Fock-type theories) to the off-diagonal superblock of the Dirac Hamiltonian add additional complexity to the problem [13]. A solution was already suggested by Foldy and Wouthuysen [11]. A sequence of unitary transformations can be used to suppress the off-diagonal blocks order by order in terms of an expansion parameter. Strictly speaking, exact decoupling is then obtained only after an infinite number of such transformations.

In physics, expansions of relativistic Hamiltonians are usually carried out order by order with respect to the inverse speed of light $1/c$, and so was the expansion of Foldy and Wouthuysen. A $1/c$-expansion is obviously advantageous as it

easily allows us to derive the nonrelativistic limit for $c \to \infty$, which should match the Schrödinger Hamiltonian. This holds for the Foldy–Wouthuysen $1/c$-expansion. Already from the free-particle Foldy–Wouthuysen transformation, the one-electron Pauli Hamiltonian emerges at second order in $1/c$, which provides the lowest-order one-electron mass velocity, Darwin, and spin–orbit corrections to the Schrödinger Hamiltonian. If the free-particle transformation is applied to the four-component many-electron Hamiltonian that includes Coulomb and Breit interactions of the electrons, the Breit–Pauli Hamiltonian will result as zeroth- to second-order terms [14–16].

The Pauli Hamiltonian is known to be useful in a perturbation theory context but produces difficulties when applied in a variational approach. We have argued [12] that all $1/c$ expansions – also the one produced by a sequence of unitary transformations as proposed by Foldy and Wouthuysen – will fail in a variational context as the true expansion parameter is actually the momentum divided by mc, which should be smaller than one for a Taylor expansion of the relativistic energy–momentum relation to converge. This, however, cannot be guaranteed as can be understood in terms of formal and physical reasons. On the one hand, high-momentum eigenfunctions cannot be excluded from a complete-basis-set representation of the $1/c$-expanded Dirac Hamiltonian, and so $p/(mc) < 1$ cannot be guaranteed for all such basis functions. On the other hand, an electron may acquire high momentum in the close vicinity of heavy nuclei that may produce a case in which $p/(mc) > 1$.

Douglas–Kroll Transformations

An alternative is the expansion of the Hamiltonian in terms of the potential v as a formal expansion parameter. It often goes without saying explicitly that $v_{\mathscr{E}}$ and not the full v is chosen as an expansion parameter. This has dramatic consequences as $v_{\mathscr{E}}$ is the even part of v, and its odd complement, $v_{\mathscr{O}}$, containing vector-potential and exchange contributions is not considered in the transformation procedure. The inclusion of vector potentials amounts to additional difficulties, which are not discussed in this chapter. Instead, we may refer the reader to Ref. [13] and references cited therein for a detailed discussion of $v_{\mathscr{O}}$ in the context of transformation techniques. Moreover, by default $v_{\mathscr{E}}$ contains only the external electrostatic potential of the nuclei, because the block-diagonal contribution from the electron–electron interaction is not easy to evaluate (it depends on the ansatz for the wavefunction approximation and requires an iterative, self-consistent determination). Hence, a well-defined standard choice is $v_{\mathscr{E}} = -\sum_A Z_A/r_A$ as the sole contribution (Z_A is the nuclear charge number of nucleus A, and r_A is the length of the difference vector of its positions to that of an electron, all in Hartree atomic units).

Expansions in terms of the inverse speed of light and of the potential have been considered by Erikson and coworkers at around 1960 [17–20], which apparently have never been recognized in quantum chemistry. In 1974, an expansion of the Dirac Hamiltonian in terms of the external electrostatic potential was proposed by

Douglas and Kroll in the appendix of their paper [21]. In the mid-1980s, Hess discovered this appendix [22–24] and combined it with a smart computational protocol to evaluate the momentum-space expressions of the electrons-only Hamiltonian in a basis of position-space one-electron functions (such as Gaussian functions used in almost all molecular quantum chemistry computer programs).

The essential insight by Hess [22] was that for practical applications eigenvalues of the squared momentum operator are required, which are known in a position-space basis that diagonalizes the matrix representation of the p^2-operator. Since the nonrelativistic kinetic energy operator $-\hbar^2\Delta/(2m)$ contains the squared momentum operator, $p^2 = -\hbar^2\Delta$, an operator in a position-space basis can be transformed into one in the p^2-basis by transforming it with the eigenvectors of the kinetic energy matrix scaled by $2m$. As a fortunate consequence, an explicit momentum-space representation of all operators is not necessary (and would have been a significant obstacle for an implementation of the Douglas–Kroll–Hess approach in a standard quantum chemistry program designed for molecular applications). It was explicitly shown by Liu and coworkers that exploiting eigenvectors of the p^2-operator corresponds to the choice of a kinetically balanced basis set and that the matrix representation of Hess' (DKH) Hamiltonian can thus be derived from the matrix form of the Dirac Hamiltonian in such a basis set [25, 26].

In his original work on the DKH Hamiltonian [23], Hess considered all terms in the transformed Hamiltonian up to second order in the external electrostatic potential, which defines the second-order DKH Hamiltonian h_+^{DKH2}. This derivation had to be slightly corrected in Ref. [24]. It was not before the year 2000 that the third-order Hamiltonian was derived and applied in quantum chemical calculations [27], followed by the correct fourth- and fifth-order DKH Hamiltonians [28] and the sixth-order one [29]. Note that the fourth- and fifth-order Hamiltonians in Ref. [27] turned out to be not correct [28]. Moreover, Ref. [27] presents results only for the third-order Hamiltonian, which, however, do not show the correct (oscillatory) convergence behavior (see below). An arbitrary-order and therefore exact numerical decoupling approach in terms of DKH Hamiltonians was then considered by us in 2004 [12].

While analytic results on the boundedness of the second-order DKH Hamiltonian could be obtained [30–32], only the first implementation of the arbitrary-order DKH approach [33] demonstrated the order-by-order convergence, and variational stability could be (numerically) investigated for high orders. The order-by-order convergence can be understood in terms of the true rather than the formal expansion parameter [12]: that is the potential (expressed in terms of matrix elements in the given p^2-basis) divided by huge energy denominators. However, we found an oscillatory convergence behavior [33]: odd DKH orders yield energy eigenvalues that are below the Dirac reference energy, while even DKH orders approach the Dirac reference from above. This behavior can be understood in view of the sign of the leading term in the truncation error [12, 33].

For the sequential order-by-order decoupling of the Dirac Hamiltonian in DKH theory, the necessary first step [12] is a free-particle Foldy–Wouthuysen transformation U_0 to generate an odd operator $\mathcal{O}_1$ that is linear in the potential,

$$h_1 = U_0 h_D U_0^\dagger = \mathscr{E}_0 + \mathscr{E}_1 + \mathscr{O}_1, \tag{9}$$

besides two four-dimensional even operators, $\mathscr{E}_0$ and $\mathscr{E}_1$ (the subscript denotes the order in the potential $v_{\mathscr{E}}$), which remain unchanged under all subsequent transformations. They define the first-order DKH Hamiltonian:

$$h_+^{\mathrm{DKH1}} = \mathscr{E}_0 + \mathscr{E}_1. \tag{10}$$

For explicit expressions of the low-order even terms, see Ref. [28].

The subsequent transformations (in principle, infinitely many of them) are chosen to eliminate the lowest-order odd term at a given step. Hence, U_1 is chosen such that $\mathscr{O}_1$ is eliminated, while new odd terms of higher order emerge. Then, U_2 eliminates $\mathscr{O}_2$ and so forth. Fortunately, each of these unitary transformations produces two even orders that remain unchanged by the higher-order unitary transformations, i.e., U_1 produces the final expression for $\mathscr{E}_2$ and $\mathscr{E}_3$, while U_2 produces $\mathscr{E}_4$ and $\mathscr{E}_5$ and so on. This has been called the $(2n+1)$-rule for producing the $(2n+1)$th-order DKH Hamiltonian from $U = U_n U_{(n-1)} \ldots U_0$.

The order-by-order elimination of odd operators in the Hamiltonian is achieved by choosing the parameter W that parametrizes the unitary transformation at a given step in such a way that the lowest-order odd operator of that step, to which W contributes, cancels. Many closed-form expressions for the parametrization of the unitary transformation are available. For example, Douglas and Kroll [21] proposed the so-called square-root parametrization,

$$U_i^{\mathrm{SQR}} = \sqrt{1 + W_i^2} + W_i, \tag{11}$$

while Nakajima and Hirao [27] employed the exponential parametrization,

$$U_i^{\mathrm{EXP}} = \exp(W_i), \tag{12}$$

which is known best in quantum chemistry.

When the exponential parametrization was applied in the derivation of the low-order DKH Hamiltonians by Nakajima and Hirao [27], it was not clear whether these low-order Hamiltonians are actually independent of the parametrization chosen. Since the analytic parametrizations are expanded in a Taylor series expansion in powers of the anti-hermitian parameter W_i, we set out [28] to study the most general unitary transformation, in which unitarity is imposed on the coefficients of a general power series expansion in terms of W_i. This ansatz covers all possible, and thus infinitely many, parametrizations of the unitary transformation. We found [28] that only up to fourth order is the DKH Hamiltonian independent of the chosen parameterization. Starting at the fifth order, DKH Hamiltonians depend on the expansion coefficients of the unitary transformation, an unfortunate effect that vanishes only at infinite order. However, the parameter dependence of the

fifth- and all higher-order DKH Hamiltonians is, for reasonable parametrizations of the unitary transformation, much smaller than the amplitude of the oscillatory convergence with increasing DKH orders [34].

It is clear that, at infinite order, any unitary transformation will exactly reproduce the spectrum of the Dirac Hamiltonian. However, this does not hold for the eigenstates. Different unitary transformations produce different DKH wavefunctions and different DKH orbitals at some given order (and also at infinite order). Only expectation values in the four-component theory are preserved by unitary transformations of the wavefunction and the property operator. The according transformation of the property operator has occasionally been omitted as the error introduced – the so-called picture-change error [35] – is small for valence-shell properties. However, it can be significant and therefore non-negligible for properties probed closed to an atomic nucleus [36]. A most prominent example, in which the picture-change error is dramatic, is the contact electron density [37], which is central to calculating the Mössbauer isomer shift [38, 39].

For properties, the $(2n + 1)$-rule does not hold and n unitary transformations are required to produce an nth-order DKH property operator [40]. A symbolic scheme for the automated derivation of arbitrary-order DKH property operators has been presented [36].

Note that the DKH expansion is *not* of the type that yields the Schrödinger Hamiltonian to lowest order (that is only achieved by considering the limiting case of $c \to \infty$). Accordingly, one does not obtain "relativistic corrections" in the DKH expansion as there is no nonrelativistic (zeroth-order) reference.

As a final remark, we should emphasize that the derivation of any DKH Hamiltonian produces a four-dimensional operator, i.e., one that contains an approximation to h_+ as well as to h_- on the block diagonal. The approximation for h_+ is then obtained by replacing the Dirac parameter matrix $\beta = \mathrm{diag}(1, 1, -1, -1)$ in all terms of the DKH expansion by the two-dimensional unit matrix.

Implementation of Douglas–Kroll Transformations

While Hess and others derived the lowest- and low-order DKH Hamiltonians manually, it became apparent that this stepwise decoupling protocol can be fully automated [12, 33], also for molecular properties [36, 40]. The derivation of an arbitrary-order DKH Hamiltonian or property operator was accomplished fully symbolically in two steps [33, 36]. First, all terms contributing to a given order were derived on a rather abstract formal level in terms of even and odd operators of a well-defined order in the external potential. Then, the resolution of the identity, $(\boldsymbol{\sigma} \cdot \mathbf{p})(\boldsymbol{\sigma} \cdot \mathbf{p})/p^2$, is used to break down all expressions into matrix products of known nonrelativistic operator matrices plus two additional types of "relativistic" matrices, namely, those of the operators $\mathbf{p} \cdot v\mathbf{p}$ and $\mathbf{p} \cdot O\mathbf{p}$ (where the omission of the vector $\boldsymbol{\sigma}$ of Pauli spin matrices in front of each momentum operator indicates the

standard spin-free one-component approximation to the DKH approach) required for the decoupled one-electron Hamiltonian and property operator O, respectively.

Unfortunately, this two-step protocol produces operator expressions of increasing length (measured by the number of matrix multiplications) with increasing order. Because of this steep scaling, DKH operators up to fourteenth order were considered in the early years [33, 36]. Peng and Hirao realized that the cost of the whole derivation can be significantly reduced by avoiding the second step [41] so that DKH calculations up to 35th order were easily possible [42,43]. We should note, however, that already the low-order DKH Hamiltonians, and in particular the original DKH2 one, provides an accurate description of valence-shell properties (see Refs. [44, 45] for two examples). Only properties probed close to an atomic nucleus [34, 36, 46] such as contact densities [37] or core excitations in X-ray and UV spectroscopies require high orders.

Amazingly, the stepwise derivation of Foldy–Wouthuysen decoupling in powers of $1/c$ had already been automated in 1968 on a Telefunken TR4 computer in ALGOL60 by deVries and Jonker [47, 48]. The even and odd decomposition of the Hamiltonian was achieved by mapping this digital structure to a binary number. In this way, the Hamiltonian could be derived "semi-symbolically" up to 8th order in $p/(mc)$ in Ref. [47] and to 10th order in $p/(mc)$ in Ref. [48].

The DKH approach is best known in its scalar-relativistic variant, in which all spin-dependent terms are separated from the scalar ones (by application of Dirac's relation) and then omitted. Clearly, omitting all Pauli spin matrices from the DKH Hamiltonian eliminates the spin–orbit coupling and a spin-averaged description emerges. The resulting scalar DKH Hamiltonian still comprises all kinematic relativistic effects (to arbitrary order in the potential). As its eigenfunctions are scalar functions, it can be easily interfaced with any nonrelativistic quantum chemistry computer program. The nonrelativistic one-electron Hamiltonian in the Fock operator is then replaced by a scalar-relativistic DKH Hamiltonian of predefined order. Various corrections to improve on the standard approximations in practical applications of DKH theory have been proposed [49–57].

Relation to Other Exact-Decoupling Approaches

A hybrid approach first proposed by Jensen [58] and then elaborated by Liu, Kutzelnigg, Saue, Visscher, Iliaš, and coworkers [59–68] offers the possibility to achieve decoupling by a single unitary transformation. It is now common to denote this approach the exact two-component approach, with acronym "X2C." The central idea is that one can construct an exact unitary transformation matrix $U = U(X)$ from the eigenvectors of the four-component Fock operator through their relation to the X-operator in matrix representation, $\mathbf{X}$,

$$C_S^{(+)} = \mathbf{X} C_L^{(+)} \quad \Rightarrow \quad \mathbf{X} = C_S^{(+)} \left(C_L^{(+)} \right)^{-1} \tag{13}$$

where $(C_L^{(+)}, C_S^{(+)})$ contain the positive-energy "(+)" eigenvectors of the four-component Fock matrix (they are the expansion coefficients of the basis-set expansions for the large "L" and small "S" two-spinors of the four-component molecular orbital). Clearly, this implies that the Fock operator needs to be diagonalized first (in a given one-electron basis set), and so the problem seems to have already been solved then (it also implies that the solution of a four-component problem is actually feasible). However, if this calculation is done only for an approximate Fock operator, in which the electron–electron interaction terms are neglected or approximated, then an efficient approximation to the basis-set representation of the exact unitary transformation can be obtained, which produces an approximate two-component Fock operator to which missing potential energy terms (most importantly, the full electron–electron interaction) are added. This procedure will produce a picture-change error for all interaction terms that were not considered in the construction of the unitary transformation. However, the resulting X2C Hamiltonian reproduces the original spectrum of the (full) four-component (reference) Fock Hamiltonian well so that it can be employed in a two-component electrons-only theory.

The abovementioned IOTC method of Barysz and Sadlej [10] is actually an extended X2C approach that involves one additional unitary transformation, namely, the free-particle Foldy–Wouthuysen transformation. Although the latter is the essential first ingredient of DKH theory, it is not mandatory in an X2C-type approach and therefore only increases the computational cost. In order to avoid confusion due to the rather general acronym "IOTC," the two-step exact-decoupling approach by Barysz and Sadlej has often been called the "BSS" approach according to the initials of the authors of an earlier paper [69] to which it is related.

It is important to understand that the computational effort for two-component decoupling approximations scales with a measure of the size of the system under consideration, e.g., with the number of basis functions. Accordingly, an efficient systematic approach to the decoupling transformation considers its atomic composition, which can be particularly easily achieved in the case of atom-centered basis functions, to set up local decoupling approximations. An atomic decomposition of the DKH unitary transformation has been proposed by us [70] and by Seino and Nakatamo [71, 72], who also developed a geometry gradient for structure optimizations [73]. An alternative is the atomic decomposition of the Hamiltonian [74–77], which, however, produces no general recipe for the transformation of off-diagonal atom–other-atom blocks in the Hamiltonian. Our derivation of the diagonal local approximation to the unitary transformation (DLU) [70] was sufficiently general to also comprise the local X2C and BSS approaches. Hence, the computational effort is reduced by a reduction of dimension of matrices subjected to multiplication, inversion, and diagonalization operations in DKH, BSS, and X2C, rather than by a limitation of the decoupling accuracy by truncation of a series expansion as in DKH2.

We have presented a highly efficient implementation of the local exact-decoupling methods in the TURBOMOLE program package along with a detailed

analysis of the performance of the different approaches [43]. Moreover, we refer the reader to a general overview [42], which provides a detailed numerical comparison of different exact-decoupling approaches.

Conclusions and Outlook

The present situation in relativistic quantum chemistry is such that all potential pitfalls associated with a four-component many-electron theory based on the Dirac one-electron Hamiltonian can be circumvented by an appropriate expansion of the molecular orbitals (spinors) into a finite one-electron basis set in such a way that all properties of the underlying Hamiltonian are respected by this expansion (kinetic balance). Even the dimension of the matrix representation of a four-component Fock operator can be limited to be about twice as large as the one for a corresponding nonrelativistic Schrödinger-based Fock Hamiltonian. Consequently, computational difficulties can no longer be held account for the development of two-component methods. In fact, four-component relativistic calculations have become routine, but since not many groups are working in this field, we can be grateful to the major effort of the DIRAC development team [78] that an open-source, freely available, highly professional, multipurpose, general, four-component molecular electronic-structure program is available.

As a consequence, we finally need to address the question whether two-component approaches are still of value in computational quantum chemistry or whether they will be eventually replaced by four-component methods valid for the whole periodic table of the elements. The discussion of this question has a long history [79] and we shall only touch upon it from the point of view of practical molecular electronic-structure calculations.

An important case can be made for sophisticated relativistic electron-correlation methods such as (four-component) multi-configurational self-consistent-field (MCSCF) [80, 81], (four-component) coupled cluster (CC) [82–84], and (four-component) density matrix renormalization group (DMRG) [85]. They require a four-index transformation which switches from one- and two-electron integrals given in the atomic-orbital basis to those in the molecular-orbital basis, in which the second-quantized Hamiltonian, that is the basis of all ab initio electron-correlation methods, is formulated. The transformation scales with the fifth power of the dimension of the problem and is thus particularly cumbersome for four-component methods due to the additional basis set for the small components of the molecular spinors. The requirements in terms of computational resources can be so large that this step may prevent one from carrying out a four-component electron-correlation calculation with an MCSCF, CC, or DMRG approach. Hence, for this step, a two-component method is beneficial as it basically requires the same effort in the four-index transformation step as a nonrelativistic approach would require.

Clearly, the X2C approach, especially in its local version [70], is most efficient for this purpose and so the next question is whether approximate and sequential

decoupling approaches will continue to have a right to exist in computational chemistry. Clearly, DKH2 will be around for some time as many developments have been based on this low-order Hamiltonian (and its accuracy for valence-shell properties such as vibrational frequencies, reaction energies, or bond lengths is undeniable). The most important technical advantage of DKH2 is the supply of basis sets for the whole periodic table of the elements that were produced by many groups in the past two decades.

Moreover, approximate two-component methods such as DKH2 require less computational effort than X2C, and so they may be beneficial for extensive calculation, if the calculation of the one-electron Hamiltonian is the limiting step, which can, however, be circumvented by introducing the DLU approximation [70].

All analysis in this conceptual overview focused mostly on the Hamiltonian and thus on the energy as the target observable. For other observables or specific electronic-structure methods, it may still be advantageous to consider a sequential decoupling protocol [26].

Apart from these computational considerations, we should not forget that, at its heart, DKH theory is an analytic tool for deriving an electrons-only Hamiltonian in the no-pair approximation of first-quantized relativistic many-electron theory. As such it will persist as the unique decoupling protocol yielding variationally stable Hamiltonians.

Acknowledgements This work was financially supported by the Swiss National Science Foundation SNF and by ETH Zurich.

References

1. Reiher M, Wolf A, Hess BA (2006) Relativistic quantum chemistry: from quantum electrodynamics to quasi-relativistic methods. In: Rieth M, Schommers W (eds) Handbook of theoretical and computational nanotechnology. American Scientific Publishers, pp 401–444
2. Reiher M (2006) Douglas-Kroll-Hess theory: a relativistic electrons-only theory for chemistry. Theor Chem Acc 116:241
3. Reiher M (2012) Relativistic Douglas-Kroll-Hess theory. WIREs: Comput Mol Sci 2:139
4. Nakajima T, Hirao K (2012) The Douglas-Kroll-Hess approach. Chem Rev 112:385
5. Reiher M, Wolf A (2015) Relativistic quantum chemistry, 2nd edn. Wiley-VCH, Weinheim
6. Reiher M, Hinze J (2003) Four-component ab initio methods for electronic structure calculations of atoms, molecules, and solids. In: Hess BA (ed) Relativistic effects in heavy-element chemistry and physics. Wiley, Chichester, pp 61–88
7. Saue T, Visscher L (2003) Four-component electronic structure methods for molecules. In: Wilson S, Kaldor U (eds) Theoretical chemistry and physics of heavy and superheavy elements. Kluwer, Dordrecht, pp 211–267
8. Eliav E, Kaldor U (2010) Four-component electronic structure methods. In: Barysz M, Ishikawa Y (eds) Relativistic methods for chemists. Challenges and advances in computational chemistry and physics, vol 10. Springer, Dordrecht, pp 279–349
9. Heully JL, Lindgren I, Lindroth E, Lundquist S, Mårtensen-Pendrill AM (1986) Diagonalisation of the Dirac Hamiltonian as a basis for a relativistic many-body procedure. J Phys B: At Mol Phys 19:2799
10. Barysz M, Sadlej AJ (2002) Infinite-order two-component theory for relativistic quantum chemistry. J Chem Phys 116:2696

11. Foldy LL, Wouthuysen SA (1950) On the Dirac theory of spin 1/2 particles and its non-relativistic limit. Phys Rev 78(1):29
12. Reiher M, Wolf A (2004) Exact decoupling of the Dirac Hamiltonian. I. General theory. J Chem Phys 121(5):2037
13. Luber S, Malkin Ondik I, Reiher M (2009) Electromagnetic fields in relativistic one-particle equations. Chem Phys 356:205
14. Chraplyvy ZV (1953) Reduction of relativistic two-particle wave equations to approximate forms. Phys Rev 91(2):388
15. Chraplyvy ZV (1953) Reduction of relativistic two-particle wave equations to approximate forms. II. Phys Rev 92(5):1310
16. Barker WA, Glover FN (1955) Reduction of relativistic two-particle wave equations to approximate forms. III. Phys Rev 99(1):317
17. Eriksen E (1958) Foldy-Wouthuysen transformation – exact solution with generalization to the 2-particle problem. Phys Rev 111:1011
18. Eriksen E (1958) Foldy-Wouthuysen transformation in closed form for spin 1/2 particle in time-independent external fields. Kongelige Norske Videnskabers Selskabs Forhandlinger 31:39
19. Eriksen E, Kolsrud M (1960) Canonical transformations of Dirac's equation to even forms. Expansion in terms of the external fields. Suppl Nuovo Cimento 18:1
20. Eriksen E (1961) Transformations of relativistic 2-particle equations. Nuovo Cimento 20:747
21. Douglas M, Kroll NM (1974) Quantum electrodynamical corrections to the fine structure of helium. Ann Phys 82:89
22. Hess BA (1985) Applicability of the no-pair equation with free-particle projection operators to atomic and molecular structure calculations. Phys Rev A 32:756
23. Hess BA (1986) Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. Phys Rev A 33:3742
24. Jansen G, Hess BA (1989) Revision of the Douglas-Kroll transformation. Phys Rev A 39(11):6016
25. Liu W (2010) Ideas of relativistic quantum chemistry. Mol Phys 108:1679
26. Li Z, Xiao Y, Liu W (2012) On the spin separation of algebraic two-component relativistic Hamiltonians. J Chem Phys 137:154114
27. Nakajima T, Hirao K (2000) The higher-order Douglas-Kroll transformation. J Chem Phys 113:7786
28. Wolf A, Reiher M, Hess BA (2002) The generalized Douglas-Kroll transformation. J Chem Phys 117:9215
29. van Wüllen C (2004) Relation between different variants of the generalized Douglas-Kroll transformation through sixth order. J Chem Phys 120:7307
30. Brummelhuis R, Siedentop H, Stockmeyer E (2002) The ground state energy of relativistic one-electron atoms according to Hess and Jansen. Documenta Mathematica 7:167
31. Siedentop H, Stockmeyer E (2005) An analytic Douglas-Kroll-Heß method. Phys Lett A 341:473
32. Siedentop H, Stockmeyer E (2006) The Douglas-Kroll-Heß method: convergence and block-diagonalization of Dirac operators. Ann Henri Poincaré 7:45
33. Reiher M, Wolf A (2004) Exact decoupling of the Dirac Hamiltonian. II. The generalized Douglas-Kroll-Hess transformation up to arbitrary order. J Chem Phys 121:10945
34. Reiher M, Wolf A (2007) Regular no-pair Dirac operators: numerical study of the convergence of high-order Douglas-Kroll-Hess transformations. Phys Lett A 360:603
35. Baerends EJ, Schwarz WHE, Schwerdtfeger P, Snijders JG (1990) Relativistic atomic orbital contractions and expansions: magnitudes and explanations. J Phys B: At Mol Phys 23:3225
36. Wolf A, Reiher M (2006) Exact decoupling of the Dirac Hamiltonian. IV. Automated evaluation of molecular properties within the Douglas-Kroll-Hess theory up to arbitrary order. J Chem Phys 124:064103
37. Mastalerz R, Lindh R, Reiher M (2008) Douglas-Kroll-Hess electron density at an atomic nucleus. Chem Phys Lett 465:157

38. Barone G, Mastalerz R, Lindh R, Reiher M (2008) Nuclear quadrupole moment of ^{119}Sn. J Phys Chem A 112:1666
39. Knecht S, Fux S, van Meer R, Visscher L, Reiher M, Saue T (2011) Mössbauer spectroscopy for heavy elements: a relativistic benchmark study of mercury. Theor Chem Acc 129:631
40. Wolf A, Reiher M (2006) Exact decoupling of the Dirac Hamiltonian. III. Molecular properties. J Chem Phys 124:064102
41. Peng D, Hirao K (2009) An arbitrary order Douglas-Kroll method with polynomial cost. J Chem Phys 130:044102
42. Peng D, Reiher M (2012) Exact decoupling of the relativistic Fock operator. Theor Chem Acc 131:1081
43. Peng D, Middendorf N, Weigend F, Reiher M (2013) An efficient implementation of two-component relativistic exact-decoupling methods for large molecules. J Chem Phys 138:184105
44. Hess BA, Kaldor U (2000) Relativistic all-electron coupled-cluster calculations on Au_2 in the framework of the Douglas-Kroll transformation. J Chem Phys 112(4):1809
45. Wolf A, Reiher M, Hess BA (2004) Correlated ab initio calculations of spectroscopic parameters of SnO within the framework of the higher-order generalized Douglas-Kroll transformation. J Chem Phys 120:8624
46. Mastalerz R, Barone G, Lindh R, Reiher M (2007) Analytic high-order Douglas-Kroll-Hess electric-field gradients. J Chem Phys 127:074105
47. de Vries E, Jonker JE (1968) Non-relativistic approximations to the Dirac Hamiltonian. Nucl Phys B 6:213
48. Jonker JE (1968) Non-relativistic approximations of the Dirac Hamiltonian II. Report IR 80 – Institute for theoretical physics. University of Groningen, The Netherlands
49. Samzow R, Hess BA (1991) Spin-orbit effects in the Br atom in the framework of the no-pair theory. Chem Phys Lett 184:491
50. Samzow R, Hess BA, Jansen G (1992) The two-electron terms of the no-pair Hamiltonian. J Chem Phys 96(2):1227
51. Park C, Almlöf JE (1994) Two-electron relativistic effects in molecules. Chem Phys Lett 231:269
52. Heß BA, Marian CM, Wahlgren U, Gropen O (1996) A mean-field spin-orbit method applicable to correlated wavefunctions. Chem Phys Lett 251:365
53. Schimmelpfennig B, Maron L, Wahlgren U, Teichteil C, Fagerli H, Gropen O (1998) On the combination of ECP-based Cl calculations with all-electron spin-orbit mean-field integrals. Chem Phys Lett 286(3–4):267
54. Schimmelpfennig B, Maron L, Wahlgren U, Teichteil C, Fagerli H, Gropen O (1998) On the efficiency of an effective Hamiltonian in spin-orbit CI calculations. Chem Phys Lett 286(3–4):261
55. Boettger JC (2000) Approximate two-electron spin-orbit coupling term for density-functional-theory DFT calculations using the Douglas-Kroll-Hess transformation. Phys Rev B 62(12):7809
56. van Wüllen C, Michauk C (2005) Accurate and efficient treatment of two-electron contributions in quasirelativistic high-order Douglas-Kroll density-functional calculations. J Chem Phys 123:204113
57. Autschbach J, Peng D, Reiher M (2012) Two-component relativistic calculations of electric-field gradients using exact decoupling methods: spin-orbit and picture-change effects. J Chem Theory Comput 8:4239
58. Jensen HJA (2005) 'Douglas–Kroll the Easy Way', talk at conference on relativistic effects in heavy elements – REHE 2005, Mülheim, Apr 2005
59. Kutzelnigg W, Liu W (2005) Quasirelativistic theory equivalent to fully relativistic theory. J Chem Phys 123:241102
60. Kutzelnigg W, Liu W (2006) Quasirelativistic theory I. Theory in terms of a quasirelativistic operator. Mol Phys 104(13–14):2225

61. Liu W, Peng D (2006) Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory. J Chem Phys 125:044102
62. Filatov M (2006) Quasirelativistic theory equivalent to fully relativistic theory. J Chem Phys 125:107101
63. Kutzelnigg W, Liu W (2006) Response to "Comment on 'Quasirelativistic theory equivalent to fully relativistic theory' ". J Chem Phys 125:107102
64. Iliaš M, Saue T (2007) An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. J Chem Phys 126:064102
65. Liu W, Kutzelnigg W (2007) Quasirelativistic theory. II. Theory at matrix level. J Chem Phys 126:114107
66. Peng D, Liu W, Xiao Y, Cheng L (2007) Making four-and two-component relativistic density functional methods fully equivalent based on the idea of "from atoms to molecule". J Chem Phys 127:104106
67. Sikkema J, Visscher L, Saue T, Ilias M (2009) The molecular mean-field approach for correlated relativistic calculations. J Chem Phys 131:124116
68. Liu W, Peng D (2009) Exact two-component Hamiltonians revisited. J Chem Phys 131:031104
69. Barysz M, Sadlej AJ, Snijders JG (1997) Nonsingular two/one-component relativistic Hamiltonians accurate through arbitrary high order in α^2. Int J Quantum Chem 65:225
70. Peng D, Reiher M (2012) Local relativistic exact decoupling. J Chem Phys 136:244108
71. Seino J, Nakai H (2012) Local unitary transformation method for large-scale two-component relativistic calculations: case for a one-electron Dirac Hamiltonian. J Chem Phys 136:244102
72. Seino J, Nakai H (2012) Local unitary transformation method for large-scale two-component relativistic calculations. II. Extension to two-electron Coulomb interaction. J Chem Phys 137:144101
73. Nakajima Y, Seino J, Nakai H (2013) Analytical energy gradients based on spin-free infinite-order Douglas-Kroll-Hess method with local unitary transformation. J Chem Phys 139:244107
74. Gagliardi L, Handy NC, Ioannou AG, Skylaris CK, Spencer S, Willetts A, Simper AM (1998) A two-centre implementation of the Douglas-Kroll transformation in relativistic calculations. Chem Phys Lett 283:187
75. Peralta JE, Scuseria GE (2004) Relativistic all-electron two-component self-consistent density functional calculations including one-electron scalar and spin-orbit effects. J Chem Phys 120(13):5875
76. Peralta JE, Uddin J, Scuseria GE (2005) Scalar relativistic all-electron density functional calculations on periodic systems. J Chem Phys 122:084108
77. Thar J, Kirchner B (2009) Relativistic all-electron molecular dynamics simulations. J Chem Phys 130:124103
78. DIRAC, a relativistic ab initio electronic-structure program, Release DIRAC14 (2014), written by Saue T, Visscher L, Jensen HJAa, Bast R. with contributions from Bakken V, Dyall KG, Dubillard S, Ekström U, Eliav E, Enevoldsen T, Faßhauer E, Fleig T, Fossgaard O, Gomes ASP, Helgaker T, Lærdahl JK, Lee YS, Henriksson J, Iliaš M, Jacob ChR, Knecht S, Komorovský S, Kullie O, Larsen CV, Nataraj HS, Norman P, Olejniczak G, Olsen J, Park YC, Pedersen JK, Pernpointner M, di Remigio R, Ruud K, Sałek P, Schimmelpfennig B, Sikkema J, Thorvaldsen AJ, Thyssen J, van Stralen J, Villaume S, Visser O, Winther T, Yamamoto S (see http://www.diracprogram.org)
79. Quiney HM, Skaane H, Grant IP (1998) Ab initio relativistic quantum chemistry: four-components good, two-components bad. Adv Quantum Chem 32:1
80. Jensen HJA, Dyall KG, Saue T, Fægri K Jr (1996) Relativistic four-component multiconfigurational self-consistent-field theory for molecules: formalism. J Chem Phys 104(11):4083
81. Abe M, Nakajima T, Hirao K (2006) The relativistic complete active-space second-order perturbation theory with the four-component Dirac Hamiltonian. J Chem Phys 125:234110
82. Eliav E, Kaldor U, Ishikawa Y (1994) Open-shell relativistic coupled-cluster method with Dirac-Fock-Breit wave functions: energies of the gold atom and its cation. Phys Rev A 49(3):1724

83. Eliav (Ilyabaev) E, Kaldor U, Ishikawa Y (1994) Relativistic coupled cluster method based on Dirac-Coulomb-Breit wavefunctions. Ground state energies of atoms with two to five electrons. Chem Phys Lett 222:82
84. Visscher L, Lee TJ, Dyall KG (1996) Formulation and implementation of a relativistic unrestricted coupled-cluster method including noniterative connected triples. J Chem Phys 105(19):8769
85. Knecht S, Legeza O, Reiher M (2014) Four-component density matrix renormalization group. J Chem Phys 140:041101

Spin Separation of Relativistic Hamiltonians

13

Zhendong Li and Wenjian Liu

Contents

Introduction . . . 412
The One-Electron h_{sf} and $h_{SO,1e}$ . . . 415
Spin Separation at Four-Component Level . . . 415
Spin Separation at Two-Component Level . . . 417
The Lowest-Order Spin-Orbit Terms . . . 423
A Short Summary . . . 429
Two-Electron Spin-Orbit Coupling . . . 432
Applications: Perturbative Treatment of SOC . . . 437
Ground-State Spin-Dependent Properties . . . 437
Excited-State Spin-Dependent Properties . . . 439
Nonadiabatic Dynamics Involving Different Spin States . . . 442
Summary . . . 445
References . . . 445

Abstract

Separating a no-pair relativistic Hamiltonian into spin-free and spin-dependent terms allows a clear distinction of scalar and spin relativistic effects, on one hand, and enables an efficient two-step treatment of such effects, on the other. That is, the spin-free Hamiltonian can be handled in the same way as the Schrödinger-Coulomb Hamiltonian, while the spin-dependent Hamiltonian can be handled either variationally or perturbatively. We here discuss the spin separation of various no-pair relativistic Hamiltonians, focusing especially on the algebraic exact two-component variants to which the Dirac identity for spin

Z. Li (✉) • W. Liu (✉)
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: zhendongli2008@gmail.com; liuwjbdf@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_4

separation cannot directly be applied. A mean-field approximation to the two-electron spin-orbit interaction is also introduced, which is sufficiently accurate for valence properties. The so-obtained operators are then combined with the analytic derivative technique, response theory, and quasi-degenerate perturbation theory for spin-free and spin-dependent properties of both ground and excited states. Nonadiabatic dynamics involving different spin states is also discussed briefly.

Keywords

Exact two-component • Spin separation • Mean-field spin-orbit Hamiltonian • Spindependent property • Nonadiabatic dynamics

Introduction

No-pair all-electron or valence-only relativistic Hamiltonians can all be written in the following generic form:

$$h_{\mathrm{rel}} = h_{\mathrm{sf}} + h_{\mathrm{sd}}(\vec{\sigma}), \tag{1}$$

where h_{sf} includes all the terms that are independent of the Pauli matrix $\vec{\sigma}$, whereas h_{sd} includes all the terms that depend on $\vec{\sigma}$. The most important term in the latter is the so-called spin-orbit coupling (SOC) which is under concern here. From a computational point of view, there exist three different ways for treating the SOC:

1. One-step approach: The entire Hamiltonian h_{rel} is used to perform two- or four-component calculations, such that the SOC is accounted for already at the self-consistent field (SCF) level for orbital optimization.
2. Two-step approach: The spin-free Hamiltonian h_{sf} is first used to perform orbital optimization, and the spin-dependent Hamiltonian h_{sd} is then invoked in the post-SCF step, either variationally or perturbatively.
3. Neglecting the SOC. That is, only the spin-free Hamiltonian h_{sf} is considered.

The first approach is most accurate and is imperative for situations (e.g., core properties) where the SOC is very strong, while the third approach is least accurate and works only if the SOC is totally negligible for the target properties. In contrast, the second, two-step approach is applicable to most real-life chemical systems where the SOC is not negligible but not too strong. In this regard, there have been several choices for h_{sf} and h_{sd}:

1. NR+so-BP: The nonrelativistic Schrödinger-Coulomb (SC) Hamiltonian combined with the Breit-Pauli (BP) SOC operator
2. sf-ZORA+so-ZORA: The spin-free and spin-orbit terms separated from the zeroth-order regular approximation (ZORA) [1]

3. sf-DKH2+AMFI: The spin-free second-order Douglas-Kroll-Hess (DKH) Hamiltonian [2, 3] combined with a mean-field approximation [4] to the two-electron spin-orbit interaction through, e.g., the atomic mean-field integrals (AMFI) [5]
4. sf-MDEQ+so-MDEQ: The spin-free and spin-orbit terms derived from the modified Dirac equation (MDEQ) [6] with the two-electron Coulomb/Gaunt/Breit interaction

All such spin-separated Hamiltonians are derived a posteriori from their parent operators $h_{\rm rel}$ by virtue of the Dirac identity

$$(\vec{\sigma}\cdot\vec{A})B(\vec{\sigma}\cdot\vec{C}) = \vec{A}\cdot(B\vec{C}) + \mathrm{i}\vec{\sigma}\cdot[\vec{A}\times(B\vec{C})]. \tag{2}$$

Given the many choices, the question remains to be "What is the best relativistic Hamiltonian that can be applied in two steps?" To be the best, it should satisfy the following requirements:

(F1) It should describe scalar relativity to infinite order in α^2, the fine structure constant.
(F2) It should describe SOC at an accuracy that can be improved readily and systematically.
(F3) It can be used both variationally and perturbatively.
(F4) It can be readily interfaced with existing wave function/density functional methods for correlation and excitation.
(F5) It can be made linear scaling with respect to the molecular size.

That the first feature (F1) can be wanted is because an infinite order $h_{\rm sf}$ still shares the same machinery (single point group, spin symmetry, and real algebra) as the SC Hamiltonian. The second feature (F2), which is important for ensuring the accuracy, is more stringent and is not fulfilled by the spin separation of the regular approximation and DKH for they include too numerous terms at high orders. The third feature (F3) dictates that there should be no unbound operators in $h_{\rm sf}$ and $h_{\rm sd}$. The fourth feature (F4) is fulfilled by spin-free two-component Hamiltonians in the Schrödinger picture but not by the sf-MDEQ Hamiltonian still in the Dirac picture. The fifth feature (F5) is also very important for large systems. Overall, none of the abovementioned relativistic Hamiltonians satisfies all the criteria (F1)–(F5). The present account is to propose a relativistic Hamiltonian that satisfies all the criteria. It is required to take the following second-quantized form

$$H_{\rm el} = H_{\rm sf} + H_{\rm sd}, \tag{3}$$

$$H_{\rm sf} = \sum_{pq}[h_{\rm sf}]_{pq}a_p^\dagger a_q + \frac{1}{2}\sum_{pqrs}\langle pq|rs\rangle a_p^\dagger a_q^\dagger a_s a_r, \tag{4}$$

$$H_{\rm sd} = \sum_{pq}[h_{\rm sd}]_{pq}a_p^\dagger a_q, \quad [h_{\rm sd}]_{pq} = [h_{{\rm SO},1e}]_{pq} + [f_{{\rm SO},2e}]_{pq}, \tag{5}$$

which implies the following approximations:

(A1) Only the bare Coulomb interaction is retained in Eq. (4), which amounts to neglecting the two-electron scalar relativistic picture change corrections. This approximation is very accurate and renders the computation of two-electron interactions the same as the nonrelativistic case.
(A2) Only the two-electron SOC of $\mathcal{O}(\alpha^2)$ is retained in the spin-dependent two-electron interactions. If wanted, the BP form of the spin-spin interaction (which may be significant for light elements) can further be added.
(A3) A mean-field approximation is adopted for the two-electron SOC, so as to obtain an effective one-electron SOC $f_{\mathrm{SO},2e}$ in Eq. (5). This arises from the fact that the SOC is heavily dominated by single excitations. Yet, it should be kept in mind that a genuine two-electron SOC must be adopted when the SOC between two configurations differing by two spin orbitals becomes important.
(A4) The screening of $h_{\mathrm{SO},1e}$, furnished by $f_{\mathrm{SO},2e}$, arises mainly from the core electrons, as reflected by the fact that the SOC matrix elements are very insensitive to valence occupations used to create the mean field [4]. Therefore, it is justified to use a spin-averaged density matrix

$$\bar{\mathbf{P}} = \frac{1}{2}(\mathbf{P}^{\alpha} + \mathbf{P}^{\beta}) \tag{6}$$

and a one-center approximation to the two-electron spin-orbit integrals. The former simplifies the formulation of $f_{\mathrm{SO},2e}$ while the latter leads naturally to a linear scaling construction of $f_{\mathrm{SO},2e}$. The matrix elements $[f_{\mathrm{SO},2e}]_{pq}$ can even be confined only to heavy elements and their local surroundings, so as to achieve an $\mathcal{O}(1)$ scaling.

Even under these restrictions, there are still some degrees of freedom in choosing h_{sf} and h_{sd} in Eq. (3). The most accurate variant of h_{sf} can be derived from the sf-MDEQ [6], whose spectrum contains both positive (PES) and negative (NES) energy states. The corresponding $h_{\mathrm{SO},1e}$ is a simple spin-orbit operator of $\mathcal{O}(\alpha^2)$ (vide post). When combined with $f_{\mathrm{SO},2e}$ to form a single perturbation operator, the post-SCF treatment of SOC will be correct to $\mathcal{O}(\alpha^2)$ if the response of the scalar NES is neglected. To get higher-order SOC, the response of the scalar NES should be accounted for via coupled-perturbed equations. Therefore, such an *explicit* and *exact* treatment of NES will make the computation significantly different from the nonrelativistic counterpart involving only PES, although spatial and spin symmetries can be used in the spin-free calculation.

To achieve feature (F4) but meanwhile maintain features (F1) and (F2), we have to introduce further approximations to simply the h_{sf} and $h_{\mathrm{SO},1e}$ derived from the MDEQ. First, the explicit treatment of the scalar NES can be avoided by using the ideas of the quasi-four-component (Q4C) [7] or exact two-component (X2C) [8] approach. In Q4C, molecular orbitals are expanded in terms only of the atomic PES, so as to avoid molecular NES. That is, Q4C works with the untransformed four-component Hamiltonian but which is represented in a special basis. In contrast,

X2C works with an effective two-component Hamiltonian for the PES alone. For the one-step account of the SOC, Q4C and X2C are computationally identical in all the aspects of simplicity, accuracy, and efficiency [9]. However, for a two-step account of the SOC, the X2C formalism offers more flexibilities and is hence employed here to construct $h_{\rm sf}$ and $h_{\rm sd}$. Second, the explicit treatment of the response of the scalar NES can be avoided by invoking a suitable approximation to the decoupling matrix $\mathbf{X}$ (29) of X2C. It turns out that the following approximation for $\mathbf{X}$ is not only simple but also sufficiently accurate:

(A5) Both the spin-free $\mathbf{X}_0$ and the kth-order spin-orbit response $\mathbf{X}_k$ are approximated at the one-electron level.

The treatment of the SOC then requires only the response of the scalar PES, precisely the same as the nonrelativistic counterpart.

Having introduced the approximations (A1)–(A5), we are now ready to derive detailed expressions for $h_{\rm sf}$, $h_{{\rm SO},1e}$ and $f_{{\rm SO},2e}$. The spin separation of the Dirac operator/matrix is first briefly recapitulated in section "Spin Separation at Two–Component Level" to introduce the notations. After highlighting some general aspects of the spin separation at the two-component level in section "Spin Separation at Two-Component Level," the lowest-order terms of $h_{{\rm SO},1e}$ are constructed explicitly in three distinct ways in section "The Lowest-Order Spin-Orbit Terms." The results are further compared therein with the spin-orbit terms of some well-known A2C Hamiltonians. A "two-component Hamiltonian ladder" is depicted in section "A Short Summary." Section "Two-Electron Spin-Orbit Coupling" is devoted to the two-electron spin-orbit interaction. A two-electron SOC operator of $\mathcal{O}(\alpha^2)$ is first introduced, followed by a mean-field approximation. The construction of the desired Hamiltonian $H_{\rm el}$ (3) is then completed. Section "Applications: Perturbative Treatment of SOC" demonstrates how to combine the so-obtained spin-orbit operators with the analytic derivative technique, response theory, and quasi-degenerate perturbation theory (QDPT) for spin-free and spin-dependent properties of both ground and excited states. Nonadiabatic dynamics involving different spin states is also discussed therein. The account ends with concluding remarks in section "Summary." The atomic units ($\hbar = m = e = 1$) are always used.

The One-Electron $h_{\rm sf}$ and $h_{{\rm SO};1e}$

Spin Separation at Four-Component Level

The Dirac equation for an electron moving in an external potential V reads

$$h^D\psi = E\psi, \quad h^D = \begin{pmatrix} V & c\vec{\sigma}\cdot\vec{p} \\ c\vec{\sigma}\cdot\vec{p} & V - 2c^2 \end{pmatrix}, \quad \psi = \begin{pmatrix} \psi^L \\ \psi^S \end{pmatrix}, \tag{7}$$

where the rest mass energy c^2 of the electron has been subtracted. To have a meaningful spin separation, it should first be noted that the Dirac equation (7) has

to be transformed to a form in which the Pauli spin matrix $\vec{\sigma}$ appears in pairs. There can be infinitely many transformations to achieve this [10]. Yet, the simplest choice is the one [6] that leaves the large-component spinor ψ^L unchanged but replaces the small-component spinor ψ^S with a *pseudo-large* spinor ϕ^L:

$$\mathcal{T} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{\vec{\sigma}\cdot\vec{p}}{2c} \end{pmatrix}, \quad \psi^S = \frac{\vec{\sigma}\cdot\vec{p}}{2c}\phi^L. \tag{8}$$

Under this transformation, the Dirac equation (7) becomes

$$h^M \psi^M = EM\psi^M, \tag{9}$$

where

$$h^M = \mathcal{T}^\dagger h^D \mathcal{T} = \begin{pmatrix} V & T \\ T & \frac{\alpha^2}{4}W - T \end{pmatrix}, \tag{10}$$

$$M = \mathcal{T}^\dagger \mathcal{T} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{\alpha^2}{2}T \end{pmatrix}, \tag{11}$$

$$\psi^M = \mathcal{T}^{-1}\psi = \begin{pmatrix} \psi^L \\ \phi^L \end{pmatrix}, \tag{12}$$

$$T = \frac{p^2}{2}, \quad W = (\vec{\sigma}\cdot\vec{p})V(\vec{\sigma}\cdot\vec{p}). \tag{13}$$

Equation (9) has been termed "modified Dirac equation" (MDEQ) [6]. It is clear that the dependence of spin in h^M (10) appears only in the potential W(13) for the pseudo-large component. In other words, the real spin dependence in the original Dirac equation (7) is not in the kinetic energy per se but in the potential for the small component. The use of the Dirac identity (2) can then be made to obtain the spin-separated Hamiltonian in the form of Eq. (1),

$$h^M = h^M_{\rm sf} + h^M_{\rm sd}, \quad h^M_{\rm sf} = \begin{pmatrix} V & T \\ T & \frac{\alpha^2}{4}W_{\rm sf} - T \end{pmatrix}, \quad h^M_{\rm sd} = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\alpha^2}{4}W_{\rm sd} \end{pmatrix}, \tag{14}$$

where

$$W_{\rm sf} = \vec{p}\cdot V\vec{p}, \quad W_{\rm sd} = \mathbb{i}\vec{\sigma}\cdot(\vec{p}V\times\vec{p}) = \vec{\sigma}\cdot[(\nabla V)\times\vec{p}]. \tag{15}$$

It is seen that the spin-dependent term $h^M_{\rm sd}$ (14) is simply an operator of $\mathcal{O}(\alpha^2 W_{\rm sd})$. For the nuclear attraction $V = -Z/r$, $W_{\rm sd}$ is equal to $\frac{2Z}{r^3}\vec{s}\cdot\vec{l}$, thereby revealing a clear connection of $W_{\rm sd}$ with the SOC.

Since the large (ψ_p^L) and pseudo-large (ϕ_p^L) components share the same symmetry properties and nonrelativistic limit, they should be expanded [8] in the same basis $\{g_\mu\}$, viz.,

$$\psi_p^L = \sum_\mu g_\mu A_{\mu p}, \quad \phi_p^L = \sum_\mu g_\mu B_{\mu p}. \tag{16}$$

This leads to the following matrix representation of the MDEQ (9):

$$\mathbf{h}^D \mathbf{C} = \mathbf{M}\mathbf{C}E, \tag{17}$$

$$\mathbf{h}^D = \begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & \frac{\alpha^2}{4}\mathbf{W} - \mathbf{T} \end{pmatrix} = \mathbf{h}_{\mathrm{sf}}^D + \mathbf{h}_{\mathrm{sd}}^D, \tag{18}$$

$$\mathbf{h}_{\mathrm{sf}}^D = \begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & \frac{\alpha^2}{4}\mathbf{W}_{\mathrm{sf}} - \mathbf{T} \end{pmatrix}, \quad \mathbf{h}_{\mathrm{sd}}^D = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\alpha^2}{4}\mathbf{W}_{\mathrm{sd}} \end{pmatrix}, \tag{19}$$

$$\mathbf{M} = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \frac{\alpha^2}{2}\mathbf{T} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}, \tag{20}$$

where the corresponding matrices are defined as

$$\begin{aligned} S_{\mu\nu} &= \langle g_\mu | g_\nu \rangle, \quad T_{\mu\nu} = \langle g_\mu | \frac{1}{2} p^2 | g_\nu \rangle, \quad V_{\mu\nu} = \langle g_\mu | V | g_\nu \rangle, \\ (W_{\mathrm{sf}})_{\mu\nu} &= \langle g_\mu | \vec{p} \cdot V \vec{p} | g_\nu \rangle, \quad (W_{\mathrm{sd}})_{\mu\nu} = \langle g_\mu | \mathrm{i}\vec{\sigma} \cdot (\vec{p} V \times \vec{p}) | g_\nu \rangle. \end{aligned} \tag{21}$$

As a matter of fact, Eq. (17) is nothing but the matrix representation of the original Dirac equation (7) in a basis where the small-component basis functions $\{f_\mu\}$ are generated from the large-component ones $\{g_\mu\}$ via the restricted kinetic balance (RKB) prescription [11]

$$f_\mu = \frac{\vec{\sigma} \cdot \vec{p}}{2c} g_\mu. \tag{22}$$

Therefore, the term "modified Dirac equation" loses its meaning in the matrix representation. In the present context, it just means that the spin separation of the Dirac equation can be done at both the operator and matrix levels.

Spin Separation at Two-Component Level

Two-component (i.e., electron-only) relativistic Hamiltonians are usually obtained by block-diagonalizing the Dirac *operator* via unitary transformation [12] or equivalently elimination of the small component [13]. However, except for the free-particle case (i.e., $V = 0$ in Eq. (7)), the so-obtained two-component Hamiltonians

do not have a closed form and are overly complicated as compared with the aesthetically simple structure of the parent Dirac operator. Because of this, various approximations must be introduced, thereby leading to a series of approximate two-component (A2C) Hamiltonians, among which DKH2 [2,3] and ZORA [1] are most popular. Being of explicit operator forms, a posteriori spin separation of such A2C Hamiltonians can readily be made by using the Dirac identity (2). Their matrix representations can then be formed in the large-component basis $\{g_\mu\}$.

At variance with the unsurmountable difficulties in the decoupling of the *operator* Dirac equation (7) (which has an infinite number of solutions), the decoupling of the *matrix* Dirac equation (17) (which has a finite number of solutions) is not only very simple but also immune of singularities [14]. If one realizes that one only solves in practice a particular matrix representation of a quantum mechanical equation rather than the equation itself, one would bethink immediately that it is the matrix formulation of two-component relativistic Hamiltonians that is actually the natural choice. Again, this can be done in various ways. The one-step decoupling of the matrix Dirac equation (17) was initiated by Dyall [15] but was formulated properly and generally only by Kutzelnigg and Liu [16]. The resulting Hamiltonian was christened "exact two-component" (X2C) [17] when going to the Schrödinger picture [18]. At variance with the one-step decoupling, both two-step [19,20] and multiple-step [21,22] exact decoupling procedures have been proposed, with the resulting Hamiltonians coined Barysz-Sadlej-Snijders (BSS) and DKH, respectively. The three types of algebraic Hamiltonians share the same decoupling condition (42) and differ only in the renormalization (and hence picture change) [8]. For a more complete summary of the matrix formulation of two-component relativistic Hamiltonians, we refer the reader to Ref. [23].

In the context of spin separation, the price to pay for the algebraic two-component Hamiltonians is that the Dirac identity (2) cannot be applied due to the lack of an explicit expression for the transformation matrix. Yet, some a priori procedure can be invoked to circumvent such a conceptual difficulty. Suppose a two-component Hamiltonian $\mathbf{h}_+$ and its spin-free part $\mathbf{h}_{+,\mathrm{sf}}$ are to be derived from $\mathbf{h}^D$ (18) and $\mathbf{h}^D_{\mathrm{sf}}$ (19), respectively, the spin-dependent part $\mathbf{h}_{+,\mathrm{sd}}$ of $\mathbf{h}_+$ can then be defined as the difference between $\mathbf{h}_+$ and $\mathbf{h}_{+,\mathrm{sf}}$, viz.,

$$\mathbf{h}_{+,\mathrm{sd}} = \mathbf{h}_+ - \mathbf{h}_{+,\mathrm{sf}}. \tag{23}$$

Since both $\mathbf{h}_+$ and $\mathbf{h}_{+,\mathrm{sf}}$ can be obtained in various ways, the so-defined $\mathbf{h}_{+,\mathrm{sd}}$ is also not unique. Here we seek for an *explicit* $\mathbf{h}_{+,\mathrm{sd}}$ inspired by the DKH type of transformation. Recall that the DKH transformation starts with a free-particle Foldy-Wouthuysen (fpFW) transformation [12] to cover all kinematic relativistic effects and then proceeds with a series of unitary transformations U_i to eliminate at each step the lowest-order operators that are odd (off-diagonal) in V. For the present purpose, we take the spin-orbit operator W_{sd} instead of the external potential V as the expansion parameter. Therefore, the first transformation in place of the fpFW is chosen to be the spin-free unitary transformation $\mathbf{U}_0$ that covers all scalar

relativistic effects. The subsequent transformations are then introduced to eliminate at each step the lowest-order odd terms in $W_{\rm sd}$. The only difference lies in that the fpFW transformation [12] is known in an explicit and closed form, whereas the $\mathbf{U}_0$ transformation for the exact decoupling of $h_{\rm sf}^D$ is only known in an implicit and matrix form. However, such subtlety is actually not immaterial, as the DKH approach can be viewed as a kind of perturbation method based on a special partition of the matrix Hamiltonian $\mathbf{h}^D$ (18), viz.,

$$\begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & \frac{\alpha^2}{4}\mathbf{W} - \mathbf{T} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{T} \\ \mathbf{T} & -\mathbf{T} \end{pmatrix} + \begin{pmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & \frac{\alpha^2}{4}\mathbf{W} \end{pmatrix}. \tag{24}$$

Two better partitions can readily be envisaged:

$$\begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & \frac{\alpha^2}{4}\mathbf{W} - \mathbf{T} \end{pmatrix} = \begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & -\mathbf{T} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\alpha^2}{4}\mathbf{W} \end{pmatrix}, \tag{25}$$

$$\begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & \frac{\alpha^2}{4}\mathbf{W} - \mathbf{T} \end{pmatrix} = \begin{pmatrix} \mathbf{V} & \mathbf{T} \\ \mathbf{T} & \frac{\alpha^2}{4}\mathbf{W}_{\rm sf} - \mathbf{T} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{\alpha^2}{4}\mathbf{W}_{\rm sd} \end{pmatrix}. \tag{26}$$

Equation (25) takes the nonrelativistic Hamiltonian as zeroth order and is therefore in the spirit of direct perturbation theory (DPT) [24], while Eq. (26) takes the (two-component) spin-free relativistic Hamiltonian as zeroth order and hence leads to a more rapidly convergent expansion series. It is clear that Eq. (26) is the best partition if the nonrelativistic machinery is to be maintained.

Before discussing the perturbative expansion of $\mathbf{h}_{+,\rm sd}$, the (one-component) spin-free Hamiltonian $\mathbf{h}_{+,\rm sf}$ must be specified, since the definition of $\mathbf{h}_{+,\rm sd}$ depends on the form of $\mathbf{h}_{+,\rm sf}$, as seen in Eq. (23). The simplest choice [25] is the one-step spin-free X2C (sf-X2C) [16], whose construction can be summarized as follows. For generality, we extend Eq. (17) to a generic generalized eigenvalue problem

$$\mathbf{hC} = \mathbf{MC}E, \tag{27}$$

$$\mathbf{h} = \begin{pmatrix} \mathbf{h}_{11} & \mathbf{h}_{12} \\ \mathbf{h}_{21} & \mathbf{h}_{22} \end{pmatrix} = \mathbf{h}^\dagger, \quad \mathbf{M} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22} \end{pmatrix} = \mathbf{M}^\dagger, \quad \mathbf{C} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}. \tag{28}$$

To decouple the PES and NES, we first introduce the following formal relations:

$$\mathbf{B} = \mathbf{XA}, \quad \mathbf{A}_- = \tilde{\mathbf{X}}\mathbf{B}_- \tag{29}$$

between the small and large coefficients for the PES and NES, respectively. The following unitary transformation matrix $\mathbf{U}_X$ can then be introduced [8]:

$$\mathbf{U}_X = \mathbf{\Omega}_N \mathbf{\Omega}_D, \quad \mathbf{\Omega}_N = \begin{pmatrix} \mathbf{R}_+^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_-^\dagger \end{pmatrix}, \quad \mathbf{\Omega}_D = \begin{pmatrix} \mathbf{I} & \mathbf{X}^\dagger \\ \tilde{\mathbf{X}}^\dagger & \mathbf{I} \end{pmatrix}, \tag{30}$$

where [18]

$$\mathbf{R}_+ = (\mathbf{S}_{11}^{-1}\tilde{\mathbf{S}}_+)^{-\frac{1}{2}} = \mathbf{S}_{11}^{-\frac{1}{2}}(\mathbf{S}_{11}^{-\frac{1}{2}}\tilde{\mathbf{S}}_+\mathbf{S}_{11}^{-\frac{1}{2}})^{-\frac{1}{2}}\mathbf{S}_{11}^{\frac{1}{2}}, \tag{31}$$

$$\mathbf{R}_- = (\mathbf{S}_{22}^{-1}\tilde{\mathbf{S}}_-)^{-\frac{1}{2}} = \mathbf{S}_{22}^{-\frac{1}{2}}(\mathbf{S}_{22}^{-\frac{1}{2}}\tilde{\mathbf{S}}_-\mathbf{S}_{22}^{-\frac{1}{2}})^{-\frac{1}{2}}\mathbf{S}_{22}^{\frac{1}{2}}, \tag{32}$$

$$\tilde{\mathbf{S}}_+ = \mathbf{S}_{11} + \mathbf{X}^\dagger\mathbf{S}_{22}\mathbf{X}, \tag{33}$$

$$\tilde{\mathbf{S}}_- = \mathbf{S}_{22} + \tilde{\mathbf{X}}^\dagger\mathbf{S}_{11}\tilde{\mathbf{X}}, \tag{34}$$

$$\tilde{\mathbf{X}} = -\mathbf{S}_{11}^{-1}\mathbf{X}^\dagger\mathbf{S}_{22}. \tag{35}$$

Note that $\mathbf{U}_X$ is related [25] to the matrix representation of the unitary operator U in a RKB basis, and the matrix identity corresponding to $UU^\dagger = U^\dagger U = 1$ reads

$$\mathbf{U}_X\mathbf{M}\mathbf{U}_X^\dagger = \mathbf{M} \text{ or } \mathbf{U}_X^\dagger\mathbf{M}^{-1}\mathbf{U}_X = \mathbf{M}^{-1}. \tag{36}$$

The renormalization matrix $\mathbf{R}_+$ (31) for PES satisfies both the normalization condition $\mathbf{R}_+^\dagger\tilde{\mathbf{S}}_+\mathbf{R}_+ = \mathbf{S}$ and the Hermiticity condition $\mathbf{R}_+^\dagger\mathbf{S} = (\mathbf{S}\mathbf{R}_+)^\dagger = \mathbf{S}\mathbf{R}_+$, which can be merged into a single equation for $\mathbf{R}_+$,

$$\mathbf{R}_+\mathbf{S}_{11}^{-1}\tilde{\mathbf{S}}_+\mathbf{R}_+ = \mathbf{I}. \tag{37}$$

In practice, $\mathbf{R}_+$ can simply be calculated as follows:

$$\tilde{\mathbf{S}}_+\mathbf{Z} = \mathbf{S}_{11}\mathbf{Z}\Lambda, \quad \mathbf{R}_+ = \mathbf{Z}\Lambda^{-\frac{1}{2}}\mathbf{Z}^{-1}, \quad \mathbf{Z}^{-1} = \mathbf{Z}^\dagger\mathbf{S}_{11}. \tag{38}$$

Similar to Eq. (37), the renormalization matrix $\mathbf{R}_-$ for NES satisfies

$$\mathbf{R}_-\mathbf{S}_{22}^{-1}\tilde{\mathbf{S}}_-\mathbf{R}_- = \mathbf{I}. \tag{39}$$

The as yet unknown matrix $\mathbf{X}$ in $\mathbf{U}_X$ (30) is to be determined by the requirement that the transformation $\mathbf{U}_X$ can block-diagonalize the matrix Dirac equation (27) in the following way:

$$(\mathbf{U}_X\mathbf{h}^D\mathbf{U}_X^\dagger)\mathbf{C}_X = \begin{pmatrix} \mathbf{h}_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_- \end{pmatrix}\mathbf{C}_X = \mathbf{M}\mathbf{C}_X E, \tag{40}$$

with

$$\mathbf{C}_X = (\mathbf{U}_X^\dagger)^{-1}\mathbf{C} = \mathbf{M}^{-1}\mathbf{U}_X\mathbf{M}\mathbf{C} = \begin{pmatrix} \mathbf{C}_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_- \end{pmatrix}. \tag{41}$$

The vanishing off-diagonal block of $\mathbf{U}_X\mathbf{h}^D\mathbf{U}_X^\dagger$ leads to the following equation for $\mathbf{X}$:

$$\mathbf{h}_{21} + \mathbf{h}_{22}\mathbf{X} = \mathbf{S}_{22}\mathbf{X}\mathbf{S}_{11}^{-1}(\mathbf{h}_{11} + \mathbf{h}_{12}\mathbf{X}). \tag{42}$$

Instead of solving this quadratic equation iteratively [26], the $\mathbf{X}$ matrix can actually be obtained simply by the definition (29) using the eigenvectors of Eq. (27),

$$\mathbf{X} = \mathbf{B}\mathbf{A}^{-1} = (\mathbf{B}\mathbf{A}^{\dagger})(\mathbf{A}\mathbf{A}^{\dagger})^{-1}. \tag{43}$$

The upper-left block of Eq. (40) defines the equation for the PES:

$$\mathbf{h}_{+}\mathbf{C}_{+} = \mathbf{S}_{11}\mathbf{C}_{+}E_{+}, \tag{44}$$

$$\mathbf{h}_{+} = \mathbf{R}_{+}^{\dagger}\mathbf{L}_{+}^{\mathrm{X}}\mathbf{R}_{+}, \quad \mathrm{X} = \mathrm{UESC}, \mathrm{SESC}, \mathrm{NESC}, \tag{45}$$

$$\mathbf{L}_{+}^{\mathrm{UESC}} = \mathbf{h}_{11} + \mathbf{h}_{12}\mathbf{X}, \tag{46}$$

$$\mathbf{L}_{+}^{\mathrm{SESC}} = \frac{1}{2}(\tilde{\mathbf{S}}_{+}\mathbf{S}_{11}^{-1}\mathbf{L}_{+}^{\mathrm{UESC}} + c.c.), \tag{47}$$

$$\mathbf{L}_{+}^{\mathrm{NESC}} = \mathbf{h}_{11} + \mathbf{h}_{12}\mathbf{X} + \mathbf{X}^{\dagger}\mathbf{h}_{21} + \mathbf{X}^{\dagger}\mathbf{h}_{22}\mathbf{X}, \tag{48}$$

$$\mathbf{C}_{+} = \mathbf{R}_{+}^{-1}\mathbf{A}_{+}, \tag{49}$$

while the lower-right block defines the equation for the NES:

$$\mathbf{h}_{-}\mathbf{C}_{-} = \mathbf{S}_{22}\mathbf{C}_{-}E_{-}, \tag{50}$$

$$\mathbf{h}_{-} = \mathbf{R}_{-}^{\dagger}\mathbf{L}_{-}^{\mathrm{X}}\mathbf{R}_{-}, \quad \mathrm{X} = \mathrm{UESC}, \mathrm{SESC}, \mathrm{NESC}, \tag{51}$$

$$\mathbf{L}_{-}^{\mathrm{UESC}} = \mathbf{h}_{22} + \mathbf{h}_{21}\tilde{\mathbf{X}}, \tag{52}$$

$$\mathbf{L}_{-}^{\mathrm{SESC}} = \frac{1}{2}(\tilde{\mathbf{S}}_{-}\mathbf{S}_{22}^{-1}\mathbf{L}_{-}^{\mathrm{UESC}} + c.c.), \tag{53}$$

$$\mathbf{L}_{-}^{\mathrm{NESC}} = \mathbf{h}_{22} + \mathbf{h}_{21}\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^{\dagger}\mathbf{h}_{12} + \tilde{\mathbf{X}}^{\dagger}\mathbf{h}_{11}\tilde{\mathbf{X}}, \tag{54}$$

$$\mathbf{C}_{-} = \mathbf{R}_{-}^{-1}\mathbf{B}_{-}. \tag{55}$$

Here the acronyms UESC, NESC, and SESC refer to the unnormalized, normalized [15], and symmetrized [9] eliminations of the small component, respectively. By plugging the matrix elements of $\mathbf{h}_{\mathrm{sf}}^{D}$ (19) and $\mathbf{M}$ (20) into Eq. (28), the above procedure gives rise to the spin-free transformation $\mathbf{U}_0$ parameterized by Eq. (30) with $\mathbf{X}_0$ (43). The transformation (41) of $\mathbf{h}_{\mathrm{sf}}^{D}$ by $\mathbf{U}_0$ then yields the desired spin-free Hamiltonian $\mathbf{h}_{+,\mathrm{sf}}$ (45) in Eq. (3) that covers all scalar relativistic effects.

The remaining task is to determine the spin-orbit part $\mathbf{h}_{\mathrm{sd}}$ (23) for $\mathbf{h}_{\mathrm{SO},1e}$ in Eq. (3). The $\mathbf{U}_0$ transformation of $\mathbf{h}^{D}$ (17) yields a transformed Hamiltonian $\mathbf{h}_{1}^{D}$ that can be written as

$$\mathbf{h}_{1}^{D} = \mathbf{U}_0\mathbf{h}^{D}\mathbf{U}_0^{\dagger} = \mathbf{U}_0\mathbf{h}_{\mathrm{sf}}^{D}\mathbf{U}_0^{\dagger} + \mathbf{U}_0\mathbf{h}_{\mathrm{sd}}^{D}\mathbf{U}_0^{\dagger} \tag{56}$$

$$= \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{O}_1, \tag{57}$$

where the subscripts count the order in $W_{\rm sd}$. The even operators $\mathcal{E}_0$ and $\mathcal{E}_1$ as well as the odd operator $\mathcal{O}_1$ read explicitly

$$\mathcal{E}_0 = \begin{pmatrix} \mathbf{E}_{+,0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{-,0} \end{pmatrix} = \begin{pmatrix} \mathbf{h}_{+,\rm sf} & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_{-,\rm sf} \end{pmatrix}, \tag{58}$$

$$\mathcal{E}_1 = \begin{pmatrix} \mathbf{E}_{+,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{-,1} \end{pmatrix}, \tag{59}$$

$$\mathcal{O}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{O}_1 \\ \mathbf{O}_1^\dagger & \mathbf{0} \end{pmatrix}, \tag{60}$$

$$\mathbf{E}_{+,1} = \frac{\alpha^2}{4}\mathbf{R}_{+,0}^\dagger \mathbf{X}_0^\dagger \mathbf{W}_{\rm sd}\mathbf{X}_0\mathbf{R}_{+,0}, \tag{61}$$

$$\mathbf{O}_1 = \frac{\alpha^2}{4}\mathbf{R}_{+,0}^\dagger \mathbf{X}_0^\dagger \mathbf{W}_{\rm sd}\mathbf{R}_{-,0}, \tag{62}$$

$$\mathbf{E}_{-,1} = \frac{\alpha^2}{4}\mathbf{R}_{-,0}^\dagger \mathbf{W}_{\rm sd}\mathbf{R}_{-,0}. \tag{63}$$

There exist two distinct schemes to further block-diagonalize $\mathbf{h}_1^D$ (56) so as to obtain $\mathbf{h}_+$ in Eq. (23). As outlined before, the DKH approach amounts to invoking a series of transformations to eliminate at each step the lowest-order odd terms in $W_{\rm sd}$, leading eventually to

$$\begin{aligned} \mathbf{h}_{bd}^{\rm DKH} &= \cdots \mathbf{U}_3\mathbf{U}_2\mathbf{U}_1\mathbf{U}_0\mathbf{h}^D\mathbf{U}_0^\dagger\mathbf{U}_1^\dagger\mathbf{U}_2^\dagger\mathbf{U}_3^\dagger\cdots \\ &= \cdots \mathbf{U}_3\mathbf{U}_2\mathbf{U}_1\mathbf{h}_1^D\mathbf{U}_1^\dagger\mathbf{U}_2^\dagger\mathbf{U}_3^\dagger\cdots \\ &= \sum_{k=0}^{\infty}\mathcal{E}_k. \end{aligned} \tag{64}$$

The upper-left block $\mathbf{h}_+$ of $\mathbf{h}_{bd}^{\rm DKH}$ is separated automatically into spin-free and spin-dependent parts

$$\mathbf{h}_{+,\rm sf} = \mathbf{E}_{+,0}, \quad \mathbf{h}_{+,\rm sd}^{\rm DKH} = \sum_{k=1}^{\infty}\mathbf{E}_{+,k}. \tag{65}$$

As the spin-orbit operator $W_{\rm sd}$ always goes with the factor $\frac{\alpha^2}{4}$, each term of $\mathbf{E}_{+,k}$ (65) is of $\mathcal{O}(\alpha^{2k}W_{\rm sd}^k)$.

Different from the multiple transformations, the second approach [19, 20] for block-diagonalizing $\mathbf{h}_1^D$ (56) invokes a single unitary transformation $\mathbf{U}_Y$ that is parameterized by Eq. (30) with the decoupling matrix $\mathbf{Y}$ subject to a condition of the same form as Eq. (42), viz.,

$$\mathbf{h}_{bd}^{\rm BSS} = \mathbf{U}_Y\mathbf{h}_1^D\mathbf{U}_Y^\dagger = \mathbf{U}_Y\mathbf{U}_0\mathbf{h}^D\mathbf{U}_0^\dagger\mathbf{U}_Y^\dagger. \tag{66}$$

At variance with this two-step approach for block-diagonalizing $\mathbf{h}^D$, the X2C Hamiltonian $\mathbf{h}_+$ is obtained in one step:

$$\mathbf{h}_{bd}^{\mathrm{X2C}} = \mathbf{U}_X \mathbf{h}^D \mathbf{U}_X^\dagger. \tag{67}$$

That is, the initial spin-free transformation $\mathbf{U}_0$ is not performed at all. Making perturbation expansions [25] of the decoupling condition (42) for $\mathbf{Y}_k$ or $\mathbf{X}_k$ and the renormalization condition (37) for $\mathbf{R}_{+,k}$ will give rise to the spin-free and spin-dependent parts of $\mathbf{h}_+$ (45) as

$$\mathbf{h}_{+,\mathrm{sf}} = \mathbf{h}_{+,0}, \quad \mathbf{h}_{+,\mathrm{sd}}^{\mathrm{X}} = \sum_{k=1}^{\infty} \mathbf{h}_{+,k}^{\mathrm{X}}, \quad \mathrm{X} = \mathrm{BSS}, \mathrm{X2C}. \tag{68}$$

The two- and one-step approaches for block-diagonalizing $\mathbf{h}^D$ are closely related to each other. That is, there exists [8] a closed relation between $\mathbf{U}_Y \mathbf{U}_0$ and $\mathbf{U}_X$, such that they can be mapped to each other uniquely. In the context of spin separation, a subtle difference in between lies in that the zeroth-order term $\mathbf{Y}_0$ of $\mathbf{Y}$ vanishes, whereas the zeroth-order term $\mathbf{X}_0$ of $\mathbf{X}$ is just the spin-free matrix (43) used for parameterizing $\mathbf{U}_0$. The former is due to the fact that $\mathbf{h}_1^D$ (56) is already diagonal in the absence of SOC. This feature of the BSS-like expansion results in some simplifications when deriving explicitly the lowest-order terms of $\mathbf{h}_{+,k}$ (see section "BSS"). However, such an advantage disappears for higher-order terms [25]. Note also that the spin-free part $\mathbf{h}_{+,\mathrm{sf}}$ is the same for Eqs. (65) and (68) and therefore does not carry a superscript.

The above general procedures can be used to construct arbitrary order DKH ($\mathbf{E}_{+,k}$), BSS ($\mathbf{h}_{+,k}^{\mathrm{BSS}}$), and X2C ($\mathbf{h}_{+,k}^{\mathrm{X2C}}$) types of spin-dependent Hamiltonians; for more details, see Ref. [25]. Truncating the expansions to finite orders leads naturally to a sequence of novel spin-dependent Hamiltonians. To reveal important distinctions between the DKH, BSS, and X2C approaches, we construct their second- and third-order terms explicitly in section "The Lowest-Order Spin-Orbit Terms."

The Lowest-Order Spin-Orbit Terms

DKH

In the DKH-like scheme, the transformation $\mathbf{U}_1$ in Eq. (64) subject to condition (36) can be parameterized by an anti-Hermitian matrix $\mathcal{W}_1$, viz.,

$$\mathbf{U}_1 = \exp(\mathcal{W}_1 \mathbf{M}^{-1}), \quad \mathcal{W}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{W}_1 \\ -\mathbf{W}_1^\dagger & \mathbf{0} \end{pmatrix}, \tag{69}$$

$$\mathbf{U}_1^\dagger = \exp(\mathbf{M}^{-1} \mathcal{W}_1^\dagger) = \exp(-\mathbf{M}^{-1} \mathcal{W}_1). \tag{70}$$

By virtue of the asymmetric Baker-Campbell-Hausdorff (BCH) expansion, an arbitrary matrix $\mathbf{N}$ can be transformed as

$$\begin{aligned}\mathbf{U}\mathbf{N}\mathbf{U}^{\dagger} &= \exp(\mathcal{W}_1\mathbf{M}^{-1})\mathbf{N}\exp(-\mathbf{M}^{-1}\mathcal{W}_1)\\ &= \mathbf{N} + [\mathcal{W}_1, \mathbf{N}]_{\mathbf{M}^{-1}} + \frac{1}{2}[\mathcal{W}_1, [\mathcal{W}_1, \mathbf{N}]_{\mathbf{M}^{-1}}]_{\mathbf{M}^{-1}} + \cdots, \end{aligned} \tag{71}$$

where the $\mathbf{M}^{-1}$-commutator is defined as

$$[\mathcal{W}_1, \mathbf{N}]_{\mathbf{M}^{-1}} = \mathcal{W}_1\mathbf{M}^{-1}\mathbf{N} - \mathbf{N}\mathbf{M}^{-1}\mathcal{W}_1. \tag{72}$$

Since $[\mathcal{W}_1, \mathbf{M}]_{\mathbf{M}^{-1}} = 0$ holds for arbitrary $\mathcal{W}_1$, it can be verified from Eq. (71) that the parametrization (69) does satisfy Eq. (36). In view of Eq. (71), the $\mathbf{U}_1$-transformation of $\mathbf{h}_1^D$ (57) can be written as

$$\begin{aligned}\mathbf{h}_2^D &= \mathbf{U}_1\mathbf{h}_1^D\mathbf{U}_1^{\dagger}\\ &= \mathbf{h}_1^D + [\mathcal{W}_1, \mathbf{h}_1^D]_{\mathbf{M}^{-1}} + \frac{1}{2}[\mathcal{W}_1, [\mathcal{W}_1, \mathbf{h}_1^D]_{\mathbf{M}^{-1}}]_{\mathbf{M}^{-1}} + \cdots\\ &= \mathcal{E}_0 + \mathcal{E}_1 + \mathcal{O}_1^{(2)} + \mathcal{E}_2 + \mathcal{O}_2^{(2)} + \mathcal{E}_3 + \cdots, \end{aligned} \tag{73}$$

with

$$\mathcal{O}_1^{(2)} = \mathcal{O}_1 + [\mathcal{W}_1, \mathcal{E}_0]_{\mathbf{M}^{-1}}, \tag{74}$$

$$\mathcal{E}_2 = [\mathcal{W}_1, \mathcal{O}_1]_{\mathbf{M}^{-1}} + \frac{1}{2}[\mathcal{W}_1, [\mathcal{W}_1, \mathcal{E}_0]_{\mathbf{M}^{-1}}]_{\mathbf{M}^{-1}}, \tag{75}$$

$$\mathcal{O}_2^{(2)} = [\mathcal{W}_1, \mathcal{E}_1]_{\mathbf{M}^{-1}}, \tag{76}$$

$$\mathcal{E}_3 = \frac{1}{2}[\mathcal{W}_1, [\mathcal{W}_1, \mathcal{E}_1]_{\mathbf{M}^{-1}}]_{\mathbf{M}^{-1}}. \tag{77}$$

By choosing $\mathcal{W}_1$ to eliminate $\mathcal{O}_1^{(2)}$, viz.,

$$[\mathcal{W}_1, \mathcal{E}_0]_{\mathbf{M}^{-1}} = -\mathcal{O}_1, \tag{78}$$

the second-order even term $\mathcal{E}_2$ becomes

$$\mathcal{E}_2 = \frac{1}{2}[\mathcal{W}_1, \mathcal{O}_1]_{\mathbf{M}^{-1}}. \tag{79}$$

The so-determined $\mathcal{W}_1$ also specifies the third-order even term $\mathcal{E}_3$ (77) in accordance with the $2n + 1$ rule [27].

Equation (78) can be recast into a Sylvester equation

$$2c^2\mathbf{W}_1\mathbf{T}^{-1}\mathbf{E}_{-,0} - \mathbf{E}_{+,0}\mathbf{S}^{-1}\mathbf{W}_1 = -\mathbf{O}_1, \tag{80}$$

in terms of the two-component quantities. The equation can readily be solved in the orthonormal basis spanned by the eigenfunctions of $\mathcal{E}_0$, viz.,

$$\mathbf{W}_1 = (\mathbf{C}_{+,0}^\dagger)^{-1}\mathbf{w}_1(\mathbf{C}_{-,0})^{-1} = \frac{\alpha^2}{2}\mathbf{S}\mathbf{C}_{+,0}\mathbf{w}_1\mathbf{C}_{-,0}^\dagger\mathbf{T}, \tag{81}$$

$$[\mathbf{w}_1]_{pq} = -\frac{[\mathbf{o}_1]_{pq}}{[E_{-,0}]_q - [E_{+,0}]_p}, \quad \mathbf{o}_1 = \mathbf{C}_{+,0}^\dagger\mathbf{O}_1\mathbf{C}_{-,0}. \tag{82}$$

Since the scalar PES and NES are well separated in energy, the denominator of $[\mathbf{w}_1]_{pq}$ (82) is never nonvanishing, such that the solution (81) is unique. In terms of the so-obtained $\mathbf{W}_1$, the second-order (79) and third-order (77) spin-orbit DKH Hamiltonians read

$$\mathbf{E}_{+,2} = c^2(\mathbf{W}_1\mathbf{T}^{-1}\mathbf{O}_1^\dagger + c.c.), \tag{83}$$

$$\mathbf{E}_{+,3} = c^2(\mathbf{W}_1\mathbf{T}^{-1}\mathbf{O}_2^\dagger + c.c.), \tag{84}$$

$$\mathbf{O}_2 = 2c^2\mathbf{W}_1\mathbf{T}^{-1}\mathbf{E}_{-,1} - \mathbf{E}_{+,1}\mathbf{S}^{-1}\mathbf{W}_1. \tag{85}$$

It has been shown [25] that both $\mathbf{E}_{+,0}$ and $\mathbf{E}_{-,0}$ are of $\mathcal{O}(\alpha^0)$, while $\mathbf{E}_{+,2}$ and $\mathbf{E}_{+,3}$ are of $\mathcal{O}(W_{\mathrm{sd}}^2\alpha^4)$ and $\mathcal{O}(W_{\mathrm{sd}}^3\alpha^6)$, respectively. The leading odd term becomes $\mathbf{O}_2$, which is of $\mathcal{O}(W_{\mathrm{sd}}^2\alpha^4)$. If the fourth-order ($\mathbf{E}_{+,4}$) and fifth-order ($\mathbf{E}_{+,5}$) terms are wanted, the $\mathbf{O}_2$ operator can be eliminated by a second transformation with $\mathbf{W}_2$ satisfying

$$2c^2\mathbf{W}_2\mathbf{T}^{-1}\mathbf{E}_{-,0} - \mathbf{E}_{+,0}\mathbf{S}^{-1}\mathbf{W}_2 = -\mathbf{O}_2, \tag{86}$$

which is of the same mathematical form as Eq. (80).

BSS

The first three orders of the BSS spin-orbit Hamiltonian can be derived via power expansions (analytic derivatives) in $\mathbf{W}_{\mathrm{sd}}$. Since $\mathbf{Y}_0 = \mathbf{0}$, the expansions of $\tilde{\mathbf{S}}_+^{\mathrm{BSS}}$ and $\mathbf{R}_+^{\mathrm{BSS}}$ (37) go as

$$\begin{aligned}\tilde{\mathbf{S}}_+^{\mathrm{BSS}} &= \tilde{\mathbf{S}}_{+,0}^{\mathrm{BSS}} + \tilde{\mathbf{S}}_{+,1}^{\mathrm{BSS}} + \tilde{\mathbf{S}}_{+,2}^{\mathrm{BSS}} + \tilde{\mathbf{S}}_{+,3}^{\mathrm{BSS}} + \cdots \\ &= \mathbf{S} + \mathbf{0} + \frac{\alpha^2}{2}\mathbf{Y}_1^\dagger\mathbf{T}\mathbf{Y}_1 + \frac{\alpha^2}{2}(\mathbf{Y}_2^\dagger\mathbf{T}\mathbf{Y}_1 + \mathbf{Y}_1^\dagger\mathbf{T}\mathbf{Y}_2) + \cdots,\end{aligned} \tag{87}$$

$$\begin{aligned}\mathbf{R}_+^{\mathrm{BSS}} &= \mathbf{R}_{+,0}^{\mathrm{BSS}} + \mathbf{R}_{+,1}^{\mathrm{BSS}} + \mathbf{R}_{+,2}^{\mathrm{BSS}} + \mathbf{R}_{+,3}^{\mathrm{BSS}} + \cdots \\ &= \mathbf{I} + \mathbf{0} - \frac{1}{2}\mathbf{S}^{-1}\tilde{\mathbf{S}}_{+,2}^{\mathrm{BSS}} - \frac{1}{2}\mathbf{S}^{-1}\tilde{\mathbf{S}}_{+,3}^{\mathrm{BSS}} + \cdots,\end{aligned} \tag{88}$$

while the expansion of $\mathbf{h}_{+}^{\mathrm{BSS}}$ reads

$$\mathbf{h}_{+}^{\mathrm{BSS}} = \mathbf{h}_{+,0}^{\mathrm{BSS}} + \mathbf{h}_{+,1}^{\mathrm{BSS}} + \mathbf{h}_{+,2}^{\mathrm{BSS}} + \mathbf{h}_{+,3}^{\mathrm{BSS}} + \cdots, \tag{89}$$

$$\mathbf{h}_{+,0}^{\mathrm{BSS}} = \tilde{\mathbf{L}}_{+,0}^{\mathrm{NESC}} = \mathbf{E}_{+,0}, \tag{90}$$

$$\mathbf{h}_{+,1}^{\mathrm{BSS}} = \tilde{\mathbf{L}}_{+,1}^{\mathrm{NESC}} = \mathbf{E}_{+,1}, \tag{91}$$

$$\mathbf{h}_{+,2}^{\mathrm{BSS}} = \tilde{\mathbf{L}}_{+,2}^{\mathrm{NESC}} + \frac{1}{2}(\mathbf{R}_{+,2}^{\mathrm{BSS}\dagger}\tilde{\mathbf{L}}_{+,0}^{\mathrm{NESC}} + c.c.), \tag{92}$$

$$\tilde{\mathbf{L}}_{+,2}^{\mathrm{NESC}} = \mathbf{O}_1\mathbf{Y}_1 + \mathbf{Y}_1^{\dagger}\mathbf{O}_1^{\dagger} + \mathbf{Y}_1^{\dagger}\mathbf{E}_{-,0}\mathbf{Y}_1, \tag{93}$$

$$\mathbf{h}_{+,3}^{\mathrm{BSS}} = \tilde{\mathbf{L}}_{+,3}^{\mathrm{NESC}} + (\mathbf{R}_{+,2}^{\mathrm{BSS}\dagger}\tilde{\mathbf{L}}_{+,1}^{\mathrm{NESC}} + \mathbf{R}_{+,3}^{\mathrm{BSS}\dagger}\tilde{\mathbf{L}}_{+,0}^{\mathrm{NESC}} + c.c.), \tag{94}$$

$$\tilde{\mathbf{L}}_{+,3}^{\mathrm{NESC}} = \mathbf{O}_1\mathbf{Y}_2 + \mathbf{Y}_2^{\dagger}\mathbf{O}_1^{\dagger} + \mathbf{Y}_2^{\dagger}\mathbf{E}_{-,0}\mathbf{Y}_1 + \mathbf{Y}_1^{\dagger}\mathbf{E}_{-,0}\mathbf{Y}_2 + \mathbf{Y}_1^{\dagger}\mathbf{E}_{-,1}\mathbf{Y}_1. \tag{95}$$

The expansion of the decoupling condition (42), which in this case becomes

$$\mathbf{O}_1^{\dagger} + (\mathbf{E}_{-,0} + \mathbf{E}_{-,1})\mathbf{Y} = \frac{\alpha^2}{2}\mathbf{T}\mathbf{Y}\mathbf{S}^{-1}(\mathbf{E}_{+,0} + \mathbf{E}_{+,1} + \mathbf{O}_1\mathbf{Y}), \tag{96}$$

can be made to determine $\mathbf{Y}_1$ and $\mathbf{Y}_2$, viz.,

$$\mathbf{E}_{-,0}\mathbf{Y}_1 - \frac{\alpha^2}{2}\mathbf{T}\mathbf{Y}_1\mathbf{S}^{-1}\mathbf{E}_{+,0} = -\mathbf{O}_1^{\dagger}, \tag{97}$$

$$\mathbf{E}_{-,0}\mathbf{Y}_2 - \frac{\alpha^2}{2}\mathbf{T}\mathbf{Y}_2\mathbf{S}^{-1}\mathbf{E}_{+,0} = -\tilde{\mathbf{O}}_2^{\dagger}, \tag{98}$$

with

$$\tilde{\mathbf{O}}_2^{\dagger} = \mathbf{E}_{-,1}\mathbf{Y}_1 - \frac{\alpha^2}{2}\mathbf{T}\mathbf{Y}_1\mathbf{S}^{-1}\mathbf{E}_{+,1}. \tag{99}$$

It has been found [25] that $\mathbf{h}_{+,2}^{\mathrm{BSS}}$ is identical with the second-order DKH Hamiltonian $\mathbf{E}_{+,2}$, viz.,

$$\mathbf{h}_{+,2}^{\mathrm{BSS}} = \mathbf{E}_{+,2} = \frac{1}{2}(\mathbf{Y}_1^{\dagger}\mathbf{O}_1^{\dagger} + c.c.) = c^2(\mathbf{W}_1\mathbf{T}^{-1}\mathbf{O}_1^{\dagger} + c.c.). \tag{100}$$

This stems from the fact that $\mathbf{h}_{+,2}^{\mathrm{BSS}}$ involves only $\mathbf{Y}_1$ and that the relation $\mathbf{Y}_1 = 2c^2\mathbf{T}^{-1}\mathbf{W}_1^{\dagger}$ or equivalently $\mathbf{W}_1 = \frac{\alpha^2}{2}\mathbf{T}\mathbf{Y}_1^{\dagger}$ holds in view of Eqs. (80) and (97). Because of this identity, the $\tilde{\mathbf{O}}_2$ (99) and $\mathbf{O}_2$ (85) operators are actually the same, thereby leading to $\mathbf{Y}_2 = 2c^2\mathbf{T}^{-1}\mathbf{W}_2^{\dagger}$. The third-order term $\mathbf{h}_{+,3}^{\mathrm{BSS}}$ can hence be written as

$$\mathbf{h}_{+,3}^{\mathrm{BSS}} = \frac{1}{2}(\mathbf{Y}_2^{\dagger}\mathbf{O}_1^{\dagger} + c.c.) = c^2(\mathbf{W}_2\mathbf{T}^{-1}\mathbf{O}_1^{\dagger} + c.c.), \tag{101}$$

which is related to the third-order DKH Hamiltonian $\mathbf{E}_{+,3}$ (84) via

$$\mathbf{h}_{+,3}^{\mathrm{BSS}} = \mathbf{E}_{+,3} + (\mathbf{E}_{+,0}\boldsymbol{\Gamma}_+ + c.c.), \tag{102}$$

where

$$\boldsymbol{\Gamma}_+ = \mathbf{S}^{-1}\boldsymbol{\Gamma}_A, \quad \boldsymbol{\Gamma}_A = \frac{1}{2}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}^\dagger),$$
$$\boldsymbol{\Gamma} = -\frac{\alpha^2}{2}\mathbf{Y}_2^\dagger\mathbf{T}\mathbf{Y}_1 = -2c^2\mathbf{W}_2\mathbf{T}^{-1}\mathbf{W}_1^\dagger. \tag{103}$$

It is seen that while $\mathbf{E}_{+,3}$ does not involve $\mathbf{W}_2$ due to the underlying $2n+1$ rule [27], the BSS-type $\mathbf{h}_{+,3}^{\mathrm{BSS}}$ does depend on $\mathbf{Y}_2$. The difference between $\mathbf{h}_{+,3}^{\mathrm{BSS}}$ and $\mathbf{E}_{+,3}$, the second term in Eq. (102), is of $\mathcal{O}(W_{\mathrm{sd}}^3\alpha^8)$, higher than the leading order of $\mathbf{h}_{+,3}^{\mathrm{BSS}}$ and $\mathbf{E}_{+,3}$ by $\mathcal{O}(\alpha^2)$. This results from the different renormalizations (i.e., picture changes) of the BSS and DKH approaches.

X2C

The lowest-order spin-orbit terms of $\mathbf{h}_+^{\mathrm{X2C}}$ can be derived in the same way as done for those of $\mathbf{h}_+^{\mathrm{BSS}}$. However, the procedure is somewhat more lengthy for the reason that the spin-free matrix $\mathbf{X}_0$ of X2C is nonvanishing. Alternatively, we can make use of the fact [8] that the transformations $\mathbf{U}_X$ and $\mathbf{U}_Y\mathbf{U}_0$ are related to each other via a picture change matrix $\boldsymbol{\Omega}$:

$$\mathbf{U}_X = \boldsymbol{\Omega}^\dagger(\mathbf{U}_Y\mathbf{U}_0), \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_+ & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_- \end{pmatrix}, \tag{104}$$

such that the X2C and BSS Hamiltonians can be related simply by

$$\mathbf{h}_+^{\mathrm{X2C}} = \boldsymbol{\Omega}_+^\dagger\mathbf{h}_+^{\mathrm{BSS}}\boldsymbol{\Omega}_+. \tag{105}$$

To determine $\boldsymbol{\Omega}$, Eq. (104) can be written more explicitly as

$$\begin{pmatrix} \mathbf{R}_+^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_-^\dagger \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{X}^\dagger \\ \tilde{\mathbf{X}}^\dagger & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Omega}_+^\dagger & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_-^\dagger \end{pmatrix}\begin{pmatrix} \mathbf{R}_+^{\mathrm{BSS}\dagger} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_-^{\mathrm{BSS}\dagger} \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{Y}^\dagger \\ \tilde{\mathbf{Y}}^\dagger & \mathbf{I} \end{pmatrix}$$
$$\times \begin{pmatrix} \mathbf{R}_{+,0}^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{-,0}^\dagger \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{X}_0^\dagger \\ \tilde{\mathbf{X}}_0^\dagger & \mathbf{I} \end{pmatrix}, \tag{106}$$

which leads to

$$\mathbf{X}(\mathbf{R}_{+,0} + \tilde{\mathbf{X}}_0\mathbf{R}_{-,0}\mathbf{Y}) = \mathbf{X}_0\mathbf{R}_{+,0} + \mathbf{R}_{-,0}\mathbf{Y}, \tag{107}$$

$$(\mathbf{R}_{+,0} + \tilde{\mathbf{X}}_0\mathbf{R}_{-,0}\mathbf{Y})\mathbf{R}_+^{\mathrm{BSS}}\boldsymbol{\Omega}_+ = \mathbf{R}_+, \tag{108}$$

where $\mathbf{R}_\pm$ represent the X2C renormalization matrices in $\mathbf{U}_X$; see Eqs. (31) and (32). Equation (107) is independent of $\mathbf{R}_\pm$ and $\mathbf{R}_\pm^{\mathrm{BSS}}$ and hence allows the calculation of $\mathbf{X}$ from $\mathbf{Y}$ or vice versa. Equation (108) then determines $\mathbf{\Omega}_+$. The perturbative expansions of Eqs. (107), (108), and (105) will establish the relation between $\mathbf{h}_{+,k}^{\mathrm{X2C}}$ and $\mathbf{h}_{+,k}^{\mathrm{BSS}}$ at each order k. For $k = 0$, Eq. (107) is just an identity as $\mathbf{Y}_0 = 0$. For $k \geq 1$, it gives rise to

$$\mathbf{X}_k = (\mathbf{R}_{-,0}\mathbf{Y}_k - [\mathbf{X}\tilde{\mathbf{X}}_0\mathbf{R}_{-,0}\mathbf{Y}]'_k)\mathbf{R}_{+,0}^{-1}, \tag{109}$$

where the prime indicates the exclusion of terms involving $\mathbf{X}_k$. This shows that $\mathbf{X}_k$ is composed generally of a *connected* term $\mathbf{Y}_k$ and some *disconnected* terms as products of lower-order $\mathbf{Y}_p$. The perturbative expansions of Eqs. (108) and (105) lead to

$$\mathbf{\Omega}_{+,0} = \mathbf{I}, \tag{110}$$

$$\mathbf{\Omega}_{+,1} = \mathbf{R}_{+,0}^{-1}(\mathbf{R}_{+,1} - \tilde{\mathbf{X}}_0\mathbf{R}_{-,0}\mathbf{Y}_1), \tag{111}$$

$$\mathbf{h}_{+,0}^{\mathrm{X2C}} = \mathbf{h}_{+,0}^{\mathrm{BSS}} = \mathbf{E}_{+,0}, \tag{112}$$

$$\mathbf{h}_{+,1}^{\mathrm{X2C}} = \mathbf{h}_{+,1}^{\mathrm{BSS}} + (\mathbf{h}_{+,0}^{\mathrm{BSS}}\mathbf{\Omega}_{+,1} + c.c.) = \mathbf{E}_{+,1} + (\mathbf{E}_{+,0}\mathbf{\Omega}_{+,1} + c.c.). \tag{113}$$

It is clear that the difference between the spin-orbit X2C and BSS (and DKH) Hamiltonians occurs already to first order. Yet, this difference is higher by $\mathcal{O}(\alpha^2)$ than the order of the Hamiltonians [25], similar to that between BSS and DKH albeit starting from third order. Such an analysis reveals clearly the distinctions between the DKH, BSS, and X2C approaches: They share the same decoupling condition and differ only in the picture change reflected by the different renormalizations. As far as lowest-order SOC are concerned, the DKH type of formulation is much preferred due to the underlying $2n + 1$ rule [27].

Relations with A2C Hamiltonians

It is also instructive to compare the present results with the well-known Pauli, ZORA, and DKH1 spin-orbit Hamiltonians:

$$h_{\mathrm{SO},1e}^{\mathrm{Pauli}} = \frac{\alpha^2}{4} W_{\mathrm{sd}}, \tag{114}$$

$$h_{\mathrm{SO},1e}^{\mathrm{ZORA}} = i\vec{\sigma} \cdot (\vec{p} \times \frac{c^2}{2c^2 - V}\vec{p}), \tag{115}$$

$$h_{\mathrm{SO},1e}^{\mathrm{DKH1}} = \alpha^2 \mathcal{Q} W_{\mathrm{sd}} \mathcal{Q}, \quad \mathcal{Q} = \frac{c^2}{\sqrt{2E_p(E_p + c^2)}}, \quad E_p = \sqrt{c^4 + c^2\vec{p}^2}. \tag{116}$$

The matrix representation of $h_{\mathrm{SO},1e}^{\mathrm{Pauli}}$ (114),

$$\mathbf{E}_{+,1} \approx \frac{\alpha^2}{4}\mathbf{W}_{\mathrm{sd}}, \tag{117}$$

amounts to replacing the sf-X2C $\mathbf{X}_0$ and $\mathbf{R}_{+,0}$ matrices in $\mathbf{E}_{+,1}$ (61) with their nonrelativistic limits, i.e., $\mathbf{X}_0 \approx \mathbf{I}$ and $\mathbf{R}_{+,0} \approx \mathbf{I}$, respectively. Similarly, the matrix representation [25] of $h_{\mathrm{SO},1e}^{\mathrm{ZORA}}$ (115),

$$\mathbf{E}_{+,1}^{\mathrm{ZORA}} = \frac{\alpha^2}{4}\mathbf{X}_0^{\mathrm{ZORA}\dagger}\mathbf{W}_{\mathrm{sd}}\mathbf{X}_0^{\mathrm{ZORA}}, \tag{118}$$

$$\mathbf{X}_0^{\mathrm{ZORA}} = (\mathbf{T} - \frac{\alpha^2}{4}\mathbf{W}_{\mathrm{sf}})^{-1}\mathbf{T}, \tag{119}$$

can also be understood as an approximation to $\mathbf{E}_{+,1}$ (61) with $\mathbf{R}_{+,0} \approx \mathbf{I}$ and $\mathbf{X}_0^{\mathrm{ZORA}}$ (119) in place of the sf-X2C $\mathbf{X}_0$. Likewise, the matrix representation of $h_{\mathrm{SO},1e}^{\mathrm{DKH1}}$ (116) also takes the same form as $\mathbf{E}_{+,1}$ (61) but with $\mathbf{X}_0$ and $\mathbf{R}_{+,0}$ determined from the free-particle part of the Dirac equation (17):

$$\begin{pmatrix} \mathbf{0} & \mathbf{T} \\ \mathbf{T} & -\mathbf{T} \end{pmatrix} \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \frac{\alpha^2}{2}\mathbf{T} \end{pmatrix} \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} E. \tag{120}$$

In sum, all these approximate spin-orbit Hamiltonians can be recovered by approximating appropriately the sf-X2C $\mathbf{X}_0$ and $\mathbf{R}_{+,0}$ matrices.

A Short Summary

Apart from the compactness of the expressions, one important feature of the $\mathbf{h}_{+,n}^{\mathrm{X}}$ (X = DKH, BSS, X2C) Hamiltonians lies in that their errors decay very fast as the order n increases [25]. A brief comparison between sf-X2C+so-DKHn and ZORA for the SOC of states of Rn^{85+} is shown in Fig. 1, where their errors relative to the solutions of the matrix Dirac equation (17) are plotted. It is seen that, at each order, sf-X2C+so-DKHn describes uniformly all the states of the same angular momentum yet of different principal quantum numbers, while ZORA works reasonably well only for the valence states. In other words, ZORA can be viewed as a "valence-electron" Hamiltonian, whereas sf-X2C+so-DKHn is truly an "all-electron" Hamiltonian. In particular, the sf-X2C+so-DKH3 Hamiltonian,

$$\mathbf{h}_{+,3}^{\mathrm{DKH}} = \mathbf{E}_{+,0} + \mathbf{E}_{+,1} + \mathbf{E}_{+,2} + \mathbf{E}_{+,3}, \tag{121}$$

is sufficiently accurate for the SOC, with the relative errors being less than 0.01% for all the states.

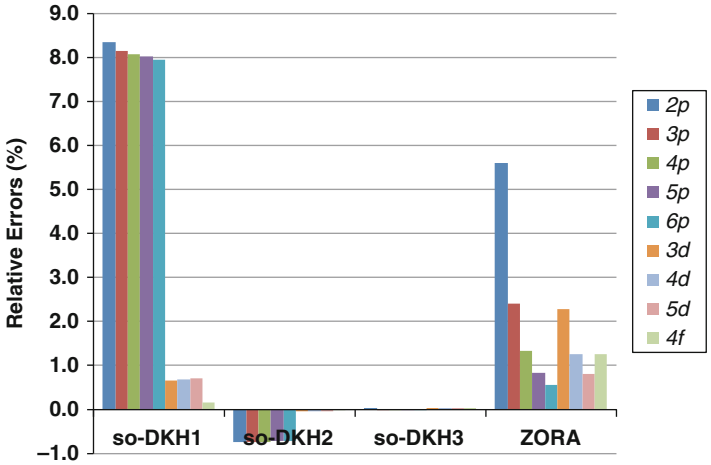


Fig. 1 Relative errors of sf-X2C+so-DKHn and ZORA for the spin-orbit splittings of states of Rn^{85+} (The data are taken from Ref. [25])

At this stage, it is to be noted that both $\mathbf{E}_{+,2}$ (83) and $\mathbf{E}_{+,3}$ (84), stemming from the spin separation though, are not yet of the desired form of Eq. (1). To achieve this, a further spin separation is to be carried out [28]. To do so, we first note that the $\mathbf{W}_{\rm sd}$ matrix in a spin-orbital basis $\{g_\mu = \chi_\mu\theta_\mu\}$ composed of spatial functions $\{\chi_\mu\}$ and spin functions $\{\theta_\mu = \alpha, \beta\}$ can be written as

$$\begin{aligned}\mathbf{W}_{\rm sd} &= \mathbb{i}(\sigma_x \otimes \mathbf{V}^x + \sigma_y \otimes \mathbf{V}^y + \sigma_z \otimes \mathbf{V}^z) \\ &= \mathbb{i}\begin{pmatrix} \mathbf{V}^z & \mathbf{V}^x - \mathbb{i}\mathbf{V}^y \\ \mathbf{V}^x + \mathbb{i}\mathbf{V}^y & -\mathbf{V}^z \end{pmatrix} \triangleq \mathbb{i}\vec{\sigma}\cdot\vec{\mathbf{V}}, \end{aligned} \tag{122}$$

$$[\mathbf{V}^m]_{\mu\nu} = \langle\chi_\mu|(\vec{p}V \times \vec{p})_m|\chi_\nu\rangle, \quad m = x, y, z, \tag{123}$$

with $\mathbf{V}^m$ being real-valued antisymmetric matrices for real-valued functions $\{\chi_\mu\}$. Since both $\mathbf{X}_0$ and $\mathbf{R}_{+,0}$ are spin free, all the first-order spin-dependent terms, including $\mathbf{E}_{+,1}$, $\mathbf{O}_1$, $\mathbf{E}_{-,1}$ and $\mathbf{W}_{+,1}$, can be rewritten in the same form as $\mathbf{W}_{\rm sd}$ (122), e.g., $\mathbf{E}_{+,1} = \mathbb{i}\vec{\sigma}\cdot\vec{\mathbf{E}}^{\rm sd}_{+,1}$ with $\vec{\mathbf{E}}^{\rm sd}_{+,1}$ in place of $\vec{\mathbf{V}}$. In this way, $\mathbf{E}_{+,2}$ (83) can further be separated into two parts by using the Dirac identity (2), viz.,

$$\mathbf{E}^{\rm sf}_{+,2} = \frac{1}{\alpha^2}(\vec{\mathbf{W}}_1\cdot\mathbf{T}^{-1}\vec{\mathbf{O}}_1^T + \vec{\mathbf{O}}_1\cdot\mathbf{T}^{-1}\vec{\mathbf{W}}_1^T), \tag{124}$$

$$\mathbf{E}^{\rm sd}_{+,2} = \frac{\mathbb{i}}{\alpha^2}\vec{\sigma}\cdot(\vec{\mathbf{W}}_1\times\mathbf{T}^{-1}\vec{\mathbf{O}}_1^T + \vec{\mathbf{O}}_1\times\mathbf{T}^{-1}\vec{\mathbf{W}}_1^T). \tag{125}$$

Similarly, $\mathbf{O}_2$ (85) can be separated into

$$\mathbf{O}_2^{\mathrm{sf}} = -(2c^2\vec{\mathbf{W}}_1 \cdot \mathbf{T}^{-1}\vec{\mathbf{E}}_{-,1} - \vec{\mathbf{E}}_{+,1} \cdot \mathbf{S}^{-1}\vec{\mathbf{W}}_1), \tag{126}$$

$$\mathbf{O}_2^{\mathrm{sd}} = -\mathbb{i}\vec{\sigma} \cdot (2c^2\vec{\mathbf{W}}_1 \times \mathbf{T}^{-1}\vec{\mathbf{E}}_{-,1} - \vec{\mathbf{E}}_{+,1} \times \mathbf{S}^{-1}\vec{\mathbf{W}}_1), \tag{127}$$

such that $\mathbf{E}_{+,3}$ (84) can be rewritten as

$$\mathbf{E}_{+,3}^{\mathrm{sf}} = \frac{1}{\alpha^2}(\vec{\mathbf{W}}_1 \cdot \mathbf{T}^{-1}\vec{\mathbf{O}}_2^{\mathrm{sd},T} + \vec{\mathbf{O}}_2^{\mathrm{sd}} \cdot \mathbf{T}^{-1}\vec{\mathbf{W}}_1^T), \tag{128}$$

$$\begin{aligned}\mathbf{E}_{+,3}^{\mathrm{sd}} = \frac{\mathbb{i}}{\alpha^2}\vec{\sigma} \cdot (&\vec{\mathbf{W}}_1\mathbf{T}^{-1}\mathbf{O}_2^{\mathrm{sf},T} - \mathbf{O}_2^{\mathrm{sf}}\mathbf{T}^{-1}\vec{\mathbf{W}}_1^T \\ &+\vec{\mathbf{W}}_1 \times \mathbf{T}^{-1}\vec{\mathbf{O}}_2^{\mathrm{sd},T} + \vec{\mathbf{O}}_2^{\mathrm{sd}} \times \mathbf{T}^{-1}\vec{\mathbf{W}}_1^T).\end{aligned} \tag{129}$$

As such, the sf-X2C+so-DKH3 Hamiltonian (121) can be rewritten in a fully spin-separated form

$$\mathbf{h}_{+,3}^{\mathrm{DKH}} = (\mathbf{E}_{+,0} + \mathbf{E}_{+,2}^{\mathrm{sf}} + \mathbf{E}_{+,3}^{\mathrm{sf}}) + \mathbb{i}\vec{\sigma} \cdot (\vec{\mathbf{E}}_{+,1}^{\mathrm{sd}} + \vec{\mathbf{E}}_{+,2}^{\mathrm{sd}} + \vec{\mathbf{E}}_{+,3}^{\mathrm{sd}}). \tag{130}$$

The constructions of $\mathbf{E}_{+,2}^{\mathrm{sf/sd}}$ and $\mathbf{E}_{+,3}^{\mathrm{sf/sd}}$ are essentially free, since all necessary quantities are already available after constructing the sf-X2C Hamiltonian $\mathbf{E}_{+,0}$ (65). The high accuracy and compactness of $\mathbf{h}_{+,3}^{\mathrm{DKH}}$ (130) arise from the spin-free and spin-dependent partitioning (26) of the Dirac matrix (17), the one-step matrix formulation (30) of the sf-X2C Hamiltonian $\mathbf{h}_{+,\mathrm{sf}}$, as well as the $2n + 1$ rule [27] in the DKH type of transformation of the spin-dependent terms. Such spin separation can also be applied to molecular properties [28].

The two-component Hamiltonians discussed so far can be collected to form a "two-component Hamiltonian ladder" (see Fig. 2), which is just the lower segment of the complete Hamiltonian ladder [23, 29]. From bottom to up, the lowest rung is the nonrelativistic SC Hamiltonian. The second rung covers BP and DPT [24], where both spin-free and spin-dependent terms are of finite orders in α^2. The third rung includes combinations of sf-X2C with spin-dependent terms of order α^2. The fourth rung includes combinations of sf-X2C with the truncated spin-dependent Hamiltonians $\mathbf{E}_{+,k}$ (65) or $\mathbf{h}_{+,k}^{\mathrm{BSS/X2C}}$ (68) [NB: Although truncated, terms of higher orders in α^2 are still present]. The highest rung includes the original X2C [= sf-X2C+so-X2C(∞)], BSS, and DKH Hamiltonians, as well as the novel sf-X2C+so-BSS(∞) and sf-X2C+so-DKH(∞) combinations. Such classification of the two-component Hamiltonians is meaningful not only conceptually but also computationally. Starting with a nonrelativistic program, the Hamiltonians in the second rung will introduce some additional terms that have been available in many program packages. A further upgrade of the zeroth-order term to the sf-X2C Hamiltonian $h_{+,\mathrm{sf}}$ will generate the Hamiltonians in the third rung. To construct the Hamiltonians

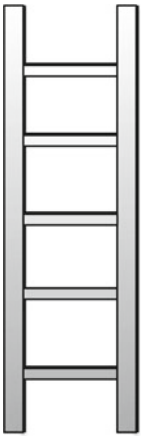


Fig. 2 The two-component relativistic Hamiltonian ladder (Source: Reprinted with permission from J. Chem. Phys. 141, 054111 (2014). © 2014, American Institute of Physics)

in the fourth rung, the responses of the decoupling matrix $\mathbf{X}$ and renormalization matrix $\mathbf{R}_+$ have to be calculated. Finally, to construct the Hamiltonians in the fifth rung, the nonrelativistic program should entirely be replaced by a two-component one. Therefore, in order to retain the nonrelativistic program and meanwhile achieve accurate treatments of relativistic effects, it is the Hamiltonians in the third and fourth rungs that should be used. By contrast, the Hamiltonians in the second rung have no particular advantages in both accuracy and efficiency and should hence be regarded as obsolete. Yet, to complete the spin-dependent Hamiltonians in the third and fourth rungs, a proper treatment of the two-electron spin-orbit coupling should further be introduced; see section "Two-Electron Spin-Orbit Coupling."

Two-Electron Spin-Orbit Coupling

Following the spin separation of the one-electron Dirac Hamiltonian (7), the spin separation of a two-body operator can be achieved in three steps: (a) The RKB transformation (8) is first carried out to pair the Pauli spin matrices, (b) the spin-dependent terms can be identified by applying the Dirac identity (2), and (c) the results can be transformed to two-component form via unitary transformations. Here, both the Coulomb and Gaunt interactions

$$G_{ij}^{C} = \frac{1}{r_{ij}}, \quad G_{ij}^{G} = -\frac{\vec{\alpha}_i \cdot \vec{\alpha}_j}{r_{ij}} \tag{131}$$

are considered (NB: the spin-orbit terms from the Breit interaction are the same as those from the Gaunt interaction). After the RKB transformation (8), we have

$$\tilde{g}_{ij}^{C} = (Z_i \otimes Z_j)^{\dagger} G_{ij}^{C} (Z_i \otimes Z_j) = \begin{pmatrix} g_{ij}^{C,00} & 0 & 0 & 0 \\ 0 & \frac{\alpha^2}{4} g_{ij}^{C,02} & 0 & 0 \\ 0 & 0 & \frac{\alpha^2}{4} g_{ij}^{C,20} & 0 \\ 0 & 0 & 0 & \frac{\alpha^4}{16} g_{ij}^{C,22} \end{pmatrix},$$

$$\begin{aligned} \tilde{g}_{ij}^{G} &= (Z_i \otimes Z_j)^{\dagger} G_{ij}^{G} (Z_i \otimes Z_j) \\ &= \begin{pmatrix} 0 & 0 & 0 & \frac{\alpha^2}{4} g_{ij}^{G,++} \\ 0 & 0 & \frac{\alpha^2}{4} g_{ij}^{G,+-} & 0 \\ 0 & \frac{\alpha^2}{4} g_{ij}^{G,-+} & 0 & 0 \\ \frac{\alpha^2}{4} g_{ij}^{G,--} & 0 & 0 & 0 \end{pmatrix}, \end{aligned} \tag{132}$$

where

$$g_{ij}^{C,00} = g_{ij}^{C}, \quad g_{ij}^{C,02} = \Pi_j g_{ij}^{C} \Pi_j,$$

$$g_{ij}^{C,20} = \Pi_i g_{ij}^{C} \Pi_i, \quad g_{ij}^{C,22} = \Pi_i \Pi_j g_{ij}^{C} \Pi_i \Pi_j, \tag{133}$$

$$g_{ij}^{G,++} = g_{ij}^{G} \Pi_i \Pi_j, \quad g_{ij}^{G,+-} = \Pi_j g_{ij}^{G} \Pi_i,$$

$$g_{ij}^{G,-+} = \Pi_i g_{ij}^{G} \Pi_j, \quad g_{ij}^{G,--} = \Pi_i \Pi^j g_{ij}^{G}, \tag{134}$$

$$g_{ij}^{C} = \frac{1}{r_{12}}, \quad g_{ij}^{G} = -\frac{\vec{\sigma}_1 \cdot \vec{\sigma}_2}{r_{12}}, \quad \Pi_i = \vec{\sigma}_i \cdot \vec{p}_i. \tag{135}$$

It is clear that in the transformed Coulomb operators (133), there is always a pair of σ in Π for the same electron, while in the transformed Gaunt terms (134), the σ in Π is paired with that in g_{ij}^{G}. Therefore, the spin separation can be achieved simply by using the Dirac identity (2), with the terms linear in $\vec{\sigma}$ being

$$\begin{aligned} g_{ij,\mathrm{SO}}^{C,20} &= [\Pi_i g_{ij}^{C} \Pi_i]_{\mathrm{SO}} = \mathbb{i} \vec{\sigma}_i \cdot (\vec{p}_i \times \frac{1}{r_{ij}} \vec{p}_i), \\ g_{ij,\mathrm{SO}}^{C,02} &= [\Pi_j g_{ij}^{C} \Pi_j]_{\mathrm{SO}} = \mathbb{i} \vec{\sigma}_j \cdot (\vec{p}_j \times \frac{1}{r_{ij}} \vec{p}_j), \\ g_{ij,\mathrm{SO}}^{G,++} &= [g_{ij}^{G} \Pi_i \Pi_j]_{\mathrm{SO}} = -\mathbb{i} (\vec{\sigma}_j - \vec{\sigma}_i) \cdot \frac{1}{r_{ij}} (\vec{p}_i \times \vec{p}_j), \\ g_{ij,\mathrm{SO}}^{G,-+} &= [\Pi_i g_{ij}^{G} \Pi_j]_{\mathrm{SO}} = -\mathbb{i} (\vec{\sigma}_i + \vec{\sigma}_j) \cdot (\vec{p}_i \times \frac{1}{r_{ij}} \vec{p}_j), \\ g_{ij,\mathrm{SO}}^{G,+-} &= [\Pi_j g_{ij}^{G} \Pi_i]_{\mathrm{SO}} = -\mathbb{i} (\vec{\sigma}_i + \vec{\sigma}_j) \cdot (\vec{p}_j \times \frac{1}{r_{ij}} \vec{p}_i), \\ g_{ij,\mathrm{SO}}^{G,--} &= [\Pi_i \Pi_j g_{ij}^{G}]_{\mathrm{SO}} = -\mathbb{i} (\vec{\sigma}_i - \vec{\sigma}_j) \cdot (\vec{p}_i \times \vec{p}_j) \frac{1}{r_{ij}}. \end{aligned} \tag{136}$$

To get two-component spin-orbit operators that are correct to $\mathcal{O}(\alpha^2)$, only a spin-free $\mathbf{U}_0$ transformation of the operators in Eq. (136) is sufficient, leading to $H_{\mathrm{sd},2e}$ composed of both spin-same-orbit (SSO) and spin-other-orbit (SOO) interactions that can conveniently be written in second-quantized form [28]:

$$H_{\mathrm{sd},2e} = H_{\mathrm{SSO},2e} + H_{\mathrm{SOO},2e}, \tag{137}$$

$$H_{\mathrm{SSO},2e} = \sum_{pqrs} \frac{\alpha^2}{4} \langle pq|G_{\mathrm{SSO}}|rs\rangle a_p^\dagger a_q^\dagger a_s a_r, \tag{138}$$

$$H_{\mathrm{SOO},2e} = \sum_{pqrs} \frac{\alpha^2}{4} \langle pq|G_{\mathrm{SOO}}|rs\rangle a_p^\dagger a_q^\dagger a_s a_r, \tag{139}$$

where

$$\begin{aligned}
&\langle pq|G_{\mathrm{SSO}}|rs\rangle \\
&= \sum_{\mu\nu\kappa\lambda} (\mathbf{X}_0\mathbf{A})^*_{\mu p} (\mathbf{A})^*_{\kappa q} \langle \mu\kappa|\mathrm{i}\vec{\sigma}_1 \cdot (\vec{p}_1 \times \frac{1}{r_{12}} \vec{p}_1)|\nu\lambda\rangle (\mathbf{X}_0\mathbf{A})_{\nu r} \mathbf{A}_{\lambda s}, \qquad (140) \\
&\langle pq|G_{\mathrm{SOO}}|rs\rangle \\
&= \sum_{\mu\nu\kappa\lambda} \Bigg[\mathbf{A}^*_{\mu p} \mathbf{A}^*_{\kappa q} \langle \mu\kappa|\mathrm{i}\vec{\sigma}_1 \cdot \frac{1}{r_{12}} (\vec{p}_1 \times \vec{p}_2)|\nu\lambda\rangle (\mathbf{X}_0\mathbf{A})_{\nu r} (\mathbf{X}_0\mathbf{A})_{\lambda s} \\
&\quad + (\mathbf{X}_0\mathbf{A})^*_{\mu p} \mathbf{A}^*_{\kappa q} \langle \mu\kappa| - \mathrm{i}\vec{\sigma}_1 \cdot (\vec{p}_1 \times \frac{1}{r_{12}} \vec{p}_2)|\nu\lambda\rangle \mathbf{A}_{\nu r} (\mathbf{X}_0\mathbf{A})_{\lambda s} \\
&\quad + \mathbf{A}^*_{\mu p} (\mathbf{X}_0\mathbf{A})^*_{\kappa q} \langle \mu\kappa| - \mathrm{i}\vec{\sigma}_1 \cdot (\vec{p}_2 \times \frac{1}{r_{12}} \vec{p}_1)|\nu\lambda\rangle (\mathbf{X}_0\mathbf{A})_{\nu r} \mathbf{A}_{\lambda s} \\
&\quad + (\mathbf{X}_0\mathbf{A})^*_{\mu p} (\mathbf{X}_0\mathbf{A})^*_{\kappa q} \langle \mu\kappa| - \mathrm{i}\vec{\sigma}_1 \cdot (\vec{p}_1 \times \vec{p}_2) \frac{1}{r_{12}} |\nu\lambda\rangle \mathbf{A}_{\nu r} \mathbf{A}_{\lambda s} \Bigg], \qquad (141)
\end{aligned}$$

where $\mathbf{A} = \mathbf{R}_{+,0}\mathbf{C}_{+,0}$ (49). Note in passing that the integral $\langle pq|G_X|rs\rangle$ ($X = \mathrm{SSO}, \mathrm{SOO}$) is itself not symmetric with respect to the exchange of particles. However, the sum $\langle pq|G_X|rs\rangle + \langle qp|G_X|sr\rangle$ does have the particle symmetry. Once the sf-X2C decoupling matrix $\mathbf{X}_0$ is made available, the operators $H_{\mathrm{SSO},2e}$ (138) and $H_{\mathrm{SOO},2e}$ (139) can be constructed and used to describe SOC between configurations differing by up to two orbitals. Again, the BP and DKH1 spin-orbit Hamiltonians can be obtained [28] by using the previous approximations for $\mathbf{X}_0$ and $\mathbf{R}_{+,0}$.

Due to the weak interplay with electron correlation, the two-electron spin-orbit operators (138) and (139) can further be simplified by invoking a mean-field

approximation, which amounts to replacing $a_p^\dagger a_q^\dagger a_s a_r$ in Eqs. (138) and (139) with $n_p(\delta_{pr}a_q^\dagger a_s - \delta_{ps}a_q^\dagger a_r)$ (i.e., keeping orbital p always occupied). The resulting effective one-electron spin-orbit operator, to be used in Eq. (5), can be written as [28]

$$f_{\mathrm{SO},2e} = \sum_{pq}(f_{\mathrm{SO},2e})_{pq}a_p^\dagger a_q, \quad (f_{\mathrm{SO},2e})_{pq} = (\mathbf{C}_{+,0}^\dagger \mathbf{F}_{\mathrm{SO},2e}\mathbf{C}_{+,0})_{pq}, \tag{142}$$

where

$$\mathbf{F}_{\mathrm{SO},2e} = \mathbf{R}_{+,0}^\dagger \tilde{\mathbf{L}}_{\mathrm{SO},2e}^{\mathrm{NESC}}\mathbf{R}_{+,0}, \tag{143}$$

$$\tilde{\mathbf{L}}_{\mathrm{SO},2e}^{\mathrm{NESC}} = \frac{\alpha^2}{4}[\mathbf{G}_{\mathrm{SO}}^{LL} + \mathbf{G}_{\mathrm{SO}}^{LS}\mathbf{X} + \mathbf{X}^\dagger\mathbf{G}_{\mathrm{SO}}^{SL} + \mathbf{X}^\dagger\mathbf{G}_{\mathrm{SO}}^{SS}\mathbf{X}], \tag{144}$$

$$[\mathbf{G}_{\mathrm{SO}}^{LL}]_{\mu\nu} = \sum_{\kappa\lambda} P_{\lambda\kappa}^{SS}[\langle\mu\kappa|[\Pi_2 g_{12}^C \Pi_2]_{\mathrm{SO}}|\nu\lambda\rangle - \langle\mu\kappa|[\Pi_2 g_{12}^G \Pi_1]_{\mathrm{SO}}|\lambda\nu\rangle], \tag{145}$$

$$\begin{aligned}[\mathbf{G}_{\mathrm{SO}}^{LS}]_{\mu\nu} = &\sum_{\kappa\lambda} P_{\lambda\kappa}^{SL}[\langle\mu\kappa|[g_{12}^G \Pi_1 \Pi_2]_{\mathrm{SO}}|\nu\lambda\rangle - \langle\mu\kappa|[g_{12}^G \Pi_1 \Pi_2]_{\mathrm{SO}}|\lambda\nu\rangle] \\ &+ \sum_{\kappa\lambda} P_{\lambda\kappa}^{LS}[\langle\mu\kappa|[\Pi_2 g_{12}^G \Pi_1]_{\mathrm{SO}}|\nu\lambda\rangle - \langle\mu\kappa|[\Pi_2 g_{12}^C \Pi_2]_{\mathrm{SO}}|\lambda\nu\rangle],\end{aligned} \tag{146}$$

$$\begin{aligned}[\mathbf{G}_{\mathrm{SO}}^{SL}]_{\mu\nu} = &\sum_{\kappa\lambda} P_{\lambda\kappa}^{LS}[\langle\mu\kappa|[\Pi_1 \Pi_2 g_{12}^G]_{\mathrm{SO}}|\nu\lambda\rangle - \langle\mu\kappa|[\Pi_1 \Pi_2 g_{12}^G]_{\mathrm{SO}}|\lambda\nu\rangle] \\ &+ \sum_{\kappa\lambda} P_{\lambda\kappa}^{SL}[\langle\mu\kappa|[\Pi_1 g_{12}^G \Pi_2]_{\mathrm{SO}}|\nu\lambda\rangle - \langle\mu\kappa|[\Pi_1 g_{12}^C \Pi_1]_{\mathrm{SO}}|\lambda\nu\rangle],\end{aligned} \tag{147}$$

$$[\mathbf{G}_{\mathrm{SO}}^{SS}]_{\mu\nu} = \sum_{\kappa\lambda} P_{\lambda\kappa}^{LL}[\langle\mu\kappa|[\Pi_1 g_{12}^C \Pi_1]_{\mathrm{SO}}|\nu\lambda\rangle - \langle\mu\kappa|[\Pi_1 g_{12}^G \Pi_2]_{\mathrm{SO}}|\lambda\nu\rangle]. \tag{148}$$

Here, the spin-free density matrices $\mathbf{P}^{XY}$ are defined as

$$\begin{aligned}\mathbf{P}^{LL} &= \mathbf{R}_{+,0}\mathbf{P}_0\mathbf{R}_{+,0}^\dagger, \quad \mathbf{P}_0 = \mathbf{C}_{+,0}\mathbf{n}\mathbf{C}_{+,0}^\dagger, \\ \mathbf{P}^{LS} &= \mathbf{P}^{LL}\mathbf{X}_0^\dagger = \mathbf{P}^{SL\dagger}, \quad \mathbf{P}^{SS} = \mathbf{X}_0\mathbf{P}^{LL}\mathbf{X}_0^\dagger.\end{aligned} \tag{149}$$

If the spin-free density matrices $\mathbf{P}_0^\alpha$ and $\mathbf{P}_0^\beta$ are further averaged as in Eq. (6), the $\mathbf{G}_{\mathrm{SO}}^{XY}$ matrices can be simplified [28] to

$$G_{\mathrm{SO}}^{XY} = \mathbb{i}\vec{\sigma}\cdot\vec{g}^{XY} = \mathbb{i}\sigma_l g^{XY,l}, \quad l \in \{x, y, z\}, \tag{150}$$

$$g_{\mu\nu}^{LL,l} = \sum_{\kappa\lambda} g_{\mu\nu,\kappa\lambda}^{G,LL,l} P_{\lambda\kappa}^{SS}, \quad g_{\mu\nu,\kappa\lambda}^{G,LL,l} = -2K_{\lambda\mu,\kappa\nu}^{l}, \tag{151}$$

$$g^{LS,l}_{\mu\nu} = \sum_{\kappa\lambda} g^{C,LS,l}_{\mu\nu,\kappa\lambda} P^{LS}_{\lambda\kappa}, \quad g^{C,LS,l}_{\mu\nu,\kappa\lambda} = -K^{l}_{\mu\lambda,\kappa\nu} - K^{l}_{\lambda\mu,\kappa\nu}, \tag{152}$$

$$g^{SL,l}_{\mu\nu} = \sum_{\kappa\lambda} g^{C,SL,l}_{\mu\nu,\kappa\lambda} P^{SL}_{\lambda\kappa} = -g^{LS,l}_{\nu\mu}, \quad g^{C,SL,l}_{\mu\nu,\kappa\lambda} = K^{l}_{\mu\lambda,\kappa\nu} + K^{l}_{\mu\lambda,\nu\kappa}, \tag{153}$$

$$g^{SS,l}_{\mu\nu} = \sum_{\kappa\lambda} (g^{C,SS,l}_{\mu\nu,\kappa\lambda} + g^{G,SS,l}_{\mu\nu,\kappa\lambda}) P^{SS}_{\lambda\kappa},$$

$$g^{C,SS,l}_{\mu\nu,\lambda\kappa} = -2(K^{l}_{\mu\nu,\kappa\lambda} + K^{l}_{\mu\nu,\lambda\kappa}), \quad g^{G,SS,l}_{\mu\nu,\lambda\kappa} = 2K^{l}_{\mu\lambda,\nu\kappa}, \tag{154}$$

where the antisymmetric K integrals are defined as

$$K^{l}_{\mu\nu,\kappa\lambda} = \epsilon_{lmn}(\mu_m \nu|\kappa_n \lambda) = -K^{l}_{\kappa\lambda,\mu\nu},$$
$$\mu_m = \partial_m \mu, \quad l,m,n \in \{x,y,z\}. \tag{155}$$

It is seen that $g^{LS,l}_{\mu\nu}/g^{SL,l}_{\mu\nu}$ and $g^{LL,l}_{\mu\nu}$ are due entirely to the Coulomb-exchange and Gaunt-exchange interactions, respectively, while $g^{SS,l}_{\mu\nu}$ (154) is a mixture of the Coulomb-direct and Gaunt-exchange interactions. The Gaunt-direct interaction vanishes due to spin averaging. If the approximations $\mathbf{X}_0 \approx \mathbf{I}$ and $\mathbf{R}_{+,0} \approx \mathbf{I}$ are further introduced, the $\mathbf{F}_{\mathrm{SO},2e}$ matrix (143) will reduce to that of the mean-field BP spin-orbit Hamiltonian [4],

$$[\mathbf{F}^{\mathrm{BP}}_{\mathrm{SO},2e}]_{\mu\nu} = \sum_{\kappa\lambda} \mathbb{i}\sigma^{l}_{\mu\nu}[g^{C,SS,l}_{\mu\nu,\lambda\kappa} + (g^{G,LL,l}_{\mu\nu,\kappa\lambda} + g^{G,SS,l}_{\mu\nu,\kappa\lambda}) + (g^{C,LS,l}_{\mu\nu,\kappa\lambda} + g^{C,SL,l}_{\mu\nu,\kappa\lambda})] P_{\lambda\kappa}$$
$$= \sum_{\kappa\lambda} \mathbb{i}\sigma^{l}_{\mu\nu}[g^{C,SS,l}_{\mu\nu,\lambda\kappa} + 3(g^{C,LS,l}_{\mu\nu,\kappa\lambda} + g^{C,SL,l}_{\mu\nu,\kappa\lambda})] P_{\lambda\kappa}. \tag{156}$$

At this level, the Gaunt-exchange interaction is just twice the Coulomb-exchange interaction. Since Eqs. (143) and (156) share the same kind of integrals $K^{l}_{\mu\nu,\kappa\lambda}$ (155), their computational costs are virtually the same, for the additional matrix multiplications involved in the former are very cheap. The computational cost for the evaluation of $K^{l}_{\mu\nu,\kappa\lambda}$ can substantially be reduced by employing the local nature of spin-orbit interactions. For instance, the one-center approximation for $K^{l}_{\mu\nu,\kappa\lambda}$ is already sufficiently accurate for the SOC of valence properties [4, 5, 30]. A two-center approximation for $K^{l}_{\mu\nu,\kappa\lambda}$ would even be better. With the so-defined $\mathbf{F}_{\mathrm{SO},2e}$ Eq. (143), every term of the Hamiltonian (3) has been specified, which can then be combined with the analytic derivative technique, response theory, or QDPT to calculate properties of many-electron systems.

Applications: Perturbative Treatment of SOC

Ground-State Spin-Dependent Properties

For a perturbed Hamiltonian $H(\lambda) = H_0 + \lambda H_1$, we have the following relation:

$$\frac{dE(\lambda)}{d\lambda} = \langle \Psi(\lambda)|H_1|\Psi(\lambda)\rangle, \tag{157}$$

provided that the wave function $|\Psi(\lambda)\rangle$ is exact and normalized, viz.,

$$H(\lambda)|\Psi(\lambda)\rangle = |\Psi(\lambda)\rangle E(\lambda), \quad \langle \Psi(\lambda)|\Psi(\lambda)\rangle = 1. \tag{158}$$

This is nothing but the Hellmann-Feynman theorem, which implies that the static property $\langle \Psi(0)|H_1|\Psi(0)\rangle$ can alternatively be calculated as the energy derivative $\frac{dE(\lambda)}{d\lambda}|_{\lambda=0}$. While this makes no difference for exact theory, the energy derivative type of formulations of molecular properties is more general [31] and can be used for approximate models, which are not fully variational or even have no explicit many-body wave functions. Taking the spin-dependent Hamiltonian $H_{\rm sd}$ (5) as the perturbation λH_1, the total energy of a many-electron system can be expressed as

$$E = E_{\rm sf} + E_{\rm sd}, \quad E_{\rm sf} = E_0, \quad E_{\rm sd} = \sum_{k=1}^{\infty} E_k, \quad E_k = \frac{1}{k!} \left. \frac{d^k E(\lambda)}{d\lambda^k} \right|_{\lambda=0}. \tag{159}$$

Here, the first term $E_{\rm sf}$ is simply the spin-free energy, while E_k represents the kth-order SOC correction. How to find the derivatives of the energy is technically the same as the previous derivation of the spin-orbit Hamiltonians via an order-by-order expansion.

Consider the Hartree-Fock (HF) model. Suppose the HF spin orbitals have been optimized with the sf-X2C Hamiltonian $H_{\rm sf}$ (4), which constitute an orthonormal basis for expanding the full Hamiltonian $H_{\rm el}$ (3) and the spin-free HF reference $|\Phi_0\rangle$ in the occupation number representation, the HF energy at a finite strength λ can then be written as

$$E_{\rm HF}(\lambda) = \langle \Phi_{\rm HF}(\lambda)|H(\lambda)|\Phi_{\rm HF}(\lambda)\rangle = \langle \Phi_0|e^{-\kappa(\lambda)} H(\lambda) e^{\kappa(\lambda)}|\Phi_0\rangle, \tag{160}$$

where

$$|\Phi_{\rm HF}(\lambda)\rangle = e^{\kappa(\lambda)}|\Phi_0\rangle, \tag{161}$$

$$\kappa(\lambda) = \sum_{pq} \kappa_{pq}(\lambda) a_p^\dagger a_q, \quad \kappa_{pq} = -\kappa_{qp}^*. \tag{162}$$

By using the BCH expansion, $E_{\mathrm{HF}}(\lambda)$ (160) can be expanded in λ as

$$E_{\mathrm{HF}}(\lambda) = E_{\mathrm{HF},0} + E_{\mathrm{HF},1}\lambda + E_{\mathrm{HF},2}\lambda^2 + E_{\mathrm{HF},3}\lambda^3 + \cdots, \tag{163}$$

$$E_{\mathrm{HF},0} = \langle\Phi_0|H_0|\Phi_0\rangle, \tag{164}$$

$$E_{\mathrm{HF},1} = \langle\Phi_0|H_1|\Phi_0\rangle + \langle\Phi_0|[H_0,\kappa_1]|\Phi_0\rangle, \tag{165}$$

$$\begin{aligned} E_{\mathrm{HF},2} = {} & \langle\Phi_0|H_2|\Phi_0\rangle + \langle\Phi_0|[H_0,\kappa_2]|\Phi_0\rangle \\ & + \langle\Phi_0|[H_1,\kappa_1]|0\rangle + \frac{1}{2}\langle\Phi_0|[[H_0;\kappa_1,\kappa_1]|\Phi_0\rangle, \end{aligned} \tag{166}$$

$$\begin{aligned} E_{\mathrm{HF},3} = {} & \langle\Phi_0|H_3|\Phi_0\rangle + \langle\Phi_0|[H_0,\kappa_3]|\Phi_0\rangle \\ & + \langle\Phi_0|[H_1,\kappa_2]|\Phi_0\rangle + \langle\Phi_0|[H_2,\kappa_1]|\Phi_0\rangle \\ & + \langle\Phi_0|[H_0;\kappa_1,\kappa_2]|\Phi_0\rangle + \frac{1}{2}\langle\Phi_0|[H_1;\kappa_1,\kappa_1]|\Phi_0\rangle \\ & + \frac{1}{3!}\langle\Phi_0|[H_0;\kappa_1,\kappa_1,\kappa_1]|\Phi_0\rangle, \end{aligned} \tag{167}$$

$$H_k = h^{\mathrm{X}}_{\mathrm{sd},k} + \delta_{k,1} f_{\mathrm{sd},1}, \quad k \geq 1, \quad \mathrm{X} = \mathrm{X2C}, \mathrm{BSS}, \mathrm{DKH}, \tag{168}$$

where a similar expansion for κ has been made (NB: $\kappa_0 = 0$), while $h^{\mathrm{X}}_{\mathrm{sd},k}$ and $f_{\mathrm{sd},1}$ represent the one-electron and mean-field (142) spin-orbit terms. The symmetrized commutators are defined as the symmetric sum of all possible permutations π of k symbols:

$$[A; B_1, B_2, \cdots, B_k] = \frac{1}{k!}\sum_{\pi\in\mathcal{S}_k}[[\cdots[A, B_{\pi(1)}], B_{\pi(2)}], \cdots, B_{\pi(k)}]. \tag{169}$$

The orbital responses κ_k in Eqs. (166) and (167) can be found from the expansion of the stationary condition

$$\begin{aligned} 0 = \frac{\partial E_{\mathrm{HF}}(\lambda)}{\partial\kappa_{pq}(\lambda)} = {} & \langle\Phi_0|[H(\lambda), a_p^\dagger a_q]|\Phi_0\rangle + \langle\Phi_0|[H(\lambda); a_p^\dagger a_q, \kappa(\lambda)]|\Phi_0\rangle \\ & + \frac{1}{2}\langle\Phi_0|[H(\lambda); a_p^\dagger a_q, \kappa(\lambda), \kappa(\lambda)]|\Phi_0\rangle + \cdots. \end{aligned} \tag{170}$$

Specifically, the first three-order equations read

$$\langle\Phi_0|[H_0, a_p^\dagger a_q]|\Phi_0\rangle = 0, \tag{171}$$

$$\langle\Phi_0|[H_0; a_p^\dagger a_q, \kappa_1]|\Phi_0\rangle = -\langle\Phi_0|[H_1, a_p^\dagger a_q]|\Phi_0\rangle, \tag{172}$$

$$\begin{aligned} \langle\Phi_0|[H_0; a_p^\dagger a_q, \kappa_2]|\Phi_0\rangle = {} & -\langle\Phi_0|[H_2, a_p^\dagger a_q]|\Phi_0\rangle - \langle\Phi_0|[H_1; a_p^\dagger a_q, \kappa_1]|\Phi_0\rangle \\ & - \frac{1}{2}\langle\Phi_0|[H_0; a_p^\dagger a_q, \kappa_1, \kappa_1]|\Phi_0\rangle. \end{aligned} \tag{173}$$

For a spin-restricted closed-shell system, the condition (171) holds for the whole manifold of the operators $\{a_p^\dagger a_q\}$, such that the $\langle\Phi_0|[H_0,\kappa_k]|\Phi_0\rangle$ $(k \geq 1)$ terms all vanish. The $\langle\Phi_0|h^X_{\mathrm{sd},1} + f_{\mathrm{sd},1}|\Phi_0\rangle$ term also vanishes in this case due to time reversal symmetry. As a result, the lowest-order spin-orbit correction to a closed-shell system starts from second-order $E_{\mathrm{HF},2}$. As for an open-shell reference $|\Phi_0\rangle$, the condition (171) holds in general only for a subset of the operators $\{a_p^\dagger a_q\}$. The $\langle\Phi_0|h^X_{\mathrm{sd},1} + f_{\mathrm{sd},1}|\Phi_0\rangle$ term also does not vanish. Therefore, the spin-orbit correction to an open-shell system appears already at first-order $E_{\mathrm{HF},1}$. It is also seen that the left-hand sides of the coupled-perturbed equations (172) and (173) are of the same structure, such that they can be solved recursively with a single solver. The situation is very similar to the inhomogeneous equations for $\mathbf{W}_1$ (80) and $\mathbf{W}_2$ (86) used for constructing the spin-orbit DKH Hamiltonians.

The above procedure can readily be extended to multiple perturbations. For instance, a second-order property (e.g., dipole polarizability or nuclear magnetic resonance shielding tensor) due to perturbations V^a and V^b can be expanded as

$$E^{ab} = \left(\frac{d^2E}{d\lambda_a d\lambda_b}\right)_{\lambda=0} = \sum_{k=0}^{\infty} E_k^{ab}, \quad E_k^{ab} = \frac{1}{k!}\frac{d^k E^{ab}(\lambda)}{d\lambda^k}|_{\lambda=0}. \tag{174}$$

where the second equality represents the expansion in the strength of SOC: E_0^{ab} represents the sf-X2C term, while E_k^{ab} $(k \geq 1)$ represent the spin-orbit corrections. Thus, similar to the Hamiltonian (68) and energy (159), the property E^{ab} can also be separated into spin-free and spin-dependent terms:

$$E^{ab} = E^{ab}_{\mathrm{sf}} + E^{ab}_{\mathrm{sd}}, \quad E^{ab}_{\mathrm{sf}} = E^{ab}_0, \quad E^{ab}_{\mathrm{sd}} = \sum_{k=1}^{\infty} E_k^{ab}. \tag{175}$$

The $2n + 1$ rule of perturbation theory may further be invoked to simplify the calculation of E_k^{ab}. The only caveat is that, to avoid picture change errors, the property operators must also be transformed accordingly with the transformations that are used to transform the four-component Hamiltonian. Various transformed operators have been discussed in Ref. [28] and are not repeated here.

Excited-State Spin-Dependent Properties

According to time-dependent perturbation theory, an excited-state property can be extracted from the response of the ground state to time-dependent perturbations characterized by the interaction operator $V(t)$

$$V(t) = \sum_k e^{-\mathrm{i}\omega_k t} \sum_b V^b(\omega_k)\epsilon_b(\omega_k), \tag{176}$$

where $V^b(\omega_k)$ represents a perturbation operator V^b associated with frequency ω_k and $\epsilon_b(\omega_k)$ is the corresponding perturbation strength. The idea has been utilized

to formulate a response theory for ground-state dynamic properties, transition properties between different states, as well as excited-state properties. It can be viewed as an extension of the analytic derivative technique. In particular, by taking the zero limit of the frequencies ω_k, ground-state static properties can be recovered. By taking spin-dependent operators as the perturbations (176), spin-dependent properties, such as spin-orbit matrix elements between two spin-free excited states, can be extracted from the residues of the response functions. However, it is to be noted that the response theory is essentially a nondegenerate perturbation theory, implying that an order-by-order expansion cannot be made for studying spin-dependent excited-state properties that mix degenerate spin-free states. Fine structure splittings of spin-free states represent such an example.

To calculate fine structures, the response theory can first be applied to a given theoretical model with SOC included. The excitation energies can then be expanded in the SOC strength λ via QDPT. Consider the configuration interaction single (CIS) model, where the single excitations $|\Phi_i^a(\lambda)\rangle$ arise from a HF reference $|\Phi_0(\lambda)\rangle$. The excitation energies are to be determined by the eigenvalue problem

$$\mathbf{H}(\lambda)\mathbf{X}_\alpha(\lambda) = \omega_\alpha(\lambda)\mathbf{X}_\alpha(\lambda), \tag{177}$$

with

$$[\mathbf{H}(\lambda)]_{ai,bj} = \langle\Phi_i^a(\lambda)|H(\lambda)|\Phi_j^b(\lambda)\rangle = \langle\Phi_i^a|e^{-\kappa(\lambda)}H(\lambda)e^{\kappa(\lambda)}|\Phi_j^b\rangle, \tag{178}$$

where $\kappa(\lambda)$ is the orbital rotation operator (170) determined by the ground-state calculation. Similar to the expansion of the HF energy (160), the CIS Hamiltonian matrix $\mathbf{H}(\lambda)$ can be expanded as

$$\mathbf{H}(\lambda) = \mathbf{H}_0 + \mathbf{H}_1\lambda + \mathbf{H}_2\lambda^2 + \cdots, \tag{179}$$

$$[\mathbf{H}_0]_{ai,bj} = \langle\Phi_i^a|H_0|\Phi_j^b\rangle, \tag{180}$$

$$[\mathbf{H}_1]_{ai,bj} = \langle\Phi_i^a|H_1 + [H_0,\kappa_1]|\Phi_j^b\rangle, \tag{181}$$

$$[\mathbf{H}_2]_{ai,bj} = \langle\Phi_i^a|H_2 + [H_0,\kappa_2] + \frac{1}{2}[H_0;\kappa_1,\kappa_1]|\Phi_j^b\rangle. \tag{182}$$

Since the scalar excited-state states are always degenerate in spin, QDPT should be applied by first decomposing the zeroth-order space into a strongly interacting subspace $\mathcal{P} = \{\mathbf{X}_P\}$ and its orthogonal complement $\mathcal{Q} = \{\mathbf{X}_Q\}$. A unitary transformation $\exp(-\mathbf{S})$ can then be introduced to decouple the two subspaces, viz., the transformed Hamiltonian

$$\bar{\mathbf{H}} = \exp(-\mathbf{S})\mathbf{H}\exp(\mathbf{S}) \tag{183}$$

which is block diagonal and $(\bar{\mathbf{H}})_{PQ} = 0$. The SOC perturbation expansions of $\mathbf{S}$ and $\bar{\mathbf{H}}$ are very similar to the DKH decoupling (73) of the PES and NES and is hence not repeated here. Yet, a few remarks should be made here. First, the interaction between the ground state and excited states vanishes at any order due

to the condition (170). In other words, the spin-orbit corrections to the scalar ground and excited states are carried out separately. However, the information of the scalar ground state is still contained in the coefficients of the spinor excited states through the ground-state orbital rotations κ_k. Second, expanding to all orders will be equivalent to X2C-HF/CIS for the spinor ground/excited states. Third, the scheme works only for a nondegenerate HF reference $|\Phi_0(\lambda)\rangle$, since the condition (171) does not hold for the whole manifold of operators $\{a_p^\dagger a_q\}$ in the case of an open-shell reference.

For the fine structure of an open-shell ground state, the matrix of $H(\lambda)$ can be constructed and diagonalized in its degenerate manifold. This can also be done for a group of energetically adjacent scalar excited states. However, for properties arising from the SOC between the scalar ground and excited states (e.g., phosphorescence), care should be taken when the excitation space is truncated. In view of QDPT, the SOC between the scalar states can be treated by diagonalizing the effective Hamiltonian $\mathbf{H}_{\text{eff}}$:

$$\mathbf{H}_{\text{eff}} = \mathbf{\Omega} + \mathbf{H}_{\text{SO}}, \tag{184}$$

where $\mathbf{\Omega}$ is a diagonal matrix with the scalar excitation energies as the diagonal elements:

$$\mathbf{\Omega} = \text{diag}\{0, \omega_1, \omega_2, \cdots, \omega_N\}. \tag{185}$$

Here the first element is for the ground state. For a degenerate ground state, there are as many zeros as the degree of degeneracy. The matrix elements between two scalar states Ψ_I and Ψ_J (including the ground state) read

$$[\mathbf{H}_{\text{SO}}]_{IJ} = \langle \Psi_I | H_{\text{sd}} | \Psi_J \rangle = \sum_{K,L} C_{K,I} C_{L,J} \langle \Phi_K | H_{\text{sd}} | \Phi_L \rangle, \tag{186}$$

where Φ_K are configuration state functions. In the case of sf-X2C FCI (full configuration interaction), the $\mathbf{H}_{\text{eff}}$ matrix can be diagonalized for the SOC of all the states, as demonstrated in Fig. 1 for the special case of one-electron systems. However, in the case of CIS, the diagonalization of $\mathbf{H}_{\text{eff}}$ will lead to an unbalanced treatment of SOC: The diagonalization accounts for spin-orbital relaxations of the ground state but not of the excited states. The latter can be seen from the second term of Eq. (181), viz., $\langle \Phi_i^a | [H_0, \kappa_1] | \Phi_j^b \rangle = \sum_{pq} (\langle \Phi_i^a | H_0 | \Phi_{jq}^{bp} \rangle - \langle \Phi_{ip}^{aq} | H_0 | \Phi_j^b \rangle)(\kappa_1)_{pq}$. The particular double excitations here with amplitudes $(\kappa_1)_{pq}$ are responsible for spin-orbital relaxations of the excited states but which are missed by the CIS-based state interaction. As a result, the energy of the ground state will be lowered much more significantly than those of the excited states when single excitations from the core orbitals are included [30]. To achieve a balanced treatment of SOC, some selection rule should be invoked for defining the active orbitals appropriate for the states under consideration.

Nonadiabatic Dynamics Involving Different Spin States

Apart from solving the stationary equation

$$H_{\rm el}\Psi = E\Psi, \quad H_{\rm el} = H_{\rm sf} + H_{\rm sd} \tag{187}$$

with the electronic Hamiltonian $H_{\rm el}$ (3), it is also of great interest to investigate nonadiabatic dynamics involving states of different spins beyond the Bohn-Oppenheimer approximation. For this purpose, the time-dependent equation can be exploited

$$\mathbb{i}\partial_t|\Psi[\vec{r},\vec{R};t]\rangle = H|\Psi[\vec{r},\vec{R};t]\rangle, \quad H = H_{\rm el} + T_N = H_{\rm sf} + H_{\rm sd} + T_N, \tag{188}$$

where the variables $\vec{r}$ and $\vec{R}$ represent the electronic and nuclear coordinates, respectively, and T_N is the nuclear kinetic energy operator. Recently, there have been growing interests in developing on-the-fly mixed quantum classical surface hopping (SH) methods for dynamical processes involving different spin states [32–34]. Different representations, spin-adiabatic or spin-diabatic, will lead to different formalisms and also different requirements for electronic structure calculations.

In the spin-adiabatic representation, where the eigenstates of $H_{\rm el}$ are taken as the basis, such that the coupling between different states stems only from the nuclear derivatives, the SH approaches developed for the nonrelativistic regime can be applied directly. The basic idea of SH is that, while the electrons are treated quantum mechanically via the electronic Hamiltonian $H_{\rm el}$, the nuclear motion is described by classical trajectories $\vec{R}(t)$, each of which evolves classically on a potential energy surface (PES) of $H_{\rm el}$, but may switch to another PES at any time with certain probability. That is, instead of solving the full quantum dynamical equation (188), the time-dependent electronic equation of motion

$$\mathbb{i}\partial_t|\Psi[\vec{r};\vec{R}(t),t]\rangle = H_{\rm el}|\Psi[\vec{r};\vec{R}(t),t]\rangle \tag{189}$$

is first solved in a basis spanned by the *instantaneous adiabatic* states $\{|\Psi_i[\vec{r};\vec{R}(t)]\rangle\}_{i=1}^M$ of $H_{\rm el}$, viz.,

$$H_{\rm el}|\Psi_i[\vec{r};\vec{R}(t)]\rangle = E_i[\vec{R}(t)]|\Psi_i[\vec{r};\vec{R}(t)]\rangle, \tag{190}$$

$$|\Psi[\vec{r};\vec{R}(t),t]\rangle = \sum_{i=1}^{M} c_i(t)|\Psi_i[\vec{r};\vec{R}(t)]\rangle. \tag{191}$$

Multiplying Eq. (189) from the left by $\langle\Psi_i[\vec{r};\vec{R}(t)]|$ and integrating over the electronic coordinates, the equation of motion for the linear coefficients $c_i(t)$ can be obtained as

$$\partial_t c_i(t) = -\mathbb{i}E_i[\vec{R}(t)]c_i(t) - \sum_{j=1}^{M} \Lambda_{ij}[\vec{R}(t),t]c_j(t), \tag{192}$$

where the coupling $\Lambda_{ij}[\vec{R}(t),t]$ reads

$$\Lambda_{ij}[\vec{R}(t),t] = \langle\Psi_i[\vec{r};\vec{R}(t)]|\partial_t|\Psi_j[\vec{r};\vec{R}(t)]\rangle_r = \vec{d}_{ij}[\vec{R}(t)]\cdot\vec{v}(t), \tag{193}$$

$$\vec{d}_{ij}[\vec{R}(t)] = \langle\Psi_i[\vec{r};\vec{R}(t)]|\frac{\partial}{\partial\vec{R}(t)}|\Psi_j[\vec{r};\vec{R}(t)]\rangle_r. \tag{194}$$

Here, $\vec{v}(t) = \partial_t\vec{R}(t)$ is the nuclear velocity, while $\vec{d}_{ji}[\vec{R}(t)]$ is the first-order nonadiabatic coupling matrix elements (fo-NACME) between states i and j. The transition probability between the two states in a small time interval Δt can be estimated by the Tully fewest-switch algorithm [35],

$$p_{ij} = \max\{0, \frac{2\Re(c_i^* c_j\,\vec{v}\cdot\vec{d}_{ij})}{|c_i|^2}\Delta t\}. \tag{195}$$

However, such a one-step treatment of SOC is computationally very demanding and is only necessary for very strong SOC. In contrast, the two-step treatment of SOC is computationally more efficient and has wider applications. In particular, the QDPT of SOC (184) can naturally be combined with SH, since the same model space $\{|\Psi_I[\vec{r};\vec{R}(t)]\}_{I=1}^M$ of scalar electronic wave functions (uppercase suffixes) for SOC can also be used for the SH dynamics. Specifically, the eigenvectors $\mathbf{U}_i[\vec{R}(t)]$ of $\mathbf{H}_{\text{eff}}$ (184) are just the coefficients of the adiabatic states (190) of H_{el} in the basis $\{|\Psi_I[\vec{r};\vec{R}(t)]\}_{I=1}^M$,

$$\mathbf{H}_{\text{eff}}[\vec{R}(t)]\mathbf{U}_i[\vec{R}(t)] = \mathbf{U}_i[\vec{R}(t)]\boldsymbol{\omega}_i[\vec{R}(t)], \tag{196}$$

$$|\Psi_i[\vec{r};\vec{R}(t)]\rangle = \sum_{I=1}^{M}|\Psi_I[\vec{r};\vec{R}(t)]\rangle U_{Ii}[\vec{R}(t)]. \tag{197}$$

The energy $E_i[\vec{R}(t)]$ (192) is simply the sum of the spin-free reference energy E_0 and the SOC-including excitation energy ω_i. The fo-NACME (194) can be written as

$$\begin{aligned}\vec{d}_{ij}[\vec{R}(t)] &= \sum_{I,J=1}^{M}(U_{Ii}^*\langle\Psi_I|)\frac{\partial}{\partial\vec{R}}(|\Psi_J\rangle U_{Jj})_r \\ &= \sum_{I,J=1}^{M}U_{Ii}^*\vec{d}_{IJ}U_{Jj} + \sum_{I=1}^{M}U_{Ii}^*\frac{\partial}{\partial\vec{R}}U_{Ij}, \quad \vec{d}_{IJ} = \langle\Psi_I|\frac{\partial}{\partial\vec{R}}|\Psi_J\rangle_r,\end{aligned} \tag{198}$$

where the fo-NACME $\vec{d}_{IJ}$ between the scalar states can be calculated at various levels of methodology [36–40]. The second term of (198) can be calculated as

$$\mathbf{U}_i^\dagger \frac{\partial}{\partial \vec{R}} \mathbf{U}_j = \frac{\mathbf{U}_i^\dagger \left(\frac{\partial}{\partial \vec{R}} \mathbf{H}_{\text{eff}} \right) \mathbf{U}_j}{\omega_j - \omega_i}, \tag{199}$$

$$\frac{\partial}{\partial \vec{R}} \mathbf{H}_{\text{eff}} = \frac{\partial}{\partial \vec{R}} \mathbf{\Omega} + \frac{\partial}{\partial \vec{R}} \mathbf{H}_{\text{SO}}, \tag{200}$$

$$\left(\frac{\partial}{\partial \vec{R}} \mathbf{\Omega} \right)_{IJ} = \delta_{IJ} \frac{\partial}{\partial \vec{R}} \omega_J, \tag{201}$$

$$\left(\frac{\partial}{\partial \vec{R}} \mathbf{H}_{\text{SO}} \right)_{IJ} = \frac{\partial}{\partial \vec{R}} \langle \Psi_I | H_{\text{sd}} | \Psi_J \rangle_r. \tag{202}$$

It is seen that the evaluation of Eq. (199) requires the energy gradients (201) of the scalar states as well as the nuclear derivatives of the SOC matrix elements (202) between two scalar states. The nuclear derivative of H_{sd} (5) involves those of $\mathbf{X}_0$, $\mathbf{R}_{0,\pm}$, and $\mathbf{W}_1$, etc., which can be obtained by differentiating their defining equations (42), (37)/(39), and (80) in exactly the same way as how the spin-orbit operators are derived. The nuclear derivative of $f_{\text{SO},2e}$ also involves the derivatives of the MO coefficients through the density matrix $\mathbf{P}$. However, such dependence does not cause much additional cost for a post-SCF approach. Note in passing that, at variance with such an explicit treatment of the second term in Eq. (198), an implicit treatment is also possible [32] during the time propagation.

Apart from the above spin-adiabatic representation, the spin-diabatic representation can also be used. Here, the wave function $|\Psi[\vec{r}; \vec{R}(t), t]\rangle$ is expanded in terms of the eigenstates of H_{sf},

$$H_{\text{sf}} |\Psi_I[\vec{r}; \vec{R}(t)]\rangle = E_I[\vec{R}(t)] |\Psi_I[\vec{r}; \vec{R}(t)]\rangle, \tag{203}$$

$$|\Psi[\vec{r}; \vec{R}(t), t]\rangle = \sum_{I=1}^{M} c_I(t) |\Psi_I[\vec{r}; \vec{R}(t)]\rangle. \tag{204}$$

As governed by both spin-orbit and derivative couplings, the equation of motion for $c_I(t)$ reads

$$\partial_t c_I(t) = -\mathbb{i} \left(E_I[\vec{R}(t)] c_I(t) + \sum_{J=1} [\mathbf{H}_{\text{SO}}]_{IJ} c_J(t) \right) - \sum_{J=1}^{M} \vec{v} \cdot \vec{d}_{IJ} c_J(t). \tag{205}$$

The relative significance of the spin-orbit and derivative couplings for nonadiabatic transitions can readily be revealed this way. This spin-diabatic representation is manifestly cheaper than the previous spin-adiabatic representation. However, one has to be aware that the SH mechanism is only valid when the nonadiabatic coupling matrix elements are only significant in a small region. This is usually the case for the fo-NACME $\vec{d}_{IJ}$ but not necessarily the case for the SOC $[\mathbf{H}_{\text{SO}}]_{IJ}$. Therefore, the

reliability and accuracy of such approach need to be verified, especially for systems containing heavy elements.

Summary

The spin separation of the algebraic X2C Hamiltonians is basically a perturbation theory based on the a priori partition (26) of the matrix Dirac Hamiltonian into a sum of spin-free and spin-dependent parts. The spin-free part defines the non-expanded sf-X2C Hamiltonian which includes scalar relativity to infinite order, while the spin-dependent part is treated as a perturbation on top of sf-X2C, thereby leading to a series of new two-component Hamiltonians, as shown in Fig. 2. In such a matrix formulation, no new integrals other than those entering the parent Dirac Hamiltonian arise and no singularities arise as well. The perturbation expansion in spin on top of sf-X2C is guaranteed to converge very fast [25, 28]. Because of this, even the lowest-order terms (esp. so-DKH3) are already very accurate. It also deserves to be mentioned that the $\mathcal{O}(\alpha^2)$ spin-orbit corrections to sf-X2C electric and magnetic properties are one-to-one correspondent to but more accurate than the BP spin-orbit corrections to the nonrelativistic ones [28], such that the conceptual and computational advantages of the BP Hamiltonian are all retained in the spin-separated Hamiltonians. In short, the generic spin-separated Hamiltonian (3) satisfies all the features (F1)–(F5) on the wish list. Further combined with the effective quantum electrodynamics Hamiltonian [23, 29], it is justified to claim that the development of relativistic Hamiltonians has been completed. One can just pick up the right Hamiltonian for the target physics and accuracy.

Acknowledgements This work was supported by the NSFC (Project Nos. 21033001, 21273011, and 21290192).

References

1. van Lenthe E, Baerends EJ, Snijders JG (1993) Relativistic regular two-component Hamiltonians. J Chem Phys 99:4597
2. Douglas M, Kroll NM (1974) Quantum electrodynamical corrections to the fine structure of helium. Ann Phys 82:89
3. Hess BA (1986) Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. Phys Rev A 33:3742
4. Hess BA, Marian CM, Wahlgren U, Groppen O (1996) A mean-field spin-orbit method applicable to correlated wavefunctions. Chem Phys Lett 251:365
5. Schimmelpfennig B (1996) Atomic spin-orbit mean-field integral program. University of Stockholm, Sweden
6. Dyall KG (1994) An exact separation of the spin-free and spin-dependent terms of the Dirac–Coulomb–Breit Hamiltonian. J Chem Phys 100:2118
7. Liu W, Peng D (2006) Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory. J Chem Phys 125:044102; 125:149901(E) (2006)
8. Liu W (2010) Ideas of relativistic quantum chemistry. Mol Phys 108:1679

9. Peng D, Liu W, Xiao Y, Cheng L (2007) Making four-and two-component relativistic density functional methods fully equivalent based on the idea of "from atoms to molecule". J Chem Phys 127:104106
10. Visscher L, van Lenthe E (1999) On the distinction between scalar and spin–orbit relativistic effects. Chem Phys Lett 306:357
11. Stanton RE, Havriliak S (1984) Kinetic balance: a partial solution to the problem of variational safety in Dirac calculations. J Chem Phys 81:1910
12. Foldy LL, Wouthuysen SA (1950) On the Dirac theory of spin 1/2 particles and its non-relativistic limit. Phys Rev 78:29
13. Kutzelnigg W (1997) Relativistic one-electron Hamiltonians 'for electrons only' and the variational treatment of the Dirac equation. Chem Phys 225:203
14. Kutzelnigg W, Liu W (2006) Quasirelativistic theory I. Theory in terms of a quasi-relativistic operator. Mol Phys 104:2225
15. Dyall KG (1997) Interfacing relativistic and nonrelativistic methods. I. Normalized elimination of the small component in the modified Dirac equation. J Chem Phys 106:9618
16. Kutzelnigg W, Liu W (2005) Quasirelativistic theory equivalent to fully relativistic theory. J Chem Phys 123:241102
17. The acronym 'X2C' (pronounced as "ecstasy") for exact two-component Hamiltonians was proposed by W. Liu after intensive discussions with H. J. Aa. Jensen, W. Kutzelnigg, T. Saue, L. Visscher during the Twelfth International Conference on the Applications of Density Functional Theory (DFT-2007), Amsterdam, August 26–30, 2007. Note that the 'exact' here emphasizes that all the solutions of the matrix Dirac equation can be reproduced up to machine accuracy. It is particularly meaningful when compared with the approximate two-component (A2C) Hamiltonians
18. Liu W, Peng D (2009) Exact two-component Hamiltonians revisited. J Chem Phys 131:031104
19. Barysz M, Sadlej AJ, Snijders JG (1997) Nonsingular two/one-component relativistic Hamiltonians accurate through arbitrary high order in α^2. Int J Quant Chem 65:225
20. Barysz M, Sadlej AJ (2001) Two-component methods of relativistic quantum chemistry: from the Douglas–Kroll approximation to the exact two-component formalism. J Mol Struct (THEOCHEM) 573:181
21. Reiher M, Wolf A (2004) Exact decoupling of the Dirac Hamiltonian. I. General theory. J Chem Phys 121:2037
22. Reiher M, Wolf A (2004) Exact decoupling of the Dirac Hamiltonian. II. The generalized Douglas–Kroll–Hess transformation up to arbitrary order. J Chem Phys 121:10945
23. Liu W (2014) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59
24. Kutzelnigg W (2002) In: Schwerdtfeger P (ed) Relativistic electronic structure theory. Part 1. Fundamentals. Elsevier, Amsterdam
25. Li Z, Xiao Y, Liu W (2012) On the spin separation of algebraic two-component relativistic Hamiltonians. J Chem Phys 137:154114
26. Liu W, Kutzelnigg W (2007) Quasirelativistic theory. II. Theory at matrix level. J Chem Phys 126:114107
27. Nakajima T, Hirao K (2000) The higher-order Douglas–Kroll transformation. J Chem Phys 113:7786
28. Li Z, Xiao Y, Liu W (2014) On the spin separation of algebraic two-component relativistic Hamiltonians: molecular properties. J Chem Phys 141:054111
29. Liu W (2014) Perspective: relativistic Hamiltonians. Int J Quantum Chem 114:983
30. Li Z, Suo B, Zhang Y, Xiao Y, Liu W (2013) Combining spin-adapted open-shell TD-DFT with spin–orbit coupling. Mol Phys 111:3741
31. Goddard JD, Osamura Y, Schaefer III HF (1994) A new dimension to quantum chemistry: analytic derivative methods in ab initio molecular electronic structure theory. Oxford University Press, New York
32. Richter M, Marquetand P, González-Vázquez J, Sola I, González L (2011) SHARC: ab Initio molecular dynamics with surface hopping in the adiabatic representation including arbitrary couplings. J Chem Theory Comput 7:1253

33. Granucci G, Persico M, Spighi G (2012) Surface hopping trajectory simulations with spin-orbit and dynamical couplings. J Chem Phys 137:22A501
34. Cui G, Thiel W (2014) Generalized trajectory surface-hopping method for internal conversion and intersystem crossing. J Chem Phys 141:124101
35. Tully JC (1990) Molecular dynamics with electronic transitions. J Chem Phys 93:1061
36. Lengsfield III BH, Saxe P, Yarkony DR (1984) On the evaluation of nonadiabatic coupling matrix elements using SA-MCSCF/CI wave functions and analytic gradient methods. I. J Chem Phys 81:4549
37. Send R, Furche F (2010) First-order nonadiabatic couplings from time-dependent hybrid density functional response theory: consistent formalism, implementation, and performance. J Chem Phys 132:044107
38. Tapavicza E, Bellchambers G, Vincent JC, Furche F (2013) Ab initio non-adiabatic molecular dynamics. Phys Chem Chem Phys 15:18336
39. Li Z, Liu W (2014) First-order nonadiabatic coupling matrix elements between excited states: a Lagrangian formulation at the CIS, RPA, TD-HF, and TD-DFT levels. J Chem Phys 141:014110
40. Li Z, Suo B, Liu W (2014) First order nonadiabatic coupling matrix elements between excited states: Implementation and application at the TD-DFT and pp-TDA levels. J Chem Phys 141:244105

Relativistic Effective Core Potentials

14

Michael Dolg

Contents

Introduction 450
General Considerations 452
All-Electron Reference Approach 452
Frozen-Core Approximation 455
Atomic Effective Core Potentials 456
Choice of the Core 457
Nodes or No Nodes? 459
Pseudopotentials 460
Local Form 463
Semilocal Form 463
Nonlocal Form 465
Adjustment 466
Model Potentials 469
Ab Initio Model Potentials 470
Model Core Potentials 471
Molecular Effective Core Potentials 471
Core Polarization Potentials 472
Valence Basis Sets 473
Calibration 474
Summary 476
References 476

Abstract

The basic ideas of the effective core potential approach allowing for valence-only quantum chemical calculations implicitly including the most important relativistic effects are briefly outlined. The model potential and pseudopotential variants

M. Dolg (✉)
Institute for Theoretical Chemistry, University of Cologne, Cologne, Germany
e-mail: m.dolg@uni-koeln.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_5

are described in their forms mostly applied for molecular electronic structure calculations. Effective core polarization potentials allowing to overcome some of the basic approximations underlying the effective core potential approach are also discussed.

Keywords

Effective core potentials • Model potentials • Pseudopotentials • Pseudo-valence orbitals • Core polarization potentials • Frozen-core approximation • Frozen-core errors • Core-valence separation • Breit interaction • Quantum electrodynamics • Valence-only Hamiltonian • Relativistic effects • Computational effort • Error-balanced basis sets • Correlation-consistent basis sets • Electronic structure • Excited states • Calibration • Copper • Roentgenium

Introduction

The effective core potential (ECP) approach is one of the oldest and still one of the most frequently used methods in relativistic quantum chemistry [1, 2]. Following chemical intuition, an atom is partitioned into a core and a valence electron system. The chemically inert core of the atom is considered to be frozen. It is removed from the explicit quantum chemical treatment, and its influence on a valence electron is modeled by an effective Hamiltonian, i.e., the ECP [3]. Thus, basic approximations underlying the ECP approach are the core-valence separation and the frozen-core approximation. On the one hand, ECPs lead to significant savings in the computations, especially for heavier atoms, since compared to an all-electron (AE) treatment only the smaller number of valence electrons has to be described explicitly. On the other hand, the corresponding valence-only model Hamiltonian can be constructed by replacing the relativistic operators with their nonrelativistic analogues, e.g., the nonrelativistic kinetic energy and the nonrelativistic Coulomb interaction, and still accounting for relativistic contributions implicitly by a suitable parametrization of the ECP to relativistic reference data.

The accuracy of a specific ECP thus depends on the accuracy of the underlying relativistic AE Hamiltonian it is designed to model. It also depends to a certain extent on the electronic structure method used to generate the suitable AE reference data for the adjustment of the ECP parameters. Last but not least, it is determined by the size of the core modeled by the ECP as well as by the chosen analytical ansatz for the ECP. In most approaches the ECP for a many-valence electron system is constructed as a sum over effective one-electron operators. Depending on whether the radial nodal structure of the AE valence orbitals is kept or simplified in the valence-only scheme, one works within the model potential (MP) [4] or the pseudopotential (PP) [5] approach.

The formal transformation from valence orbitals with the correct number of radial nodes used at the AE or MP level to the so-called pseudo-valence orbitals of the PP schemes exhibiting less radial nodes allows for the usage of smaller

valence basis sets and thus leads to considerable computational savings. However, the elimination of radial nodes from the valence orbitals also leads to changes in the electron interaction between the valence electrons, which is not in all cases sufficiently compensated by the PP parametrization. Therefore, when using the same core-valence separation, the MP approach is potentially more accurate than the PP approach. Concerning the latter, aside from smaller frozen-core errors, small-core PPs usually exhibit smaller errors due to the pseudo-orbital transformation than large-core ones and are thus preferred for accurate calculations.

When going from an atom to a molecule, it is usually assumed that the atomic contributions behave additively, i.e., the molecular ECP is a superposition of the atomic ECPs. This assumption allows for an atomic parametrization of the ECPs, as well as an atomic optimization of the corresponding valence basis sets. In view of the about 120 elements in the periodic table as well as the various possible choices for the core size and the modeled relativistic Hamiltonian, this assumption is mandatory for efficiently generating consistent sets of ECPs which can be applied for all combinations of elements occurring in molecules.

Corrections to the assumption of superposition of atomic ECPs are nevertheless necessary for large cores, which exhibit in addition to the leading Coulomb repulsion between the cores modeled as point charges also deviations in their Coulomb interaction due to their extended and polarizable electron distributions as well as a noticeable additional repulsion due to the Pauli principle. Typically, the corresponding Pauli repulsion and mutual charge distribution penetration corrections can be approximated by those pairwise interactions between frozen spherical atomic cores, which go beyond their simple Coulomb point charge repulsion [6].

The ECPs further can be combined with so-called core polarization potentials (CPPs), which are effective one- and two-electron operators and allow to correct somewhat for both the frozen-core approximation and the core-valence separation [7]. CPPs thus take into account the static polarizabilities of the atomic cores, as well as their dynamic polarizabilities, i.e., core-valence correlation [8]. In contrast to ECPs, the CPPs for many-valence electron systems are not merely sums of one-electron operators. Moreover, although constructed from atomic contributions, the molecular CPPs are not just superpositions of atomic terms. CPPs can be adjusted to both AE and experimental reference data.

Besides a parametrization based on ab initio AE reference data, the PP approach, usually then applied in combination with CPPs, also offers the possibility to adjust the free parameters in the analytical ansatz for the PP to experimental data [3]. Semiempirical energy-adjusted PPs based on experimental data for one-valence electron atoms and ions were derived for main group elements [9] as well as transition metals in configurations with closed d^{10} shells [6]. Difficulties arise for transition metals with open d and f shells, since due to the difficulties to accurately account for valence electron correlation effects a rigorous semiempirical adjustment is only possible for one-valence electron systems. However, when going from the corresponding one-valence electron ions to the neutral atom, large frozen-core errors arise, especially for elements with open d and f shells. Therefore, due to their more

general applicability to all elements of the periodic table, the current article mainly focuses on the ab initio PP approaches.

The present article will briefly discuss the at present most frequently used ECPs, as well as CPPs, for standard molecular quantum chemical calculations. PPs used for quantum Monte Carlo (QMC) calculations or PPs used in connection with plane wave basis sets for density functional theory (DFT) calculations of solids are not described here. The list of references provided in this article is merely a selection of some representative articles, and as such it is incomplete. Further information on the topic, including the items omitted here, can be found in numerous review articles; cf., e.g., Schwerdtfeger [10] or Cao and Dolg [11] and the references cited therein.

General Considerations

A fundamental question before setting up an ECP approach is which AE Hamiltonian should be modeled by it. Various approximate relativistic many-electron schemes are at hand nowadays to generate suitable AE reference data for the ECP adjustment. In order to avoid any bias by finite basis sets, it is desirable to work with atomic structure codes using a finite difference scheme and to optimize suitable valence basis sets after the ECP has been constructed.

All-Electron Reference Approach

Consider a generic AE Hamiltonian for an atom with n electrons, i.e.,

$$\hat{H} = \sum_{i=1}^{n} \hat{h}(i) + \sum_{i=j+1}^{n} \sum_{j=1}^{n-1} \hat{g}(i,j), \tag{1}$$

where i and j stand for electron indices. Here and in the following, the expression $X(i)$ denotes the dependence of the preceding term X on the position vector $\mathbf{r}_i$, whereas the dependence on only the distance is denoted by $X(r_i)$. Atomic units are used in the equations.

For a relativistic calculation, the one-electron operator $\hat{h}$ might be chosen as the Dirac (D) Hamiltonian

$$\hat{h}_D(i) = c\hat{\boldsymbol{\alpha}}_i \cdot \hat{\mathbf{p}}_i + (\hat{\boldsymbol{\beta}}_i - \mathbf{I}_4)c^2 + \hat{V}(r_i), \tag{2}$$

where the rest energy c^2 of the electron was subtracted in order to get the same zero of energy as in the nonrelativistic case. Here, c stands for the velocity of light ($c \approx 137.0359895$ a.u.). $\mathbf{I}_4$ corresponds to the 4×4 unit matrix. $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ denote the 4×4 Dirac matrices

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} \mathbf{0}_2 & \hat{\boldsymbol{\sigma}} \\ \hat{\boldsymbol{\sigma}} & \mathbf{0}_2 \end{pmatrix} \qquad \text{and} \qquad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & -\mathbf{I}_2 \end{pmatrix}, \tag{3}$$

which can be written in terms of the three-component vector of the 2×2 Pauli matrices $\hat{\boldsymbol{\sigma}}$,

$$\hat{\boldsymbol{\sigma}}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{\boldsymbol{\sigma}}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \hat{\boldsymbol{\sigma}}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{4}$$

the 2×2 unit matrix $\mathbf{I}_2$, and the 2×2 zero matrix $\mathbf{0}_2$. $\hat{\mathbf{p}}_i = -i\hat{\boldsymbol{\nabla}}_i$ is the momentum operator acting on the i-th electron, with the vector differential (del or nabla) operator $\hat{\boldsymbol{\nabla}}_i = (\partial/\partial x_i, \partial/\partial y_i, \partial/\partial z_i)$.

Frequently, instead of the Coulomb point charge model

$$\hat{V}(r) = -\frac{Z}{r} \tag{5}$$

a finite nuclear model is used, e.g., a Gaussian-type charge distribution

$$\rho_\lambda(r) = \rho_{0,\lambda} \exp(-\eta_\lambda r^2) \qquad \text{with} \qquad 4\pi \int_0^\infty dr\, r^2 \rho_\lambda(r) = Z_\lambda. \tag{6}$$

The parameter η_λ can be determined from the nuclear radius R_λ, which is related to the nuclear mass according to

$$\eta_\lambda = 3/(2R_\lambda^2) \qquad \text{with} \qquad R_\lambda = 2.2677 \times 10^{-5} M_\lambda^{1/3}. \tag{7}$$

Other nuclear models, e.g., a hard sphere nucleus or a nucleus with a Fermi-type charge distribution, are also in use. It is important to note that the accuracy of modern ECPs can be so high that the parametrization can also include finite nucleus effects which are noticeable for heavy elements.

For the two-electron terms $\hat{g}$, the simplest choice is the nonrelativistic Coulomb (C) interaction

$$\hat{g}_C(i,j) = \frac{1}{r_{ij}}. \tag{8}$$

Inclusion of the magnetic interaction between the electrons leads to the Gaunt (G) interaction

$$\hat{g}_{CG}(i,j) = \frac{1}{r_{ij}} - \frac{\hat{\boldsymbol{\alpha}}_i \cdot \hat{\boldsymbol{\alpha}}_j}{r_{ij}}, \tag{9}$$

which includes the most important correction of the nonrelativistic Coulomb (C) interaction. Including the retardation of the interaction due to the finite velocity of light leads to the Breit (B) interaction in its low-frequency limit for the exchanged photon

$$\hat{g}_{CB}(i,j) = \frac{1}{r_{ij}} - \frac{1}{2r_{ij}}\left[\hat{\boldsymbol{\alpha}}_i \cdot \hat{\boldsymbol{\alpha}}_j + \frac{(\hat{\boldsymbol{\alpha}}_i \cdot \mathbf{r}_{ij})(\hat{\boldsymbol{\alpha}}_j \cdot \mathbf{r}_{ij})}{r_{ij}^2}\right]. \tag{10}$$

Inserted together with the Dirac Hamiltonian $\hat{h}_D$ (Eq. 2) into the generic Hamiltonian (Eq. 1), these three choices for $\hat{g}$ lead to the Dirac-Coulomb (DC), Dirac-Coulomb-Gaunt (DG), and Dirac-Coulomb-Breit (DCB) Hamiltonian, respectively. For heavy atoms, the latter Hamiltonian has to be augmented to include also low-order effects from quantum electrodynamics (QED), i.e., a frequency-dependent expression of the Breit interaction, the electron self-energy, and the vacuum polarization, leading to the best Hamiltonian which can currently be applied routinely in atomic structure calculations, e.g., using computer codes such as GRASP (general relativistic atomic structure package) [12]. These are usually carried out at the multi-configuration finite difference Dirac-Hartree-Fock (MCDHF) level to obtain suitable reference data for the determination of the ECPs.

When answering the initial question which relativistic Hamiltonian the valence-only approach should model, one has to consider the errors inherent to the ECP approach as well as possible errors due to the use of finite basis sets and the applied computational method in practical calculations. It is well known that the total nonrelativistic energy of an atom roughly increases with the nuclear charge to the second power, whereas relativistic corrections to it increase with the nuclear charge to the fourth power. In contrast to this, for a fixed number of valence electrons, the elements of a group of the periodic table are treated on equal footing within the ECP schemes, and the corresponding ECP errors are roughly the same for all members of the group. Thus, whereas for light elements the DC point nucleus Hamiltonian or approximations to it are usually sufficiently accurate, reference data based on at least the finite nucleus DCB Hamiltonian becomes mandatory for heavy elements. Since nowadays atomic electronic structure codes featuring the DC and DCB Hamiltonians are at hand, modern ECPs use the corresponding AE reference data. Some older, but nevertheless still very popular, sets of ECPs are based on approximate relativistic schemes such as the Wood-Boring [13] or Cowan-Griffin [14] relativistic HF approaches.

Table 1 lists relative energies of selected low-lying electronic states of roentgenium (Rg, eka-Au, $Z = 111$) calculated at the HF and MCDHF level using various forms of the Hamiltonian. The corresponding differential effects, i.e., contributions to the energy differences, are provided in Table 2. It is clear that for an element as heavy as Rg relativistic contributions arising from the Dirac one-particle Hamiltonian Δ_{DC} are huge and have to be included. This is also obvious from the corresponding huge orbital contraction ($7s_{1/2}$ vs. 7s) and noticeable

Table 1 Energies (eV) relative to the ground J level of selected low-lying electronic states of roentgenium (Rg) ($Z = 111$) [12]

Configuration		J	P	HF	DHF/DC			
							+B lfl	+B+QED
				pn	pn	fn	fn	fn
Rg	$d^9\ s^2$	5/2	+	0	0	0	0	0
		3/2	+	0	2.3516	2.3663	2.2975	2.3000
	$d^{10}\ s^1$	1/2	+	−6.2398	2.9685	2.7419	2.6740	2.6107
Rg^-	$d^{10}\ s^2$	0	+	−6.3094	−0.1471	−0.2157	−0.2583	−0.2793
Rg^+	$d^8\ s^2$	4	+	14.5812	9.7898	8.9056	8.9450	8.9670

HF Hartree-Fock (nonrelativistic), *DHF* Dirac-Hartree-Fock (relativistic), *DC* Dirac-Coulomb Hamiltonian, *+B* perturbative treatment of the Breit interaction, *lfl* low-frequency limit, *QED* low-order quantum electrodynamic contributions (self-energy, vacuum polarization), *pn* point nucleus, *fn* finite nucleus

Table 2 Various contributions to energies (eV) relative to the ground J level of selected low-lying electronic states of roentgenium (Rg) ($Z = 111$) [12]

Configuration		J	P	Δ_{DC}	Δ_{fn}	Δ_B	Δ_{QED}
Rg	$d^9\ s^2$	3/2	+	2.3516	0.0147	−0.0688	0.0025
	$d^{10}\ s^1$	1/2	+	9.2083	−0.2266	−0.0679	−0.0633
Rg^-	$d^{10}\ s^2$	0	+	6.1623	−0.0686	−0.0426	−0.0210
Rg^+	$d^8\ s^2$	4	+	−4.7914	−0.8842	0.0394	0.0220

Δ_{DC}, DHF/DC pn vs. HF pn; Δ_{fn}, DHF/DC fn vs. DHF/DC pn; Δ_B, DHF/DC+B lfl fn vs. DHF/DC fn; Δ_{QED}, DHF/DC+B+QED fn vs. DHF/DC+B lfl fn; cf. Table 1 for other abbreviations

expansion ($6d_{5/2}$ vs. 6d) of the valence orbitals depicted in Fig. 1. However, also the contributions due to a finite nuclear model Δ_{fn}, the Breit two-electron interaction Δ_B, and even low-order quantum electrodynamic contributions Δ_{QED} are above the accuracy typically achieved for modern ECP approaches in such energy differences, i.e., 0.01 eV or better.

Frozen-Core Approximation

The most straightforward approach to formally reduce the explicit treatment to the valence electron system is the frozen-core approximation. The valence-only Hamiltonian can be written as

$$\hat{H}_v = \sum_{i=1}^{n_v} \left[\hat{t}(i) + \hat{V}_{cv}(i) \right] + \sum_{i=j+1}^{n_v} \sum_{j=1}^{n_v-1} \hat{g}(i,j). \tag{11}$$

Here n_v denotes the number of explicitly treated valence electrons, which adds up with the number of core electrons n_c to the total number of electrons $n = n_v + n_c$. $\hat{t}$ represents the kinetic energy operator in the one-electron Hamiltonian $\hat{h}$ in Eq. 1,

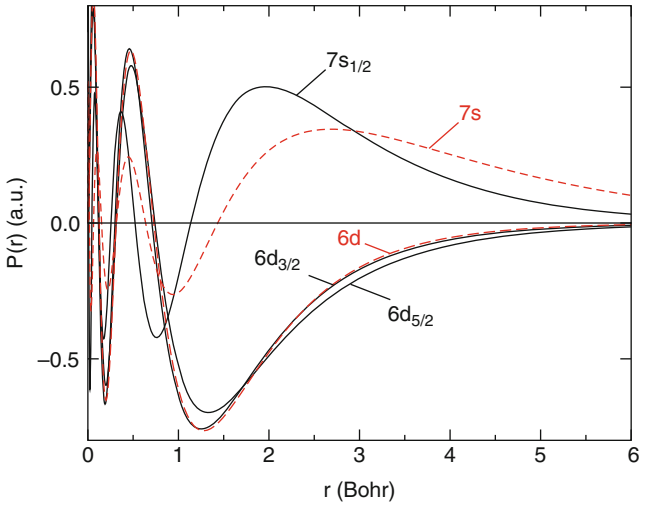


Fig. 1 Large components of the relativistic $6d_{3/2}$, $6d_{5/2}$, and $7s_{1/2}$ MCDHF/DC AE spinors of roentgenium (Rg) ($Z = 111$) in its $6d^9\ 7s^2$ ground state configuration in comparison to corresponding nonrelativistic HF orbitals [12]. For the corresponding orbital energies, cf. Fig. 2

and $\hat{V}_{cv}$ is the nonlocal potential for the interaction of a valence electron with the nucleus and the fixed core electron system. Besides the Coulomb potential arising from the nucleus, the latter contains a sum over Coulomb operators $\hat{J}_k$ and exchange operators $\hat{K}_k$ constructed from the $n_c/2$ doubly occupied core orbitals, i.e., in the case of a point nucleus, it reads

$$\hat{V}_{cv}(i) = -\frac{Z}{r_i} + \sum_{k=1}^{n_c/2} \left[2\hat{J}_k(i) - \hat{K}_k(i) \right]. \tag{12}$$

Clearly, the computational savings are small, if they are present at all, since the core orbitals have to be determined in advance and the full primitive AE basis sets are required to correctly describe the valence orbitals. However, atomic frozen-core AE calculations are a useful tool to assess the potential accuracy of ECPs applying a corresponding core definition.

Atomic Effective Core Potentials

Effective core potentials are constructed to model the DCB finite nucleus Hamiltonian, or some approximation of it, by using a parametrized valence-only Hamiltonian circumventing or at least simplifying the explicit construction of $\hat{V}_{cv}$ from AE solutions. Hereby, a suitable compromise between computational efficiency and accuracy is sought. The valence-only Hamiltonian is thus kept simple, but still offers enough flexibility to compensate the approximations underlying the

ECP approach by a suitable adjustment to AE reference data. In most of the ECP schemes, the nonrelativistic kinetic energy is used in $\hat{h}$, and the nonrelativistic Coulomb interaction between electrons is assumed for $\hat{g}$. In the resulting, formally nonrelativistic valence-only model Hamiltonian

$$\hat{H}_v = -\frac{1}{2}\sum_{i=1}^{n_v} \nabla_i^2 + \sum_{i=j+1}^{n_v} \sum_{j=1}^{n_v-1} \frac{1}{r_{ij}} + \sum_{i}^{n_v} \hat{V}_{cv}(i) \tag{13}$$

the core-electron interaction terms $\hat{V}_{cv}$ have to be chosen so that results of calculations using the underlying AE Hamiltonian are reproduced as closely as possible in calculations using a corresponding method for approximately solving the Schrödinger equation, e.g., by matching results of AE calculations at the DHF level and corresponding ECP calculations at the two-component HF level. The whole machinery of nonrelativistic quantum mechanics can be used if a spin-free form of $\hat{V}_{cv}$ is used. Either spin-orbit (SO) coupling is neglected or averaged out at the level of the AE reference calculations prior to the parametrization of a one-component ECP, or the two-component ECP adjusted to AE reference data including SO effects is averaged after the adjustment. The resulting scalar-relativistic valence-only Hamiltonians might lead to slightly different results.

Choice of the Core

A basic question, besides which Hamiltonian the valence-only approach should model, is how large the atomic core to be replaced by the ECP should be. Larger cores make the calculations less expensive, whereas smaller cores usually allow for a higher accuracy due to smaller frozen-core errors as well as a better transferability from the atom to the molecule. A suitable compromise has to be found for routine calculations. Criteria for assigning shells to the core and valence space, respectively, are orbital energies and the shape of the corresponding radial functions. Some hints for an efficient choice of the core can also be obtained by performing atomic AE frozen-core calculations. It frequently turns out that the chemical core, e.g., the core implied by the ordering of the elements in the periodic system, is not a very accurate choice, and smaller cores are preferable.

A spatial criterium for separating core and valence orbitals based on the extension of the radial functions leads to smaller cores and usually to more accurate valence-only schemes than the energetic criterium based on orbital energies, which favors larger cores. Thus, a separation including all occupied shells of a given main quantum number n, as well as those with higher main quantum numbers, in the valence is favored nowadays. In the case of nd ($n = 3, 4, 5, 6$) transition metals, this corresponds to ns, np, nd, and $(n+1)$s in the valence space, whereas for nf elements ($n = 4$ lanthanides, $n = 5$ actinides) ns, np, nd, nf, $(n + 1)$s, $(n + 1)$p, $(n + 1)$d, and $(n + 2)$s are attributed to the valence space for accurate parametrizations [15]. For heavy main group elements, e.g., those which have one or more filled d shells,

a choice analogous to the one made for the transition metals provides a sufficiently high accuracy.

Table 3 lists frozen-core errors for a small-core (1s – 5f) and medium-core (1s – 6p) approximation for roentgenium (Rg). Only the small-core choice can provide an accuracy of 0.01 eV or better. The medium-core approximation is suggested by the orbital energies plotted in Fig. 2, both at the nonrelativistic and the relativistic level. However, the radial overlap of 6d and the 6s, 6p semicore orbitals is obviously too large to allow high accuracy, and, e.g., noticeable errors in energy differences depending on the 6d occupation number arise. Clearly, the large-core (1s – 6d) option is not possible for Rg due to the relativistically induced d^9s^1 ground state configuration. The differential relativistic effect in the $d^{10}s^1$ - d^9s^1 energy difference amounts to more than 9 eV; cf. Table 2. Figure 2 shows the strong stabilization of 7s and the destabilization of 6d due to dominating direct and indirect relativistic effects, respectively. The corresponding orbital contraction and respective expansion are visualized by the $\langle r \rangle$ expectation values in Fig. 3. The large-core choice is also too inaccurate for the lighter homologue Au despite its $d^{10}s^1$ ground state configuration,

Table 3 Frozen-core errors (eV) in total MCDHF/DC energies relative to the ground J level of selected low-lying electronic states of roentgenium (Rg) ($Z = 111$) using the Rg d^9 s^2 ground state configuration core orbitals [12]

Configuration		J	P	Δ_{14}	Δ_{16}
Rg	d^{10} s^1	1/2	+	0.0053	0.0187
Rg^-	d^{10} s^2	0	+	0.0021	0.0198
Rg^+	d^8 s^2	4	+	0.0023	0.0316

Δ_{14}, 14 core shells frozen (1s – 5f); Δ_{14}, 16 core shells frozen (1s – 6p)

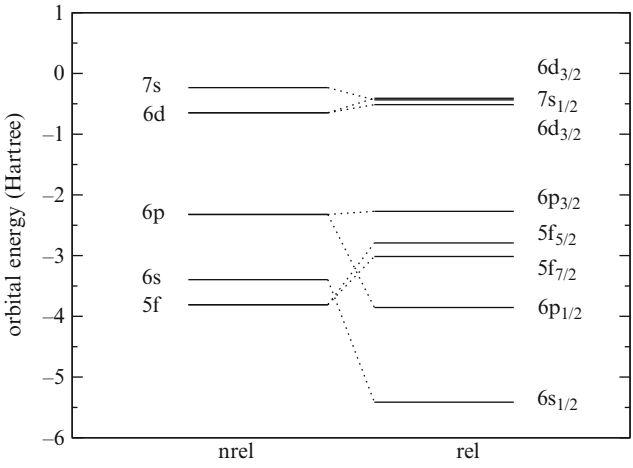


Fig. 2 Orbital energies of roentgenium (Rg) ($Z = 111$) in its $6d^9$ $7s^2$ ground state configuration from nonrelativistic HF and relativistic MCDHF/DC AE calculations [12]

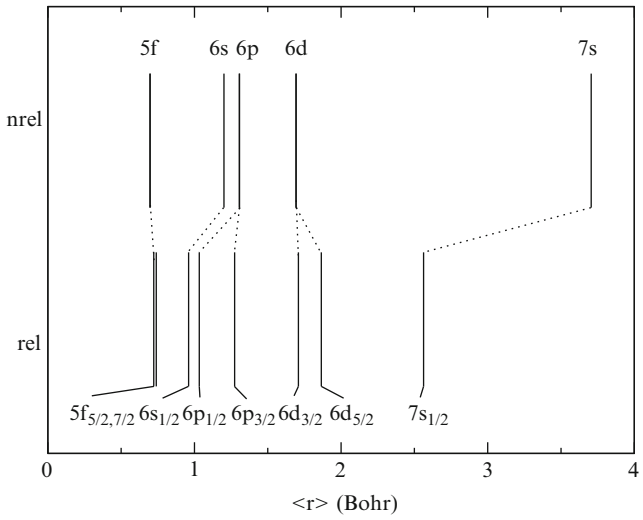


Fig. 3 Radial expectation values $\langle r \rangle$ of roentgenium (Rg) ($Z = 111$) in its $6d^9\ 7s^2$ ground state configuration from nonrelativistic HF and relativistic MCDHF/DC AE calculations [12]

but it works reasonably well for the still lighter members of the group Ag and especially Cu [6].

In special cases it is even possible to define open-shell cores, which correspond to an average over all states arising from the core electron configuration. In the case of lanthanides as well as (heavier) actinides, the f-in-core approximation can be successfully used in electronic structure calculations to treat an atom/ion with a given valency, corresponding to a fixed f occupation number, in molecules [11].

Nodes or No Nodes?

Another important question is if the radial nodal structure of the AE valence orbitals should be kept in the valence-only scheme or if a formal transformation to pseudo-valence orbitals with less radial nodes should be performed. Keeping the radial nodal structure unchanged has the advantage of being able to use unchanged AE operators in the calculations, whereas for pseudo-valence orbitals at least the operators sampling mainly the core region have to be transformed or effective operators have to be constructed. A typical and important example is the SO term of Pauli-type relativistic Hamiltonians, e.g., the Wood-Boring Hamiltonian, which exhibit a Z/r^3 dependence. If such operators are used in calculations with pseudo-valence orbitals, the (effective) nuclear charge Z is merely an adjustable parameter and loses it physical meaning. It also turned out, especially for large cores, that valence correlation energies as well as, e.g., multiplet splittings are more accurate

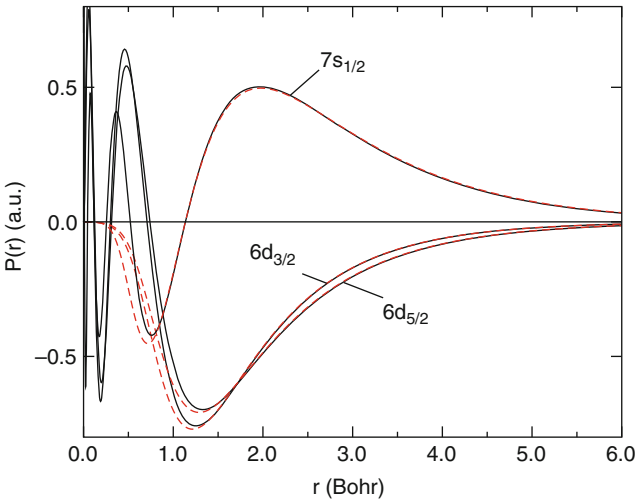


Fig. 4 Large components of the $6d_{3/2}$, $6d_{5/2}$, and $7s_{1/2}$ MCDHF/DC AE spinors of roentgenium (Rg) ($Z = 111$) in its $6d^9\ 7s^2$ ground state configuration in comparison to corresponding pseudo-valence orbitals of a MCDHF/DC+B-adjusted small-core PP (Q = 19) [12, 18]

when the radial nodal structure of the valence orbitals is kept unchanged. This is related to a possible overestimation of exchange integrals, especially between valence orbitals for which a different number of radial nodes was removed [16]. On the other hand, it was found that for energy differences of chemical interest the related errors often cancel [17].

The main advantage of using pseudo-valence orbitals is the reduced basis set requirements, since the oscillations of the valence orbitals in the core region resulting from explicit orthogonality constraints with respect to the core orbitals are replaced by a smooth shape. A comparison between AE valence orbitals [12] and the corresponding pseudo-valence orbitals of an energy-consistent PP [18] is provided for roentgenium in Figs. 4 and 5. As seen from Fig. 4, the pseudo-valence orbitals virtually agree with the corresponding AE valence orbitals in the spatial valence region, e.g., for distances of more than 1.5 Bohr from the nucleus. In the spatial core region, which is emphasized by the logarithmic scale of the r axis in Fig. 5, the pseudo-valence orbitals decay quickly and smoothly when approaching the nucleus. The two options, keeping the number of radial nodes unchanged and reducing it, lead to the model potential (MP) [4] and pseudopotential (PP) [19, 20] approach, respectively.

Pseudopotentials

Pseudopotentials (PPs), in a semiempirical form, were historically the first [3] and, in their ab initio form, are still the most widely used [10] ingredients for the core-electron interaction $\hat{V}_{cv}$ in the valence-only Hamiltonian (Eq. 13) of an atom.

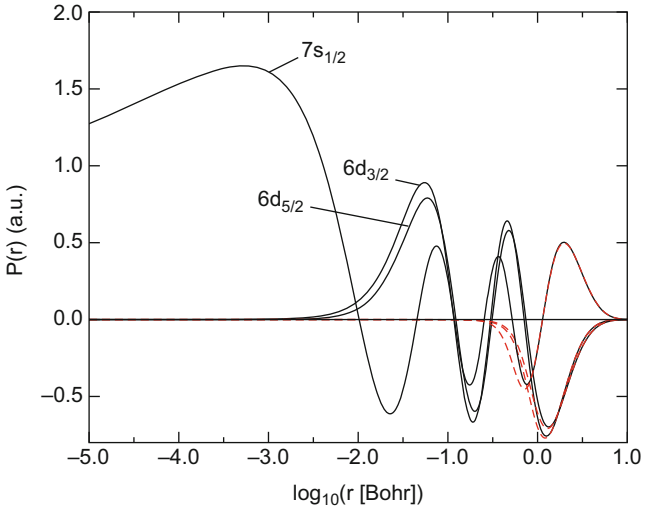


Fig. 5 As in Fig. 4, but for a logarithmic scale of the radius r

The theoretical foundation for the PP approach was set in 1959 by Phillips and Kleinman within an effective one-valence electron framework [5]. A generalization to many-valence electron cases was provided in 1968 by Weeks and Rice [21]. In the so-called generalized Phillips-Kleinman (GPK) equation, the valence electron Hamiltonian $\hat{H}_v$ is supplemented by the GPK PP $\hat{V}^{GPK}$,

$$\left(\hat{H}_v + \hat{V}^{\mathrm{GPK}}\right)|\Phi_p\rangle = E_v|\Phi_p\rangle. \tag{14}$$

Here, E_v denotes the total valence energy, and $|\Phi_p\rangle$ is a many-electron pseudo-valence eigenfunction. The term pseudo used here means that, e.g., Φ_p is built from orbitals which may have a different radial nodal structure than the AE valence orbitals. The original AE valence eigenfunction Φ_v is assumed to fulfill the following Schrödinger equation:

$$\hat{H}_v|\Phi_v\rangle = E_v|\Phi_v\rangle. \tag{15}$$

Φ_v and Φ_p are related by a projection operator $1 - \hat{P}_c$, which projects out any core components in the wave function it acts on, i.e.,

$$|\Phi_v\rangle = (1 - \hat{P}_c)|\Phi_p\rangle. \tag{16}$$

By substituting Φ_v in Eq. 15 by this expression, acting from the left with $1 - \hat{P}_c$, exploiting the idempotency of $\hat{P}_c$, and comparing to Eq. 14, one arrives the following expression for the GPK PP

$$\hat{V}^{\mathrm{GPK}} = -\hat{H}_v\hat{P}_c - \hat{P}_c\hat{H}_v + \hat{P}_c\hat{H}_v\hat{P}_c + E_v\hat{P}_c, \tag{17}$$

which is a nonlocal, energy-dependent many-electron operator still requiring the notion of the solutions for the core electron system.

Eqs. 14, 15, 16, and 17 are by no means easier to solve than corresponding frozen-core HF AE equations. They demonstrate, however, that the correct total valence energy can be obtained by a valence electron eigenfunction which has a simplified form, e.g., a simpler radial nodal structure of the underlying orbitals. An explicit orthogonality of the core and valence electron wave functions is not required anymore. This can be more easily seen by looking at the case of a single valence electron outside a closed-shell core treated at the HF level. The valence Hamiltonian $\hat{H}_v$ then reduces to a Fock operator $\hat{F}$; the AE and pseudo-valence wave functions Φ_v and Φ_p then correspond to the AE valence and PP pseudo-valence orbitals φ_v and φ_p, respectively. The total valence energy E_v is replaced by the valence orbital energy ϵ_v. Assuming that the core orbitals φ_c are also eigenfunctions of the Fock operator $\hat{F}$ with eigenvalues ϵ_c, one can set up the so-called Phillips-Kleinman (PK) equation

$$\left(\hat{F} + \sum_c (\epsilon_v - \epsilon_c)|\varphi_c\rangle\langle\varphi_c|\right)|\varphi_p\rangle = \epsilon_v|\varphi_p\rangle. \tag{18}$$

Any pseudo-valence orbital φ_p which is a mixture of the original AE valence orbital φ_v and the core orbitals φ_c satisfies this equation and yields the correct eigenvalue ϵ_v. One can now seek solutions for which the energetically lowest one is free of radial nodes and thus requires only a reduced basis set for an accurate description.

However, for practical calculations, both the PK and the GPK equations are of no use, since in both cases the core-type solutions also have to be known. Moreover, restricting the pseudo-valence orbitals to be linear combinations of the original valence orbitals and the core orbitals leads to too compact radial functions and to related errors in molecular calculations, e.g., too short bond distances and too high binding energies. Therefore, for practical applications, further simplifications are necessary, and, in principle, the (formal) admixture of virtual orbitals has also to be allowed when building the pseudo-valence orbitals.

Essentially all PP approaches neglect the many-electron character and the energy dependence of the PK and GPK potentials, i.e., they use sums of effective one-electron Hamiltonians. The leading term of the core-valence interaction is the Coulomb interaction between the core, assumed as a point charge $Q = Z - n_c$, and the valence electron i

$$\hat{V}_{cv}^{PP}(i) = -\frac{Q}{r_i} + \Delta\hat{V}_{cv}(i). \tag{19}$$

The analytical forms of the correction $\Delta\hat{V}_{cv}$ are discussed in the following.

Local Form

Historically, a local potential of Yukawa type was used by Hellmann [3] in his 1935 study of the K atom:

$$\Delta\hat{V}_{cv}(i) = \frac{2.74}{r} e^{-1.16r}. \tag{20}$$

One disadvantage of the local ansatz is, e.g., that the valence energies for a given pseudo-main quantum number $\tilde{n}$ increase with the angular momentum quantum number l. This ordering is not always fulfilled, e.g., the (n−1)d^1 ^{2}D states of Ca^+, Sr^+, and Ba^+ are found experimentally below the np^1 ^{2}P states.

Semilocal Form

For the deviation from the Coulomb potential, a semilocal form is used in modern PPs [20], e.g., for scalar-relativistic calculations, the ansatz is

$$\Delta\hat{V}_{cv}(i) = \sum_{l=0}^{l=\infty} V_l(r_i)\hat{P}_l(i). \tag{21}$$

Here, $\hat{P}_l$ denotes an angular momentum projection operator based on spherical harmonics $|lm\rangle$

$$\hat{P}_l(i) = \sum_{m=-l}^{m=l} |lm\rangle\langle lm|. \tag{22}$$

Usually, the radial potentials V_l are very similar for angular momenta which are higher than those present in the core. Thus, a simpler approximate ansatz can be used, i.e.,

$$\Delta\hat{V}_{cv}(i) \cong V_L(r_i) + \sum_{l=0}^{L-1}[V_l(r_i) - V_L(r_i)]\hat{P}_l(i), \tag{23}$$

where $L-1$ denotes the largest angular momentum of any of the core orbitals.

Relativistic calculations including SO coupling require a modified semilocal ansatz [22, 23] such as

$$\Delta\hat{V}_{cv}(i) = \sum_{l=0}^{l=\infty} \sum_{j=|l-\frac{1}{2}|}^{j=l+\frac{1}{2}} V_{lj}(r_i)\hat{P}_{lj}(i), \tag{24}$$

where now $\hat{P}_{lj}$ denotes the projector on spinor spherical harmonics $|ljm>$

$$\hat{P}_{lj}(i) = \sum_{m=-j}^{m=j} |ljm\rangle\langle ljm|. \tag{25}$$

Again, the summation over angular symmetry may be truncated at L, when $L-1$ stands for the highest angular quantum number present in the core, i.e.,

$$\Delta\hat{V}_{cv}(i) \cong V_L(r_i) + \sum_{l=0}^{l=L-1} \sum_{j=|l-\frac{1}{2}|}^{j=l+\frac{1}{2}} [V_{lj}(r_i) - V_L(r_i)]\hat{P}_{lj}(i). \tag{26}$$

SO coupling usually lowers the symmetry and therefore makes quantum chemical investigations more expensive. It is thus advisable to use SO coupling as late as possible in the calculations. The options are, e.g., to perform a SO-CI after standard SO averaged configuration interaction (CI) or coupled cluster (CC) calculations for the interacting states in the basis of correlated scalar-relativistic (LS or ΛS) states, to perform the SO-CI in the basis of determinants and thus include electron correlation and SO coupling on the same footing, or to include it from the very beginning in the SCF process. The latter one-step approach is closest to modeling the underlying two- or four-component Hamiltonian, but it is also more costly than the two previous two-step procedures.

For the two-step procedures in which SO coupling is treated after the scalar-relativistic HF or even CI level, the relativistic PP in Eqs. 24 or 26 is splitted up in a spin-free (spin-dependent terms neglected or averaged, av) and a spin-dependent (spin-orbit, so) part

$$\Delta\hat{V}_{cv}(i) = \Delta\hat{V}_{av}(i) + \Delta\hat{V}_{so}(i). \tag{27}$$

The one-component PP $\Delta\hat{V}_{av}$ can be obtained from the two-component one by applying the relations for the projection operators

$$\hat{P}_l(i) = \sum_j \hat{P}_{lj}(i) = \hat{P}_{l,|l-\frac{1}{2}|}(i) + \hat{P}_{l,l+\frac{1}{2}}(i) \tag{28}$$

and radial potentials

$$V_l(r_i) = \frac{1}{2l+1}[l\,V_{l,|l-\frac{1}{2}|}(r_i) + (l+1)V_{l,l+\frac{1}{2}}(r_i)]. \tag{29}$$

The scalar-relativistic PP $\Delta\hat{V}_{av}$ formally corresponds to the PPs defined in Eqs. 21 or 23. The corresponding SO term reads as

$$\Delta\hat{V}_{so}(i) = \sum_{l=1}^{L-1} \frac{\Delta V_l(r_i)}{2l+1}[l\,\hat{P}_{l,l+1/2}(i) - (l+1)\,\hat{P}_{l,l-1/2}(i)]. \tag{30}$$

Here, ΔV_l is the difference between the relativistic radial potentials for the corresponding l value

$$\Delta V_l(r_i) = V_{l,l+\frac{1}{2}}(r_i) - V_{l,|l-\frac{1}{2}|}(r_i). \tag{31}$$

A simpler form of the SO PP $\Delta\hat{V}_{so}$, which is especially suited for use in SO-CI calculations following a scalar-relativistic HF solution, was proposed by Pitzer and Winter [24]:

$$\Delta\hat{V}_{so}(i) = \sum_{l=1}^{L-1} \frac{2\Delta V_l(r_i)}{2l+1} \hat{P}_l(i)\hat{\mathbf{l}}_i \cdot \hat{\mathbf{s}}_i \hat{P}_l(i). \tag{32}$$

An alternative way to include SO coupling in PP calculations is the usage of a Pauli-type SO operator

$$\Delta\hat{V}_{so}(i) = \frac{\alpha^2}{2} \frac{Z_{\text{eff}}}{r_i^3} \hat{\mathbf{l}}_i \cdot \hat{\mathbf{s}}_i. \tag{33}$$

In contrast to Eq. 32, this form of the SO term is not variationally stable and can only be used in perturbation theory. Moreover, due to the altered shape of the pseudo-valence orbitals in the core region, the parameter Z_{eff} takes values which are far from those of an effective nuclear charge.

The angular parts of the ECPs given in Eqs. 23 and 26 are fixed in the semilocal ansatz. The radial potentials however have to be parametrized. In order to be able to perform calculations on multicenter systems, where mainly Gaussian basis sets are applied, the radial potentials are usually represented by linear combinations of Gaussian functions multiplied by powers of r

$$V_m(r_i) = \sum_k A_{km} r_i^{n_{km}} e^{-a_{km} r_i^2} \qquad \text{with} \qquad m = l, lj, L. \tag{34}$$

The free parameters A_{km}, a_{km} have to be determined by a suitable adjustment to AE reference data, as will be discussed in section "Adjustment". For integrability, $n_{km} \geq -2$ is required. Integrals over semilocal PPs as well as the corresponding SO terms for Cartesian Gaussian basis functions were derived, e.g., by McMurchie and Davidson [25] and Pitzer and Winter [24, 26].

Nonlocal Form

The integral evaluation over semilocal pseudopotential operators is quite involved and may become quite costly if many centers in the system are bearing PPs. Pélissier

et al. therefore proposed a transformation of the atomic semilocal part $V_l \hat{P}_l$ of the PPs in Eqs. 21 and 23 into a nonlocal form [27]

$$\hat{\tilde{V}}_l(i) = \sum_j \sum_k A_{jk} |g_j(i)\rangle \langle g_k(i)|. \tag{35}$$

Here, the $|g_j\rangle$ are orthonormalized linear combinations of Cartesian Gaussian functions

$$\langle \mathbf{r}|g_j\rangle = \sum_k B_{jk} x^{l_{jx}} y^{l_{jy}} z^{l_{jz}} e^{-b_{jk} r^2} \qquad \text{with} \qquad l_{jx} + l_{jy} + l_{jz} = l. \tag{36}$$

The parameters in Eq. 35 can be determined by minimizing the sum S of squared differences of matrix elements over $\hat{\tilde{V}}_l$ and V_l for a large atom-centered basis set $\{|\chi_j\}\rangle$ of appropriate angular symmetry l

$$S = \sum_{jk} \left(\langle \chi_j | \hat{\tilde{V}}_l | \chi_k \rangle - \langle \chi_j | V_l | \chi_k \rangle \right)^2. \tag{37}$$

The nonlocal operator $\hat{\tilde{V}}_l$ can further be cast into a simpler form by diagonalizing the matrix of the coefficients A_{jk}:

$$\hat{\tilde{V}}_l(i) = \sum_j \bar{A}_j |\bar{g}_j(i)\rangle \langle \bar{g}_j(i)|. \tag{38}$$

Here, $\bar{A}_j$ and $|\bar{g}_j\rangle$ denote the resulting eigenvalues and eigenvectors of the $[A_{jk}]$ matrix in the original $\{|g_j\rangle\}$ basis. The integral evaluation in molecular calculations is thus reduced to the calculation of overlap integrals between the molecular basis and the expansion basis $\{|\bar{g}\rangle\}$. Similarly, the evaluation of derivatives of PP matrix elements with respect to the nuclear coordinates, which are needed in energy gradients for geometry optimizations, involves only the computation of derivatives of these overlap integrals.

Adjustment

For the adjustment of the free parameters in the radial PPs of the semilocal ansatz, two approaches differing in the type of AE reference data are in use today, i.e., the shape-consistent approach based on one-particle energies and wave functions (e.g., orbital energies and orbitals or spinor energies and spinors) or the energy-consistent approach relying exclusively on total valence energies. Thus, whereas the shape-consistent approach makes use of quantities defined only within an

effective one-particle picture, the energy-consistent approach is based on quantum mechanically observable energy differences, i.e., excitation energies, ionization potentials, and electron affinities, and makes no use of wave function information.

Shape-Consistent Approach

The main idea of the shape-consistent approach [28, 29] is to select one suitable atomic reference state, to keep the orbital energy ϵ_p of the PP calculation at the value of the AE reference calculation ϵ_v and to reproduce with the pseudo-valence orbital $\varphi_{p,lj}$ exactly the radial shape of the AE valence orbital $\varphi_{v,lj}$ in the spatial valence region, e.g., outside a reasonably chosen critical radius ($r \geq r_c$). In the spatial core region ($r < r_c$), the pseudo-valence orbital is represented by an auxiliary function f_{lj}, i.e.,

$$\varphi_{p,lj}(r) = \begin{cases} \varphi_{v,lj}(r) \text{ for } r \geq r_c \\ f_{lj}(r) \text{ for } r < r_c \end{cases} \qquad \text{and} \qquad \epsilon_{p,lj} = \epsilon_{v,lj}. \tag{39}$$

Usually, a polynomial is used for f_{lj}, which is required to be radially nodeless and smooth, fulfills certain matching conditions at r_c, and leads to a normalized pseudo-valence orbital. In the case of reference data taken from, e.g., MCDHF/DC calculations, the renormalized upper (large) components of the spinors are used for $\varphi_{v,lj}$. The lower (small) components influence the radial density mainly in the core region, which is modified by the usage of f_{lj} anyhow, and can thus be neglected.

Once the pseudo-valence orbital $\varphi_{p,lj}$ has been constructed, the radial part $V_{lj}(r)$ of the corresponding semilocal PP $\Delta\hat{V}_{cv}$ can be determined for each lj-value from the radial Fock equation for the chosen atomic reference state:

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V_{lj}^{PP}(r) + \hat{W}_{p,lj}(\{\varphi_{p',l'j'}\})\right]\varphi_{p,lj}(r) = \epsilon_{v,lj}\varphi_{p,lj}(r). \tag{40}$$

Here, the radial kinetic energy operator is represented by the first two terms in the square brackets, whereas the last term $\hat{W}_{p,lj}$ stands for the effective valence Coulomb and exchange potential acting on $\varphi_{p,lj}$. The radial potential V_{lj} can be determined pointwise on a grid by inversion. It is usually approximated by a linear combination of Gaussian functions, possibly multiplied by powers of r; cf. Eq. 34. By repeating this procedure for each lj-combination, the complete PP in semilocal form can be constructed. The construction of scalar-relativistic PPs and their radial potentials V_l, e.g., based on Cowan-Griffin [14] HF reference data, is performed in analogous fashion using the AE valence orbital $\varphi_{v,l}$ and its orbital energy $\epsilon_{v,l}$.

In order to generate compact Gaussian expansions for the radial potentials V_l or V_{lj}, Durand and Barthelat proposed to minimize the following operator norm [30, 31]:

$$||\hat{\mathcal{O}}|| = \left\langle \varphi_p | \hat{\mathcal{O}}^2 | \varphi_p \right\rangle^{1/2} \quad \text{with} \quad \hat{\mathcal{O}} = \tilde{\epsilon}_v |\tilde{\varphi}_p\rangle\langle\tilde{\varphi}_p| - \epsilon_v |\varphi_p\rangle\langle\varphi_p|. \tag{41}$$

Here, the quantities without tilde correspond to the exact V_l or V_{lj} from a radial Fock equation such as Eq. 40, whereas those with tilde are calculated with an analytical potential $\tilde{V}_l$ or $\tilde{V}_{lj}$.

One advantage of the shape-consistent approach is that (in optimal cases) only one converged reference state is needed to derive a PP. Thus, even derivations of 0-valence electron potential modeling, e.g., rare gas elements, are possible. On the other hand, the restriction to one reference state may also lead to a bias of the PP between this state and others which might be equally important. Here, a suitable averaging might help. Another drawback is the requirement that the pseudo-valence orbitals in Eq. 39 have to be nodeless due to the necessary inversion of the radial Fock Eq. 40. Thus, the adjustment sometimes has to be performed for highly charged ions, which might lead to noticeable frozen-core errors when the PP is applied for the neutral atom or lower charged ions. Prescriptions to overcome this restriction were advised in the generalized relativistic effective core potential (GRECP) ansatz [32]. However, due to the necessary extension of the semilocal PP by nonlocal terms, the approach leads to a more complicated valence-only Hamiltonian and is thus not widely used.

The most popular shape-consistent PP sets using the semilocal ansatz, which is applicable for most standard quantum chemistry codes, are those of Hay and Wadt [33], Christiansen and collaborators [34], and Stevens, Krauss, and coworkers [35]. Whereas the first set is based on scalar-relativistic Cowan-Griffin AE HF calculations, the latter two sets use DHF/DC AE reference data. The set of Christiansen and collaborators covers essentially all elements of the periodic table. For a complete list of references to these PP sets, cf. Ref. [11].

Energy-Consistent Approach

In the energy-consistent approach [10, 15], first an analytical ansatz for the radial potentials V_l or V_{lj} in the PP is made. The free parameters therein are then adjusted to reproduce best the total valence energies of a multitude of electronic states of the neutral atom and its (not too highly charged) ions. This is usually achieved by minimizing the following quantity:

$$S = \sum_I w_I (E_I^{PP} - E_I^{AE} + \Delta E_{\text{shift}})^2 := \text{min.} \tag{42}$$

Here, E_I^{AE} and E_I^{PP} represent the total valence energies for a state I at the AE and PP level. The weights w_I could in principle be used to increase the accuracy for states of special interest, but they are usually set to one or to values leading to equal weights for all configurations included in the fit. The global shift ΔE_{shift} amounts usually to a small fraction of the total valence energies. It allows for each state I a systematic deviation of the total PP valence energy from the AE reference value (e.g., the sum of all valence ionization potentials for the ground state), but usually leads to much better energy differences between the states included in the adjustment (e.g., the lower ionization potentials, excitation energies, and the electron affinity).

A disadvantage of the energy-consistent approach is that a quite large number of states, which are partly not necessarily of direct physical interest, have to be included in the adjustment. For example, in order to describe semicore shells well, energetically high-lying states with excitations and ionizations from these have to be included in the reference data set. Sometimes, the reference data set has to be unwillingly restricted due to convergence difficulties. On the other hand, one advantage is that a large number of states of interest can be described in a balanced way and that the reference energies might also include contributions which are not treated self-consistently at the AE reference level, e.g., the Breit interaction or QED contributions.

Energy-consistent PPs adjusted to Wood-Boring AE HF reference data are available for most main groups as well as d- and f-transition metals. More recent parametrizations are based on MCDHF/DC+B or even MCDHF/DC+B+QED AE reference data and are available for heavy main group elements [36], d-transition metals [37, 38], some actinides [15], as well as some superheavy elements [39]. For a complete list of references to these PP sets, cf. Ref. [11].

Model Potentials

Model potentials (MPs) [4], in contrast to pseudopotentials, retain the correct radial nodal structure of the AE valence orbitals. They are based on the Huzinaga-Cantu equation [40]

$$\left(\hat{F} + \sum_c (-2\epsilon_c)\langle\varphi_c||\varphi_c\rangle\right)|\varphi_v\rangle = \epsilon_v|\varphi_v\rangle \tag{43}$$

and bridge the gap between PPs and AE frozen-core calculations. A shift operator added to the Fock operator $\hat{F}$, i.e., the second term in parentheses in Eq. 43, moves the energies of the core orbitals upward by $-2\epsilon_c$, so that a desired valence orbital is the lowest energy solution. Hereby, unlike in the PP approach where core and virtual orbitals are admixed to the valence orbital to yield a nodeless pseudo-valence orbital, the MP valence orbital keeps the nodal structure of the corresponding AE valence orbital. It is obvious that the primitive basis sets used in MP calculations have to be larger than those of PP calculations.

The interaction between a valence electron i and the atomic core is written in the MP approach as

$$\hat{V}_{cv}^{\mathrm{MP}}(i) = -\frac{Q}{r_i} + \Delta\hat{V}_C(i) + \Delta\hat{V}_X(i) + \hat{P}(i). \tag{44}$$

Here, $\Delta\hat{V}_C(i)$ is the Coulomb (C) interaction between the core electrons and the valence electron

$$\Delta\hat{V}_C(i) = -\frac{n_c}{r_i} + 2\sum_c \hat{J}_c(i) \tag{45}$$

and $\Delta\hat{V}_X(i)$ stands for the corresponding exchange (X) interaction

$$\Delta\hat{V}_X(i) = -\sum_c \hat{K}_c(i). \tag{46}$$

$\hat{J}_c(i)$ and $\hat{K}_c(i)$ denote for the Coulomb and exchange operators for the core orbital $|\varphi_c\rangle$. The shift operator

$$\hat{P}(i) = \sum_c (-2\epsilon_c)|\varphi_c(i)\rangle\langle\varphi_c(i)| \tag{47}$$

prevents the MP valence orbitals to collapse into core-like solutions.

Ab Initio Model Potentials

One of the most successful variants of the MP approach is the ab initio model potentials (AIMPs) developed by Seijo, Barandiarán, and coworkers [41]. The Coulomb interaction of a valence electron and the core is represented by a local spherically symmetric potential

$$\Delta\hat{V}_C(i) = \Delta V_C(r_i) = \frac{1}{r_i}\sum_k C_k e^{-\alpha_k r_i^2}. \tag{48}$$

A least-squares fit to the AE potential is used to determine the parameters α_k, C_k under the constraint $\sum_k C_k = Z - Q = n_c$ enforcing the correct asymptotic behavior. A spectral representation is used for the nonlocal exchange part

$$\Delta\hat{V}_X(i) = \sum_{p,q} |\chi_p(i)\rangle A_{pq} \langle\chi_q(i)|. \tag{49}$$

This MP operator yields the same Coulomb and exchange (one-center) integrals as the AE reference calculation, provided that sufficiently accurate expansions are used in Eqs. 48 and 49. The shift operator in Eq. 47 is constructed with core orbitals represented in sufficiently large AE basis sets.

Scalar-relativistic effects are taken into account implicitly. Using a Cowan-Griffin [14] scalar-relativistic AE reference calculation, the parameters in the Coulomb term (Eq. 45) and the shift operator (Eq. 47) are modified accordingly. In addition, the spectral representation of the mass-velocity and Darwin terms are added to the nonlocal representation of the exchange part (Eq. 49).

In order to take also SO coupling into account, an effective one-electron operator similar to the form proposed by Pitzer and Winter [24] for PPs is added [42]:

$$\Delta\hat{V}_{cv,so}(i) = \sum_l \left(\sum_k \frac{B_{lk}}{r_i^2} e^{-\beta_{lk} r_i^2} \right) \hat{P}_l(i) \hat{\mathbf{l}}_i \cdot \hat{\mathbf{s}}_i \hat{P}_l(i). \tag{50}$$

Here, $\hat{\mathbf{l}}_i = \hat{\mathbf{r}}_i \times \hat{\mathbf{p}}_i$ and $\hat{\mathbf{s}}_i$ denote the operators of orbital angular momentum and spin, respectively. The coefficients B_{lk} and exponents β_{lk} are determined by means of a least-squares fit to the radial components of the Wood-Boring [13] AE SO term.

Model Core Potentials

The model core potential (MCP) approach advocated by Klobukowski and coworkers [43, 44] uses a simpler, entirely local approximation for the Coulomb and exchange terms than the AIMP approach, i.e.,

$$\Delta\hat{V}_C(i) + \Delta\hat{V}_X(i) = -\frac{Q}{r_i} \sum_k A_k r_i^{n_k} e^{-\alpha_k r_i^2}. \tag{51}$$

The values of n_k are restricted to 0 and 1. The model potential parameters A_k and α_k are determined by a simultaneous fit to orbital energies and corresponding radial functions for a given reference state.

Molecular Effective Core Potentials

The molecular valence-only Hamiltonian for a system with N atoms

$$\hat{H}_v = -\frac{1}{2} \sum_{i=1}^{n_v} \nabla_i^2 + \sum_{i=j+1}^{n_v} \sum_{j=1}^{n_v-1} \frac{1}{r_{ij}} + \sum_i^{n_v} \hat{V}_{cv}(i) + V_{cc} \tag{52}$$

is usually constructed by using a superposition of the atomic potentials

$$\hat{V}_{cv}(i) = \sum_{I=1}^{N} \hat{V}_{cv}^I(i). \tag{53}$$

Here and in the following, I and J stand for nuclear indices. The interaction between the nuclei and cores I, J

$$V_{cc} = \sum_{J=I+1}^{N} \sum_{I=1}^{N-1} \left[\frac{Q_I Q_J}{r_{IJ}} + \Delta V_{cc}^{IJ}(r_{IJ}) \right] \tag{54}$$

has as leading term the Coulomb interaction between the point charges Q_I and Q_J, which equal the nuclear charges or the core charges depending on whether the centers are treated at the AE or ECP level. The pairwise corrections $\Delta V_{cc}^{IJ}(r_{IJ})$ describe deviations from the Coulomb repulsion of the point charges, e.g., for mutually penetrating cores, where besides modified electrostatic contributions also orthogonality constraints and the Pauli repulsion between the electron shells localized on different cores have to be considered.

For large cores, the point charge approximation used in Eq. 54 is not sufficiently accurate, and core-core/nucleus repulsion corrections have to be added, e.g.,

$$\Delta V_{cc}^{IJ}(r_{IJ}) = B_{IJ} e^{-b_{IJ} r_{IJ}}. \tag{55}$$

The parameters B_{IJ}, b_{IJ} can be fitted to the deviations of the core-core and/or core-nucleus interactions from the point charge model as obtained from AE HF or DHF calculations for pairs of interacting frozen cores [6].

Core Polarization Potentials

For large, easily polarizable cores, a core polarization potential (CPP) $\hat{V}_{\text{cpp}}$ can be added to the Hamiltonians (Eqs. 13 and 52) in order to account for static core polarization effects occurring already at the HF level as well as for dynamic core polarization effects related to core-valence correlation, i.e.,

$$\hat{V}_{\text{cpp}} = -\frac{1}{2} \sum_I \alpha_I \hat{\mathbf{f}}_I^2. \tag{56}$$

Here, α_I is the static dipole polarizability of the core I, and $\hat{\mathbf{f}}_I$ denotes the electric field produced by all valence electrons as well as all other cores and nuclei at the core I

$$\hat{\mathbf{f}}_I = -\sum_i \frac{\mathbf{r}_{iI}}{r_{iI}^3} + \sum_{J \neq I} Q_J \frac{\mathbf{r}_{JI}}{r_{JI}^3}. \tag{57}$$

For short distances between the polarizing particles and the polarized core, problems arise due to the breakdown of this expression. The electric field is therefore multiplied by a cutoff function F, i.e.,

$$\hat{\mathbf{f}}_I = -\sum_i \frac{\mathbf{r}_{iI}}{r_{iI}^3} F(r_{iI}, \delta_e^I) + \sum_{J \neq I} Q_J \frac{\mathbf{r}_{JI}}{r_{JI}^3} F(r_{JI}, \delta_c^I). \tag{58}$$

A suitable ansatz for F was proposed by Meyer and coworkers [8] in the framework of AE calculations and was successfully adopted for large-core PPs [7]:

$$F(r_{iI}, \delta^I) = [1 - exp(-\delta^I r_{iI}^2)]^n \qquad \text{with} \qquad x = i, I. \tag{59}$$

The dipole polarizability α_I of the core I can be, e.g., evaluated at the coupled HF or DHF level, whereas the cutoff parameter δ_I can be fitted to match ionization potentials of one-valence electron systems from correlated calculations. Alternatively, the parameters can also be taken from experimental data, if this is available with sufficiently high accuracy.

It should be noted that unlike the ECPs (PPs, MPs), the CPP of a many-valence electron atom has one- and two-electron contributions and the CPP for a multicenter system is not a simple superposition of atomic one-electron potentials. The square of the electric field $\hat{\mathbf{f}}_I$ in Eq. 56 leads to modified electron-electron and core-core/nuclei interactions. Integrals over the above CPP operator for Cartesian Gaussian basis functions have been derived and implemented by Schwerdtfeger and Silberbach [45].

The usage of CPPs is especially popular for semiempirical large-core PPs, e.g., when treating alkaline and alkaline earth elements as one- and two-valence electron atoms, respectively [7, 46]. CPPs are also used in combination with semiempirical large-core PPs for main group elements [9] or when treating the group 11 and 12 metals with closed d^{10} shell as one- and two-valence electron atoms, respectively [6]. In the latter case, the extension to include higher-order effects, e.g., the quadrupole polarizability, has also been explored [47]. Their usage of CPPs together with ab initio small-core PPs as well as with MPs in general has not yet been sufficiently investigated, despite their quite successful applications at the AE level [8].

Valence Basis Sets

For the MP approach, the computational savings arising from the use of specially optimized valence basis sets are relatively small, since the MP valence orbitals keep all radial nodes which are present in their AE counterparts. The situation is quite different for the PP approach, since here the number of radial nodes in the pseudo-valence orbitals is reduced for most angular quantum numbers with respect to the AE case. Since the shape of the PP in the core region is to a certain extent arbitrary, each PP requires an individually optimized valence basis set. The usage of (truncated) AE basis sets or valence basis sets adopted from another PP may lead to significant errors in the total valence energies and in consequence possibly also to errors in the calculated atomic and molecular properties. The PP approach has the advantage to lead to relatively small basis set superposition errors. This advantage however may be lost when basis sets not optimized for the specific PP are used. Many of the published sets of PPs come together with corresponding valence basis sets, which therefore should be used in unchanged form or might be further extended, e.g., by adding diffuse, polarization, and correlation functions.

During the last decades, it became quite popular to use series of basis sets, e.g., in order to estimate the basis set limit from results of, e.g., double-, triple-, and quadruple-zeta quality by extrapolation techniques. For energy-consistent PPs, two types of such systematic basis sets have been developed for a large number of elements, i.e., the correlation-consistent generalized contracted basis sets of Peterson and collaborators (e.g., cc-pVXZ-PP and cc-pwCVXZ-PP, X = D, T, Q, 5) [36–38] and the error-balanced segmented contracted basis sets of Weigend and coworkers (def2-XVP and dhf-XVP, X = S, TZ, QZ) [48, 49]. The focus in the latter basis sets was also put on their performance in two-component HF and DFT calculations (dhf-XVP-2c, X = S, TZ, QZ). Further references to these and other basis sets for ECPs are provided in Ref. [11].

Calibration

A successful adjustment of an ECP for an atom at a certain level of theory, e.g., at the HF or DHF level, does not guarantee its success in calculations using different computational methods, e.g., correlated approaches, or when the ECP is used in a molecular environment. Calibration calculations with a rigorous comparison to results of accurate AE calculations, preferentially using the Hamiltonian the ECP aims to model and basis sets of comparable quality, or experimental data are therefore mandatory for ECPs.

Figure 6 summarizes results for the ground state of Cu_2 when Cu is treated as a one-valence electron ion [6]. The underlying large-core PP replacing the $1s^2 \ldots 3d^{10}$ core is designed to model scalar-relativistic DHF/DC results. The core-core repulsion correction elongates the bond by more than 0.3 Bohr, whereas the inclusion of static core polarization and core-valence correlation by means of a CPP shortens the bond by more than 0.2 Bohr. After including valence correlation with a local correlation density functional, the bond distance agrees within 0.1 Bohr with the experimental value. The better performance of the PP at the HF level without any corrections is most likely due to error cancellations. This becomes obvious from the corresponding results for the binding energy and the vibrational frequency, which approach the experimental values very systematically upon adding the corrections.

An example for a correlated atomic calibration study of a MCDHF/DC+B-adjusted small-core PP for the roentgenium atom is provided in Table 4. AE DCB Fock-space coupled cluster (FSCC) results of Eliav et al. [50] using very large basis sets are compared to corresponding PP MRCI data using standard basis sets [18], which are applicable also in molecular calculations [39]. The agreement with the AE reference data is on the average within 0.15 eV (3 %) provided that, as necessary in the FSCC approach, the spinors are taken from the closed-shell $d^{10}\ s^2\ ^1S_0$ anion ground state. Larger deviations occur when spinors individually optimized for each configuration are used in the PP MRCI calculations, which points to substantial relaxation effects not covered by the FSCC approach. The deviations between AE

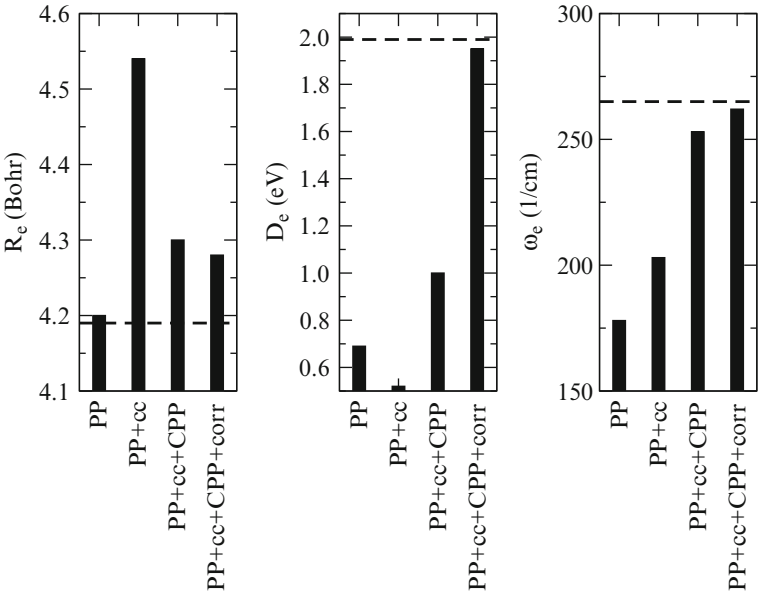


Fig. 6 Equilibrium distance R_e, binding energy D_e, and vibrational frequency ω_e of Cu_2 from HF calculations with a DHF/DC-adjusted one-valence electron PP, including a core-core (cc) repulsion correction, a core polarization potential (CPP), and valence correlation (corr) by means of the Vosko-Wilk-Nusair local correlation functional [6]. Experimental values are indicated by the *horizontal dashed lines*

Table 4 Electron affinity (EA), excitation energies (EE), and ionization potentials (IP) of roentgenium (Rg) ($Z = 111$) from MRCI PP calculations in comparison to AE DCB FSCC results (eV) [18,50]

		AE FSCC		PP MRCI	
		DC	DCB	Basis a	Basis b
EA	$d^9\ s^2\ ^2D_{5/2} - d^{10}\ s^2\ ^1S_0$	1.542	1.565	1.588	0.991
EE	$d^9\ s^2\ ^2D_{5/2} - d^9\ s^2\ ^2D_{3/2}$	2.719	2.687	2.629	2.511
	$d^9\ s^2\ ^2D_{5/2} - d^{10}\ s^1\ ^1S_{1/2}$	3.006	2.953	2.691	3.070
IP	$d^9\ s^2\ ^2D_{5/2} - d^8\ s^2\ ^3D_4$	10.57	10.60	10.859	9.852
	$d^9\ s^2\ ^2D_{5/2} - d^9\ s^1\ ^3D_3$	12.36	12.33	12.408	11.925
	$d^9\ s^2\ ^2D_{5/2} - d^{10}\ ^1S_0$	15.30	15.23	14.989	14.912
m.a.d.				0.15	0.39
m.r.d.				3.0	13.6

AE FSCC: 26s21p16d9f7g5h, atomic spinors from $d^{10}\ s^2\ ^1S_0$
PP MRCI basis a: (12s11p9d3f)/[5s7p5d3f], atomic spinors from $d^{10}\ s^2\ ^1S_0$
PP MRCI basis b: (12s11p9d3f)/[5s7p5d3f], individual atomic spinors for each configuration
Mean absolute deviations (m.a.d., eV), mean relative deviations (m.r.d., %)

Table 5 Bond distances R_e (Å) and force constants k_e (N/m) for the monohydride of roentgenium from AE and PP calculations [39]

Method	DHF		MP2	
	R_e	k_e	R_e	k_e
AE, DCG	1.520	453.10	1.502	477.95
PP, DC+B	1.518	451.98	1.498	477.45

and PP data in these calculations are much larger than in the PP fit: in the case of Rg, the energy adjustment was performed for 37 nonrelativistic configurations corresponding to 309 J levels. The mean absolute errors in the total valence energies were below 0.01 eV both for the nonrelativistic configurations and the J levels.

Table 5 lists results for the equilibrium distance and the force constant of the monohydride of roentgenium using the same MCDHF/DC+B-adjusted small-core PP for roentgenium in two-component HF and MP2 calculations [39]. Comparison is made to corresponding AE DHF/DCG results, i.e., the Gaunt term was used instead of the full Breit term. In view of the very large relativistic bond length contraction of about 0.5 Å the obtained accuracy of better than 0.004 Å (0.27 %) of the PP results for the bond distances is quite satisfactory. The agreement of the force constants is better than 0.25 %.

Summary

The effective core potential approach, i.e., its model potential and pseudopotential variants, and the effective core polarization potential approach have been briefly reviewed. Despite the development of efficient, i.e., not too costly but still accurate, relativistic all-electron schemes, effective core potentials will still remain the workhorse for relativistic quantum chemical calculations on larger systems. The possibility to include besides the Dirac relativity also the Breit interaction as well as quantum electrodynamic effects implicitly in the calculations for heavy elements renders the approach quite attractive, besides the computational savings due to the reduced number of electrons to deal with.

References

1. Pyykkö P (1978) Relativistic quantum chemistry. Adv Quantum Chem 11:353–409
2. Pyykkö P (1988) Relativistic effects in structural chemistry. Chem Rev 88:563–594
3. Hellmann H (1935) A new approximation method in the problem of many electron electrons. J Chem Phys 3:61
4. Bonifacic V, Huzinaga S (1974) Atomic and molecular calculations with the model potential method. I. J Chem Phys 60:2779–2786
5. Phillips JC, Kleinman L (1959) New method for calculating wave functions in crystals and molecules. Phys Rev 116:287–294

6. Stoll H, Fuentealba P, Dolg M, Flad J, v. Szentpály L, Preuß H (1983) Cu and Ag as one-valence-electron atoms: pseudopotential results for Cu_2, Ag_2, CuH, AgH, and the corresponding cations. J Chem Phys 79:5532–5542
7. Fuentealba P (1982) On the reliability of semiempirical pseudopotentials: dipole polarizability of the alkali atoms. J Phys B: At Mol Phys 15:L555–L558
8. Müller W, Flesch J, Meyer W (1982) Treatment of intershell correlation effects in ab initio calculations by use of core polarization potentials. Method and application to alkali and alkaline earth atoms. J Chem Phys 80:3297–3310
9. Igel-Mann G, Stoll H, Preuss H (1988) Pseudopotentials for main group elements (IIIa through VIIa). Mol Phys 65:1321–1328
10. Schwerdtfeger P (2003) Relativistic pseudopotentials. In: Kaldor U, Wilson, S (eds) Progress in theoretical chemistry and physics: theoretical chemistry and physics of heavy and superheavy elements. Kluwer Academic, Dordtrecht, pp 399–438
11. Cao X, Dolg M (2012) Relativistic pseudopotentials: their development and scope of applications. Chem Rev 112:403–480
12. Dyall KG, Grant IP, Johnson CT, Parpia FA, Plummer EP (1989) GRASP – a general-purpose relativistic atomic structure program. Comput Phys Commun 55: 425–456
13. Wood JH, Boring AM (1978) Improved Pauli Hamiltonian for local-potential problems. Phys Rev B 18:2701–2711
14. Cowan RD, Griffin DC (1976) Approximate relativistic corrections to atomic radial wave functions. J Opt Soc Am 66:1010–1014
15. Cao X, Dolg M (2012) Relativistic pseudopotentials. In: Barysz M, Ishikawa Y (eds) Relativistic methods for chemists. Challenges and advances in computational physics, vol 10. Wiley, Chichester, pp 200–210
16. Pittel B, Schwarz WHE (1977) Correlation energies from pseudopotential calculations. Chem Phys Lett 46:121–124
17. Dolg M (1996) On the accuracy of valence correlation energies in pseudopotential calculations. J Chem Phys 104:4061–4067
18. Hangele T, Dolg M, Hanrath M, Cao X, Schwerdtfeger P (2012) Accurate relativistic energy-consistent pseudopotentials for the superheavy elements 111 to 118 including quantum electrodynamic effects. J Chem Phys 136:214105-1-11
19. Schwarz WHE (1968) Das kombinierte Näherungsverfahren. I. Theoretische Grundlagen. Theor Chim Acta 11:307–324
20. Kahn LR, Goddard WA (1968) A direct test of the validity of the use of pseudopotentials in molecules. Chem Phys Lett 2:667–670
21. Weeks JD, Rice SA (1968) Use of pseudopotentials in atomic structure calculations. J Chem Phys 49:2741–2755
22. Lee YS, Ermler WC, Pitzer KS (1977) Ab initio effective core potentials including relativistic effects. I. Formalism and applications to the Xe and Au atoms. J Chem Phys 67:5861–5876
23. Hafner P, Schwarz WHE (1978) Pseudopotential approach including relativistic effects. J Phys B: At Mol Phys 11:217–233
24. Pitzer RM, Winter NW (1988) Electronic-structure methods for heavy-atom molecules. J Phys Chem 92:3061–3063
25. McMurchie LE, Davidson ER (1981) Calculation of integrals over ab initio pseudopotentials. J Comput Phys 44:289–301
26. Pitzer RM, Winter NW (1991) Spin-orbit (core) and core potential integrals. Int J Quantum Chem 40:773–780
27. Pélissier M, Komiha N, Daudey JP (1988) One-center expansion for pseudopotential matrix elements. J Comput Chem 9:298–302
28. Durand P, Barthelat JC (1974) New atomic pseudopotentials for electronic structure calculations of molecules and solids. Chem Phys Lett 27:191–194
29. Christiansen PA, Lee YS, Pitzer KS (1979) Improved ab initio effective core potentials for molecular calculations. J Chem Phys 71:4445:4450

30. Durand P, Barthelat JC (1975) A theoretical method to determine atomic pseudopotentials for electronic structure calculations of molecules and solids. Theor Chim Acta 38:283–302
31. Barthelat JC, Durand P (1978) Recent progress of pseudo-potential methods in quantum chemistry. Gazz Chim Ital 108:225–236
32. Titov AV, Mosyagin NS (1999) Generalized relativistic effective core potential: theoretical grounds. Int J Quantum Chem 71:359–401
33. Hay PJ, Wadt WR (1985) Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg. J Chem Phys 82:270–282
34. Pacios LF, Christiansen PA (1985) Ab initio relativistic effective potentials with spin-orbit operators. I. Li through Ar. J Chem Phys 82:2664–2671
35. Stevens WJ, Basch H, Krauss M (1984) Compact effective potentials and efficient shared-exponent basis sets for the first- and second-row atoms. J Chem Phys 81:6026–6033
36. Peterson KA, Figgen D, Goll E, Stoll H, Dolg M (2003) Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-d group 16–18 elements. J Chem Phys 119:11113–111123
37. Peterson KA, Figgen D, Dolg M, Stoll H (2007) Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4d elements Y – Pd. J Chem Phys 126:124101-1-12
38. Figgen D, Peterson KA, Dolg M, Stoll H (2009) Energy-consistent pseudopotentials and correlation consistent basis sets for the 5d elements Hf – Pt. J Chem Phys 130:164108-1-12
39. Hangele T, Dolg M (2013) Accuracy of relativistic energy-consistent pseudopotentials for superheavy elements 111–118: molecular calibration calculations. J Chem Phys 138:044104-1–8
40. Huzinaga S, Cantu AA (1971) Theory of separability of many-electron systems. J Chem Phys 55:5543–5549
41. Seijo L, Barandiarán Z (1999) The ab initio model potential method: a common strategy for effective core potential and embedded cluster calculations. In: Leszczynski J (ed) Computational chemistry: reviews of current trends, vol 4. World Scientific, Singapore, pp 55–152
42. Seijo L (1995) Relativistic ab initio model potential calculations including spin-orbit effects through the Wood-Boring Hamiltonian. J Chem Phys 102:8078–8088
43. Klobukowski M, Huzinaga S, Sakai Y (1999) Model core potentials: theory and application. In: Leszczynski J (ed) Computational chemistry: reviews of current trends, vol 3. World Scientific, Singapore, pp 49–74
44. Zheng T, Klobukowski M (2009) New model core potential for gold. J Chem Phys 130:204107-1–12
45. Schwerdtfeger P, Silberbach H (1998) Multicenter integrals over long-range operators using Cartesian Gaussian functions. Phys Rev A 37:2834–2842
46. Fuentealba P, Preuss H, Stoll H, v Szentpály L (1982) A proper account of core polarization with pseudopotentials: single valence-electron alkali compounds. Chem Phys Lett 89:418–422
47. Stoll H, Fuentealba P, Schwerdtfeger P, Flad J, v Szentpály L, Preuß H (1984) Cu and Ag as one-valence-electron atoms: CI results and quadrupole corrections for Cu_2, Ag_2, CuH, AgH. J Chem Phys 81:2732–2736
48. Weigend F, Ahlrichs R (2005) Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. Phys Chem Chem Phys 7:3297–3305
49. Weigend F, Baldes A (2010) Segmented contracted basis sets for one- and two-component Dirac-Fock effective core potentials. J Chem Phys 133:174102-1–11
50. Eliav E, Kaldor, U, Schwerdtfeger P, Hess, BA, Ishikawa Y (1994) Ground state electron configuration of element 111. Phys Rev Lett 73:3203–3206

Part IV

Relativistic Wave Functions and Density Functionals

Wenjian Liu

Basic Structures of Relativistic Wave Functions

15

Sihong Shao, Zhendong Li, and Wenjian Liu

Contents

Introduction 482
Block Structures 484
Matrix Ring 485
Tracy-Singh Product 486
Internal Symmetries 490
Permutation 490
Angular Momenta 492
Space Inversion 493
Application to a Reduced 2-Electron System 494
Summary 495
References 495

Abstract

It is shown that relativistic many-body Hamiltonians and wave functions can be expressed systematically with Tracy-Singh products for partitioned matrices. The latter gives rise to the usual notion for a relativistic N-electron wave function: a column vector composed of 2^N blocks, each of which consists of 2^N components formed by the Kronecker products of N one-electron 2-spinors. Yet, the noncommutativity of the Tracy-Singh product dictates that the chosen serial

S. Shao (✉)
LMAM and School of Mathematical Sciences, Peking University, Beijing, China
e-mail: sihong@math.pku.edu.cn

Z. Li (✉) • W. Liu
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: zhendongli2008@gmail.com; liuwjbdf@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_7

ordering of electronic coordinates cannot be altered when antisymmetrizing a Tracy-Singh product of 4-spinors. It is further shown that such algebraic representation uncovers readily the internal symmetries of the relativistic Hamiltonians and wave functions, which are crucial for deriving the electron-electron coalescence conditions.

Keywords
Relativistic Hamiltonian • Relativistic wave function • Tracy-Singh product • Block structure • Internal symmetry

Introduction

The Dirac equation [1] describes the relativistic motion of an electron, which has spin $1/2$ and charge -1, and further acquires an angular momentum quantum number corresponding to a quantized atomic orbital when bounded to an atomic nucleus. The 1-electron Dirac Hamiltonian $\hat{h}^D$ is usually written in the atomic units as

$$\hat{h}^D = c\vec{\alpha} \cdot \vec{p} + \beta c^2 + \phi I_4, \tag{1}$$

where

$$\vec{\alpha} = \begin{pmatrix} 0 & \vec{\sigma} \\ \vec{\sigma} & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \tag{2}$$

$$\vec{p} = -\mathbb{i}\nabla, \tag{3}$$

$$\phi = -\sum_A \frac{Z_A}{r_A}, \quad r_A = |\vec{r} - \vec{r}_A|. \tag{4}$$

Here, the constant $c = 1/\alpha$ is the speed of light, $\vec{\alpha}$ and β are the usual 4-by-4 Dirac matrices, $\vec{\sigma}$ is the vector of the Pauli spin matrices, $\vec{p}$ is the linear momentum operator, and ϕ denotes the nuclear attraction. The stationary 1-electron Dirac equation reads

$$\hat{h}^D \Psi_D(\vec{r}) = E\Psi_D(\vec{r}), \tag{5}$$

which is an eigenvalue problem and has the same form as the 1-electron Schrödinger equation

$$\hat{h}^S \Psi_S(\vec{r}) = E\Psi_S(\vec{r}), \tag{6}$$

$$\hat{h}^S = -\frac{1}{2}\nabla^2 + \phi. \tag{7}$$

The N-electron Dirac equation can formally be written in the following form:

$$\hat{H}\Psi_D(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) = E\Psi_D(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N), \tag{8}$$

where the Hamiltonian $\hat{H}$ is defined in the framework of clamped nuclei as

$$\hat{H} = \sum_{k}^{N} \hat{h}_k^D + \sum_{k>l}^{N} \hat{g}_{kl}. \tag{9}$$

Here, $\hat{h}_k^D$ is the Dirac operator (9) for electron k, and the electron-electron interaction operator $\hat{g}_{kl}$ can be put into a generic form

$$\hat{g}_{kl} = d_C \frac{1}{r_{kl}} + d_G \frac{\vec{\alpha}_k \cdot \vec{\alpha}_l}{r_{kl}} + d_R \frac{(\vec{\alpha}_k \cdot \hat{r}_{kl})(\vec{\alpha}_l \cdot \hat{r}_{kl})}{r_{kl}} \tag{10}$$

with $\hat{r}_{kl} = \vec{r}_{kl}/r_{kl}$. The Coulomb interaction is represented by $\hat{g}_{kl}$ with $(1, 0, 0)$ for the coefficients (d_C, d_G, d_R). Likewise, the Coulomb-Gaunt and Coulomb-Breit interactions are recovered by the coefficients $(1, -1, 0)$ and $(1, -1/2, -1/2)$, respectively. However, it must be pointed out that Eq. (8) is only formal and has a vague meaning. Both the structure in which $\hat{h}_k^D$ and $\hat{g}_{kl}$ constitute $\hat{H}$ (9) and the way how $\hat{H}$ acts on Ψ_D are not clearly shown therein.

Apart from the fact that the first quantized N-electron Dirac equation (8) has some unphysical ingredients and hence cannot be solved as it stands [2–4], it also differs from the N-electron Schrödinger equation in at least two mathematical aspects.

The first main difference lies in that $\hat{h}^D$ (1) is a 4-by-4 matrix operator, while $\hat{h}^S$ (7) is only a scalar operator. That is, the Dirac equation (5) is a set of four equations with the unknown $\Psi_D(\vec{r})$ being a 4-component complex vector (4-spinor), but the Schrödinger equation (6) is just an equation with the unknown $\Psi_S(\vec{r})$ being a complex scalar. When going from the 1-electron to the N-electron case, such difference becomes even more enhanced. Actually, for an N-electron system, $\Psi_D(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$ has 4^N components, each of which depends on $3N$ spatial coordinates of the electrons,

$$\Psi_D^{X_1 X_2 \ldots X_N}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N), \quad X_k \in \{L\alpha, L\beta, S\alpha, S\beta\}, \tag{11}$$

but $\Psi_S(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$ still has only one component (i.e., $1^N \equiv 1$). Moreover, the components (11) are not completely independent, since antisymmetry principle dictates that they must satisfy the following relation:

$$\begin{aligned} &\Psi_D^{X_1 X_2 \ldots X_k \ldots X_l \ldots X_N}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_k, \ldots, \vec{r}_l, \ldots, \vec{r}_N) \\ &= -\Psi_D^{X_1 X_2 \ldots X_l \ldots X_k \ldots X_N}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_l, \ldots, \vec{r}_k, \ldots, \vec{r}_N). \end{aligned} \tag{12}$$

Therefore, the Dirac equation (8) is mathematically much harder than the Schrödinger equation. For instance, the electron-electron coalescence conditions

for $\Psi_D(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$ were derived only recently [5], but those for $\Psi_S(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$ have been known since 1957 [6].

A second main difference lies in that the spectrum of the 1-electron Dirac operator $\hat{h}^D$ (1) consists of a point spectrum, a positive energy continuum, and a negative energy continuum and is therefore unbounded from below, while the spectrum of the 1-electron Schrödinger operator $\hat{h}^S$ does not contain any negative energy state and is bounded from below. This negative energy continuum brings challenges to both mathematical analysis and numerical solutions of the N-electron Dirac equation (8). In numerics, the so-called variational collapse is quite troublesome although practical solutions have been available [7–9]. In theory, the spectrum of the 1-electron Dirac operator is well understood [10], but little is known on the spectrum of the N-electron ($N \geq 2$) Dirac operator [11]. Even for the simplest 2-electron Dirac-Coulomb operator (i.e., $(d_C, d_G, d_R) = (1, 0, 0)$ in Eq. (10)), the question whether it has bound states is still open [5]. Obviously, the first step toward a theoretical understanding of the spectrum of the N-electron Dirac operator (9) is to uncover the structure of the operator itself.

In this chapter, we will examine the basic structures of the N-electron Dirac operator and the corresponding wave function $\Psi_D(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$, by virtue of matrix rings and Tracy-Singh products. In this representation, the N-electron Dirac operator and other typical operators have explicit forms that can readily be used in direct calculations (e.g., verification of commutation relations). Their actions on 4-spinors can also be made crystal clear. It is further shown that such algebraic representation can uncover the internal symmetries of the relativistic Hamiltonians and wave functions, which are crucial for understanding the electron-electron coalescence conditions.

Block Structures

Equation (1) can be written more explicitly in block form:

$$\hat{h} = \begin{pmatrix} \phi + c^2 & c(\vec{\sigma} \cdot \vec{p}) \\ c(\vec{\sigma} \cdot \vec{p}) & \phi - c^2 \end{pmatrix}, \quad \Psi(\vec{r}) = \begin{pmatrix} \psi^L(\vec{r}) \\ \psi^S(\vec{r}) \end{pmatrix}. \tag{13}$$

For brevity, the superscript (resp. subscript) "D" in $\hat{h}$ (resp. $\Psi(\vec{r})$) has been dropped. For a positive energy solution, the lower component $\psi^S(\vec{r})$ of $\Psi(\vec{r})$ is smaller (in the mean sense) than the upper component $\psi^L(\vec{r})$ by a factor of c^{-1} [10]. They are hence often called small and large components, respectively. In particular, the former vanishes in the nonrelativistic limit. The N-electron wave function $\Psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)$ is usually written also in block form with 2^N blocks, each of which consists of 2^N components formed by the Kronecker products of N one-electron 2-spinors. However, it is not clear how this structure is related to the N-electron Hamiltonian (9). To reveal this, we invoke matrix ring and Tracy-Singh product for partitioned matrices.

Matrix Ring

In mathematics, a *ring* is an algebraic structure consisting of a set of addition and multiplication operations. The set is required to be an abelian group under addition and a semigroup (where no element has to have an inverse) under multiplication, such that multiplication distributes over addition. All the 2-by-2 complex matrices form a *matrix ring* $K_1 = M_2(\mathbb{C})$ under the matrix addition and multiplication, with the four matrices $\{I_2, \sigma_x, \sigma_y, \sigma_z\}$ being the basis. Likewise, the set of 2-by-2 matrices with entries from ring K_1 is defined as another matrix ring $\tilde{K}_1 = M_2(K_1)$ under the matrix addition and multiplication. The elements of $\tilde{K}_1$, which are originally 4-by-4 matrices of complex numbers, are now treated as 2-by-2 matrices with each entry being a 2-by-2 matrix in K_1. A *scalar multiplication* between rings K_1 and $\tilde{K}_1$ can then be introduced:

$$A = (a_{ij}) \in K_1, \quad \mathbf{A} = (A_{ij}) \in \tilde{K}_1; \tag{14}$$

$$A\mathbf{A} := (AA_{ij})_{ij} \in \tilde{K}_1, \quad \mathbf{A}A := (A_{ij}A)_{ij} \in \tilde{K}_1, \tag{15}$$

where the notation $(AA_{ij})_{ij}$ means that the ij-th entry of $A\mathbf{A}$ is $A\mathbf{A}_{ij}$ calculated by the multiplication in K_1. This represents a generalization of the usual scalar multiplication between $\mathbb{C}$ and K_1, i.e., $aA = (aa_{ij})_{ij} \in K_1$ and $Aa = (a_{ij}a)_{ij} \in K_1$ for $a \in \mathbb{C}$ and $A = (a_{ij}) \in K_1$. Note that $A\mathbf{A}$ is generally different from $\mathbf{A}A$ for the multiplication in K_1 is not commutative. However, for the special case $A_{ij} \in \{I_2, 0_2\}$, the identity $A\mathbf{A} = \mathbf{A}A$ always holds. The matrices $\mathbf{E}_{ij}$ defined as

$$[\mathbf{E}_{ij}]_{kl} = \delta_{ik}\delta_{jl}I_2 \tag{16}$$

form the standard basis of $\tilde{K}_1$. Thus, an arbitrary 4-by-4 matrix $\mathbf{A}$ can be reexpressed as

$$\mathbf{A} = \sum_{i,j=1}^{2} A_{ij}\mathbf{E}_{ij} = \sum_{i,j=1}^{2} \mathbf{E}_{ij}A_{ij}, \quad A_{ij} \in K_1. \tag{17}$$

The following associative law can also readily be verified:

$$(A\mathbf{A})B = A(\mathbf{A}B), \quad \forall A, B \in K_1, \quad \forall \mathbf{A} \in \tilde{K}_1. \tag{18}$$

In terms of the above definitions and relationships, any 4-by-4 matrix operator can be decomposed into a "block operator" in $\tilde{K}_1$ that transforms linearly or simply interchanges the blocks ψ^L and ψ^S of the Dirac 4-spinor $(\psi^L, \psi^S)^T$ and a "component operator" in K_1 that acts on the components of every block. For instance, the Dirac matrix $\vec{\boldsymbol{\alpha}}$ (2) can be decomposed as

$$\vec{\boldsymbol{\alpha}} = \begin{pmatrix} 0 & \vec{\sigma} \\ \vec{\sigma} & 0 \end{pmatrix} = \vec{\sigma} \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix} = \vec{\sigma}\boldsymbol{\gamma}^5 = \boldsymbol{\gamma}^5\vec{\sigma}, \tag{19}$$

where $\boldsymbol{\gamma}^5 \in \tilde{K}_1$ is the "block operator" and $\vec{\sigma} \in K_1$ the "component operator." Likewise, the one-electron Dirac Hamiltonian $\hat{h}$ (13) can be reexpressed as

$$\hat{h} = c(\vec{\sigma} \cdot \vec{p})\boldsymbol{\gamma}^5 + c^2\boldsymbol{\beta} + \phi\mathbf{I}_4, \tag{20}$$

where $\vec{\sigma} \cdot \vec{p} \in K_1$, $\boldsymbol{\gamma}^5, \boldsymbol{\beta}, \mathbf{I}_4 \in \tilde{K}_1$, and the nuclear attraction ϕ is just a scalar. Such a formalism is advantageous in that the part of an operator that retains the block structure of the wave function, e.g., $\boldsymbol{\gamma}^5$ in Eq. (19), and the part that mixes the internal components of a block, e.g., $\vec{\sigma}$ in Eq. (19), are separated out. Moreover, it allows an easy and consistent extension to the many-body case through the Tracy-Singh product [12, 13], which is a generalization of the Kronecker product for partitioned matrices.

Tracy-Singh Product

Given matrices $\mathbf{A}$ of size $m \times n$ and $\mathbf{B}$ of size $p \times q$, the Kronecker product is defined as

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})_{ij}, \tag{21}$$

meaning that the ij-th subblock of the mp-by-nq product $\mathbf{A} \otimes \mathbf{B}$ is the p-by-q matrix $a_{ij}\mathbf{B}$. The following identity holds for the Kronecker product

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}) \tag{22}$$

if the matrix products $\mathbf{AC}$ and $\mathbf{BD}$ are allowed to be formed. Let $m = \sum_{i=1}^{\tau} m_i$, $n = \sum_{j=1}^{\kappa} n_j$, $p = \sum_{k=1}^{\lambda} p_k$, and $q = \sum_{l=1}^{\mu} q_l$ be partitions of integers m, n, p, q, and $\mathbf{A}$ and $\mathbf{B}$ be so partitioned as well, viz., $\mathbf{A} = (A_{ij})$ and $\mathbf{B} = (B_{ij})$ with A_{ij} and B_{kl} being the submatrices of sizes $m_i \times n_j$ and $p_k \times q_l$, respectively. Then the Tracy-Singh product [12, 13] is defined as

$$\mathbf{A} \circ \mathbf{B} = (A_{ij} \circ \mathbf{B})_{ij} = ((A_{ij} \otimes B_{kl})_{kl})_{ij}, \tag{23}$$

meaning that the ij-th subblock of the mp-by-nq product $\mathbf{A} \circ \mathbf{B}$ is the $m_i p$-by-$n_j q$ matrix $A_{ij} \circ \mathbf{B}$, the kl-th subblock of which is equal to the $m_i p_k$-by-$n_j q_l$ matrix $A_{ij} \otimes B_{kl}$. As such, the Tracy-Singh product is just the pairwise Kronecker product of the partitioned matrices. Note in passing that $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{A} \circ \mathbf{B}$ are of the same size $mp \times nq$. They are related by simple permutation matrices [12] for equal-sized partitioning and become identical for non-partitioned matrices (i.e., $m_i = n_j = p_k = q_l = 1$).

As the two-electron relativistic wave function $\Psi = (\Psi^{LL}, \Psi^{LS}, \Psi^{SL}, \Psi^{SS})^T$ belongs to $\mathbb{C}^{4^2}$ with each block $\Psi^{XY}(X, Y \in \{L, S\})$ in $\mathbb{C}^4$, we shall use rings $K_2 = M_{2^2}(\mathbb{C})$ and $\tilde{K}_2 = M_{2^2}(K_2)$ to express the two-electron Hamiltonian

$$\hat{H} = \sum_{k=1}^{2}(c\vec{\alpha}_k \cdot \vec{p}_k + \beta_k c^2 + \phi_k) + \hat{g}_{12}, \tag{24}$$

$$\hat{g}_{12} = d_C \frac{1}{r_{12}} + d_G \frac{\vec{\alpha}_1 \cdot \vec{\alpha}_2}{r_{12}} + d_R \frac{(\vec{\alpha}_1 \cdot \hat{r}_{12})(\vec{\alpha}_2 \cdot \hat{r}_{12})}{r_{12}}. \tag{25}$$

Note that the multiplication between K_2 and $\tilde{K}_2$ can be defined in the same way as Eq. (15) for K_1 and $\tilde{K}_1$. Actually, we have the following mixed-product property, which is the key for the extension of the above formalism to the many-body case.

Proposition 1. *For $A, C \in K_1$ and $\mathbf{B}, \mathbf{D} \in \tilde{K}_1$, we have*

$$(A\mathbf{B}) \circ (C\mathbf{D}) = (A \otimes C)(\mathbf{B} \circ \mathbf{D}), \tag{26}$$

with $A \otimes C \in K_2$ and $\mathbf{B} \circ \mathbf{D} \in \tilde{K}_2$.

Proof. Direct algebraic calculation shows that

$$\begin{aligned}
&(A\mathbf{B}) \circ (C\mathbf{D}) \\
&= \begin{pmatrix} AB_{11} & AB_{12} \\ AB_{21} & AB_{22} \end{pmatrix} \circ \begin{pmatrix} CD_{11} & CD_{12} \\ CD_{21} & CD_{22} \end{pmatrix} \\
&= \begin{pmatrix} AB_{11} \circ \begin{pmatrix} CD_{11} & CD_{12} \\ CD_{21} & CD_{22} \end{pmatrix} & AB_{12} \circ \begin{pmatrix} CD_{11} & CD_{12} \\ CD_{21} & CD_{22} \end{pmatrix} \\ AB_{21} \circ \begin{pmatrix} CD_{11} & CD_{12} \\ CD_{21} & CD_{22} \end{pmatrix} & AB_{22} \circ \begin{pmatrix} CD_{11} & CD_{12} \\ CD_{21} & CD_{22} \end{pmatrix} \end{pmatrix} \\
&= \begin{pmatrix} AB_{11} \otimes CD_{11} & AB_{11} \otimes CD_{12} & AB_{12} \otimes CD_{11} & AB_{12} \otimes CD_{12} \\ AB_{11} \otimes CD_{21} & AB_{11} \otimes CD_{22} & AB_{12} \otimes CD_{21} & AB_{12} \otimes CD_{22} \\ AB_{21} \otimes CD_{11} & AB_{21} \otimes CD_{12} & AB_{22} \otimes CD_{11} & AB_{22} \otimes CD_{12} \\ AB_{21} \otimes CD_{21} & AB_{21} \otimes CD_{22} & AB_{22} \otimes CD_{21} & AB_{22} \otimes CD_{22} \end{pmatrix} \\
&= \begin{pmatrix} (A \otimes C)(B_{11} \otimes D_{11}) & (A \otimes C)(B_{11} \otimes D_{12}) & (A \otimes C)(B_{12} \otimes D_{11}) & (A \otimes C)(B_{12} \otimes D_{12}) \\ (A \otimes C)(B_{11} \otimes D_{21}) & (A \otimes C)(B_{11} \otimes D_{22}) & (A \otimes C)(B_{12} \otimes D_{21}) & (A \otimes C)(B_{12} \otimes D_{22}) \\ (A \otimes C)(B_{21} \otimes D_{11}) & (A \otimes C)(B_{21} \otimes D_{12}) & (A \otimes C)(B_{22} \otimes D_{11}) & (A \otimes C)(B_{22} \otimes D_{12}) \\ (A \otimes C)(B_{21} \otimes D_{21}) & (A \otimes C)(B_{21} \otimes D_{22}) & (A \otimes C)(B_{22} \otimes D_{21}) & (A \otimes C)(B_{22} \otimes D_{22}) \end{pmatrix} \\
&= (A \otimes C) \begin{pmatrix} B_{11} \circ \mathbf{D} & B_{12} \circ \mathbf{D} \\ B_{21} \circ \mathbf{D} & B_{22} \circ \mathbf{D} \end{pmatrix} \\
&= (A \otimes C)(\mathbf{B} \circ \mathbf{D}),
\end{aligned}$$

where the identity (22) for the Kronecker product has been used in the fifth line. □

By defining

$$\vec{\alpha}_1 := \vec{\boldsymbol{\alpha}} \circ \mathbf{I}_4, \quad \vec{\alpha}_2 := \mathbf{I}_4 \circ \vec{\boldsymbol{\alpha}}, \tag{27a}$$

$$\beta_1 := \boldsymbol{\beta} \circ \mathbf{I}_4, \quad \beta_2 := \mathbf{I}_4 \circ \boldsymbol{\beta}, \tag{27b}$$

$$\vec{\sigma}_1 := \vec{\sigma} \otimes I_2, \quad \vec{\sigma}_2 := I_2 \otimes \vec{\sigma}, \tag{27c}$$

and further noticing the relations (19) and (26), we have

$$\vec{\alpha}_k = \vec{\sigma}_k \mathbf{C}_k, \quad \mathbf{C}_k \in \tilde{K}_2, \quad k = 1, 2 \tag{28}$$

in terms of the $\mathbf{C}_k$ operators

$$\mathbf{C}_1 = \boldsymbol{\gamma}^5 \circ \mathbf{I}_4 = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix} \circ \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & I_4 & 0 \\ 0 & 0 & 0 & I_4 \\ I_4 & 0 & 0 & 0 \\ 0 & I_4 & 0 & 0 \end{pmatrix}, \tag{29}$$

$$\mathbf{C}_2 = \mathbf{I}_4 \circ \boldsymbol{\gamma}^5 = \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} \circ \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & I_4 & 0 & 0 \\ I_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_4 \\ 0 & 0 & I_4 & 0 \end{pmatrix}. \tag{30}$$

For example, Eq. (28) for $k = 1$ can be verified as follows

$$\begin{aligned} \vec{\alpha}_1 &= \vec{\boldsymbol{\alpha}} \circ \mathbf{I}_4 = (\vec{\sigma}\boldsymbol{\gamma}^5) \circ (I_2 \mathbf{I}_4) \\ &= (\vec{\sigma} \otimes I_2)(\boldsymbol{\gamma}^5 \circ \mathbf{I}_4) = \vec{\sigma}_1 \mathbf{C}_1. \end{aligned} \tag{31}$$

Consequently, we arrive at

$$\vec{\alpha}_k \cdot \vec{p}_k = (\vec{\sigma}_k \cdot \vec{p}_k)\mathbf{C}_k, \quad k = 1, 2, \tag{32}$$

$$\vec{\alpha}_1 \cdot \vec{\alpha}_2 = (\vec{\sigma}_1 \cdot \vec{\sigma}_2)\mathbf{C}_1\mathbf{C}_2 = (\vec{\sigma}_1 \cdot \vec{\sigma}_2)\mathbf{C}_{12}, \tag{33}$$

$$\begin{aligned} (\vec{\alpha}_1 \cdot \hat{r}_{12})(\vec{\alpha}_2 \cdot \hat{r}_{12}) &= [(\vec{\sigma}_1 \cdot \hat{r}_{12})\mathbf{C}_1][(\vec{\sigma}_2 \cdot \hat{r}_{12})\mathbf{C}_2] \\ &= (\vec{\sigma}_1 \cdot \hat{r}_{12})(\vec{\sigma}_2 \cdot \hat{r}_{12})\mathbf{C}_1\mathbf{C}_2 \\ &= (\vec{\sigma}_1 \cdot \hat{r}_{12})(\vec{\sigma}_2 \cdot \hat{r}_{12})\mathbf{C}_{12}, \end{aligned} \tag{34}$$

where

$$\mathbf{C}_{12} = \mathbf{C}_1\mathbf{C}_2 = \mathbf{C}_2\mathbf{C}_1 = \begin{pmatrix} 0 & 0 & 0 & I_4 \\ 0 & 0 & I_4 & 0 \\ 0 & I_4 & 0 & 0 \\ I_4 & 0 & 0 & 0 \end{pmatrix}. \tag{35}$$

Note that use of the commutation relation

$$A\mathbf{C} = \mathbf{C}A, \quad \forall A \in K_2, \ \forall \mathbf{C} \in \mathcal{G} \tag{36}$$

has been made in the above. Here $\mathcal{G}$ is the abelian group

$$\mathcal{G} = \{\mathbf{E}, \mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_{12}\} \tag{37}$$

formed from the direct product of groups $\{\mathbf{E}, \mathbf{C}_1\}$ and $\{\mathbf{E}, \mathbf{C}_2\}$, with $\mathbf{E}$ being

$$\mathbf{E} = \mathbf{I}_4 \circ \mathbf{I}_4 = \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} \circ \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} = \begin{pmatrix} I_4 & 0 & 0 & 0 \\ 0 & I_4 & 0 & 0 \\ 0 & 0 & I_4 & 0 \\ 0 & 0 & 0 & I_4 \end{pmatrix}. \tag{38}$$

Therefore, the two-electron Hamiltonian $\hat{H}$ (24) can be expressed as

$$H = \sum_{k=1}^{2} [c\vec{\sigma}_k \cdot \vec{p}_k \mathbf{C}_k + \phi_k \mathbf{E}] + c^2 \mathbf{B}_{12} + V_C \mathbf{E} + V_B \mathbf{C}_{12}, \tag{39}$$

$$\mathbf{B}_{12} = \boldsymbol{\beta} \circ \mathbf{I}_4 + \mathbf{I}_4 \circ \boldsymbol{\beta} = \begin{pmatrix} 2I_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2I_4 \end{pmatrix}, \tag{40}$$

$$V_C = d_C \frac{1}{r_{12}}, \quad V_B = d_G \frac{\vec{\sigma}_1 \cdot \vec{\sigma}_2}{r_{12}} + d_R \frac{(\vec{\sigma}_1 \cdot \hat{r}_{12})(\vec{\sigma}_2 \cdot \hat{r}_{12})}{r_{12}}, \tag{41}$$

where both $\vec{\sigma}_k \cdot \vec{p}_k$ and V_B belong to K_2, whereas the $\mathbf{E}, \mathbf{C}_k, \mathbf{C}_{12}$, and $\mathbf{B}_{12}$ matrices belong to $\tilde{K}_2$.

The above formulation can straightforwardly be generalized to the N-electron case $\{\mathbb{D}^N, K_N, \tilde{K}_N\}$ with $\mathbb{D}^N := \mathbb{C}^{4^N}$, $K_N := M_{2^N}(\mathbb{C})$, $\tilde{K}_N := M_{2^N}(K_N)$. For instance, the N-body counterpart of Eq. (27a) is simply

$$\vec{\alpha}_i := \underbrace{\mathbf{I}_4 \circ \mathbf{I}_4 \circ \cdots \circ \mathbf{I}_4}_{i-1} \circ \vec{\boldsymbol{\alpha}} \circ \underbrace{\mathbf{I}_4 \circ \cdots \circ \mathbf{I}_4}_{N-i}, \quad i = 1, \ldots, N, \tag{42}$$

which reads, e.g.,

$$\vec{\alpha}_1 := \vec{\boldsymbol{\alpha}} \circ \mathbf{I}_4 \circ \mathbf{I}_4, \quad \vec{\alpha}_2 := \mathbf{I}_4 \circ \vec{\boldsymbol{\alpha}} \circ \mathbf{I}_4, \quad \vec{\alpha}_3 := \mathbf{I}_4 \circ \mathbf{I}_4 \circ \vec{\boldsymbol{\alpha}} \tag{43}$$

for $N = 3$.

Likewise, a 4-spinor product "$\psi_i(1)\psi_j(2)\ldots\psi_k(N)$" should also be understood as a Tracy-Singh product $\psi_i(1)\circ\psi_j(2)\circ\cdots\circ\psi_k(N)$.

Internal Symmetries

The previous formalism can also be applied to such operators as permutation $\hat{\mathcal{P}}_{12}$, angular momenta $\hat{j}_{12}^2$, and $\hat{j}_{12,z}$, as well as space inversion $\hat{\mathcal{I}}$. Given the explicit forms, the internal symmetries of a 2-electron system emerge naturally.

Permutation

The operator $\hat{\mathcal{P}}_{12}$ for permuting electrons 1 and 2 can be expressed as

$$\hat{\mathcal{P}}_{12} = \hat{\pi}_{12}\hat{\Pi}_{12}\mathbf{\Pi}_{12}, \quad \hat{\Pi}_{12} \in K_2, \quad \mathbf{\Pi}_{12} \in \tilde{K}_2, \tag{44}$$

where $\hat{\pi}_{12}$ interchanges the spatial coordinates, i.e.,

$$\hat{\pi}_{12} f(\vec{r}_1, \vec{r}_2) = f(\vec{r}_2, \vec{r}_1), \tag{45}$$

while the "component operator" $\hat{\Pi}_{12}$ acts on the blocks $\Psi^{XY}(X, Y \in \{L, S\})$ of Ψ,

$$\hat{\Pi}_{12} = \frac{1}{2}(I_4 + \vec{\sigma}_1 \cdot \vec{\sigma}_2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{46}$$

The "block operator" $\mathbf{\Pi}_{12}$ in Eq. (44) is given as

$$\mathbf{\Pi}_{12} = \begin{pmatrix} I_4 & 0 & 0 & 0 \\ 0 & 0 & I_4 & 0 \\ 0 & I_4 & 0 & 0 \\ 0 & 0 & 0 & I_4 \end{pmatrix}. \tag{47}$$

Direct algebraic calculations show that

$$\hat{\Pi}_{12}\vec{\sigma}_1 = \vec{\sigma}_2\hat{\Pi}_{12}, \tag{48}$$

$$\hat{\Pi}_{12}\vec{\sigma}_2 = \vec{\sigma}_1\hat{\Pi}_{12}, \tag{49}$$

$$\hat{\Pi}_{12}(\vec{\sigma}_1 \cdot \vec{\sigma}_2) = (\vec{\sigma}_1 \cdot \vec{\sigma}_2)\hat{\Pi}_{12}, \tag{50}$$

$$\hat{\Pi}_{12}(\vec{\sigma}_1 \cdot \hat{r}_{12})(\vec{\sigma}_2 \cdot \hat{r}_{12}) = (\vec{\sigma}_1 \cdot \hat{r}_{12})(\vec{\sigma}_2 \cdot \hat{r}_{12})\hat{\Pi}_{12}, \tag{51}$$

$$\hat{\Pi}_{12} V_B = V_B \hat{\Pi}_{12}. \tag{52}$$

Similarly, we have the following relations between $\mathbf{\Pi}_{12} \in \tilde{K}_2$ and the elements of group $\mathcal{G}$ (37):

$$\mathbf{\Pi}_{12}\mathbf{C}_1 = \mathbf{C}_2\mathbf{\Pi}_{12}, \tag{53}$$

$$\mathbf{\Pi}_{12}\mathbf{C}_2 = \mathbf{C}_1\mathbf{\Pi}_{12}, \tag{54}$$

$$\mathbf{\Pi}_{12}\mathbf{C}_{12} = \mathbf{C}_{12}\mathbf{\Pi}_{12}. \tag{55}$$

As an illustration, we construct a two-electron Slater determinant in terms of the so-defined operator $\hat{\mathcal{P}}_{12}$ (44):

$$|\psi_i \psi_j\rangle = \frac{1}{\sqrt{2}}(1 - \hat{\mathcal{P}}_{12})\psi_i(1) \circ \psi_j(2) \tag{56}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} \psi_i^L(1) \otimes \psi_j^L(2) - \hat{\pi}_{12}\hat{\Pi}_{12}\psi_i^L(1) \otimes \psi_j^L(2) \\ \psi_i^L(1) \otimes \psi_j^S(2) - \hat{\pi}_{12}\hat{\Pi}_{12}\psi_i^S(1) \otimes \psi_j^L(2) \\ \psi_i^S(1) \otimes \psi_j^L(2) - \hat{\pi}_{12}\hat{\Pi}_{12}\psi_i^L(1) \otimes \psi_j^S(2) \\ \psi_i^S(1) \otimes \psi_j^S(2) - \hat{\pi}_{12}\hat{\Pi}_{12}\psi_i^S(1) \otimes \psi_j^S(2) \end{pmatrix} \tag{57}$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} \psi_i^L(1) \otimes \psi_j^L(2) - \psi_j^L(1) \otimes \psi_i^L(2) \\ \psi_i^L(1) \otimes \psi_j^S(2) - \psi_j^L(1) \otimes \psi_i^S(2) \\ \psi_i^S(1) \otimes \psi_j^L(2) - \psi_j^S(1) \otimes \psi_i^L(2) \\ \psi_i^S(1) \otimes \psi_j^S(2) - \psi_j^S(1) \otimes \psi_i^S(2) \end{pmatrix} \tag{58}$$

$$= \frac{1}{\sqrt{2}}[\psi_i(1) \circ \psi_j(2) - \psi_j(1) \circ \psi_i(2)] = -|\psi_j \psi_i\rangle, \tag{59}$$

where we have used

$$\hat{\pi}_{12}\hat{\Pi}_{12}\psi_i^X(1) \otimes \psi_j^Y(2) = \hat{\pi}_{12}\hat{\Pi}_{12} \begin{pmatrix} \psi_i^{X\alpha}(\vec{r}_1)\psi_j^{Y\alpha}(\vec{r}_2) \\ \psi_i^{X\alpha}(\vec{r}_1)\psi_j^{Y\beta}(\vec{r}_2) \\ \psi_i^{X\beta}(\vec{r}_1)\psi_j^{Y\alpha}(\vec{r}_2) \\ \psi_i^{X\beta}(\vec{r}_1)\psi_j^{Y\beta}(\vec{r}_2) \end{pmatrix} \tag{60}$$

$$= \hat{\pi}_{12} \begin{pmatrix} \psi_i^{X\alpha}(\vec{r}_1)\psi_j^{Y\alpha}(\vec{r}_2) \\ \psi_i^{X\beta}(\vec{r}_1)\psi_j^{Y\alpha}(\vec{r}_2) \\ \psi_i^{X\alpha}(\vec{r}_1)\psi_j^{Y\beta}(\vec{r}_2) \\ \psi_i^{X\beta}(\vec{r}_1)\psi_j^{Y\beta}(\vec{r}_2) \end{pmatrix} \tag{61}$$

$$= \begin{pmatrix} \psi_i^{X\alpha}(\vec{r}_2)\psi_j^{Y\alpha}(\vec{r}_1) \\ \psi_i^{X\beta}(\vec{r}_2)\psi_j^{Y\alpha}(\vec{r}_1) \\ \psi_i^{X\alpha}(\vec{r}_2)\psi_j^{Y\beta}(\vec{r}_1) \\ \psi_i^{X\beta}(\vec{r}_2)\psi_j^{Y\beta}(\vec{r}_1) \end{pmatrix} \tag{62}$$

$$= \psi_j^Y(1) \otimes \psi_i^X(2), \quad X, Y \in \{L, S\}. \tag{63}$$

It is seen from Eqs. (56) and (59) that the action of $\hat{\mathcal{P}}_{12}$ on a Tracy-Singh product of two 4-spinors is equivalent to permute the functions $\psi_p(\cdot)$ rather than the spatial coordinates. So is the action of $\hat{\pi}_{12}\hat{\Pi}_{12}$ on a Kronecker product between 2-spinors (see Eq. (63)). This feature stems directly from the noncommutativity of the Tracy-Singh or Kronecker products. The second equality of Eq. (59) emphasizes that the permutation of the orbitals in the determinant $|\psi_i \psi_j\rangle$ is antisymmetric in the common sense, such that the familiar Slater-Condon rule can directly be used for evaluating matrix elements over determinants. A close inspection of Eq. (58) further reveals the following relationship for the components $\Psi^{X_1X_2}$:

$$\Psi^{X_1X_2}(\vec{r}_1, \vec{r}_2) = -\Psi^{X_2X_1}(\vec{r}_2, \vec{r}_1), \quad X_k \in \{L\alpha, L\beta, S\alpha, S\beta\}. \tag{64}$$

It follows that the components $\Psi^{X_1X_2}(\vec{r}_1, \vec{r}_2)$ with $X_1 = X_2$ are antisymmetric in the common sense, but $\Psi^{X_1X_2}(\vec{r}_1, \vec{r}_2)$ with $X_1 \neq X_2$ do not have any permutation symmetry and are not all independent. As for the blocks Ψ^{XY} of the already antisymmetrized function $\Psi(\vec{r}_1, \vec{r}_2)$, the following relationship

$$\hat{\pi}_{12}\hat{\Pi}_{12}\Psi^{XY}(\vec{r}_1, \vec{r}_2) = -\Psi^{YX}(\vec{r}_1, \vec{r}_2), \quad X, Y \in \{L, S\} \tag{65}$$

can directly be deduced from the requirement $\hat{\mathcal{P}}_{12}\Psi(\vec{r}_1, \vec{r}_2) = -\Psi(\vec{r}_1, \vec{r}_2)$. Therefore, the antisymmetry relation (64) holds only for the individual components but not for the blocks. The notation $\Psi^{XY}(\vec{r}_1, \vec{r}_2) = -\Psi^{YX}(\vec{r}_2, \vec{r}_1)$ $(X, Y \in \{L, S\})$ often adopted in the literature is thus rather misleading and should be abandoned.

Angular Momenta

The one-electron spin operator reads

$$\vec{s} = \frac{1}{2}\vec{\Sigma} = \frac{1}{2}\begin{pmatrix} \vec{\sigma} & 0 \\ 0 & \vec{\sigma} \end{pmatrix} = \frac{1}{2}\vec{\sigma}\mathbf{I}_4. \tag{66}$$

To be consistent with the convention (27), the spin angular momenta for two electrons should be written as

$$\vec{s}_1 = \frac{1}{2}(\vec{\sigma}\mathbf{I}_4) \circ \mathbf{I}_4 = \frac{1}{2}\vec{\sigma}_1\mathbf{E}, \tag{67}$$

$$\vec{s}_2 = \frac{1}{2}\mathbf{I}_4 \circ (\vec{\sigma}\mathbf{I}_4) = \frac{1}{2}\vec{\sigma}_2\mathbf{E}. \tag{68}$$

The total angular momentum $\vec{j}_{12}$ is then given by

$$\vec{j}_{12} = \left(\vec{l}_{12} + \frac{1}{2}\vec{\sigma}_1 + \frac{1}{2}\vec{\sigma}_2\right)\mathbf{E}. \tag{69}$$

The commutation relations

$$[\vec{j}_{12}, \mathbf{C}_1] = [\vec{j}_{12}, \mathbf{C}_2] = [\vec{j}_{12}, \mathbf{C}_{12}] = 0 \tag{70}$$

follow simply from that the group $\mathcal{G}$ (37) is abelian.

Space Inversion

The space inversion operator for the relative coordinate $\vec{r}_{12}$ can be written as

$$\hat{\mathcal{I}} = \hat{\lambda}\mathbf{\Lambda}, \quad \hat{\lambda} f(\vec{r}_{12}) = f(-\vec{r}_{12}), \tag{71}$$

where the "block operator" $\mathbf{\Lambda} \in \tilde{K}_2$ reads

$$\mathbf{\Lambda} = \boldsymbol{\beta} \circ \boldsymbol{\beta} = \begin{pmatrix} I_4 & 0 & 0 & 0 \\ 0 & -I_4 & 0 & 0 \\ 0 & 0 & -I_4 & 0 \\ 0 & 0 & 0 & I_4 \end{pmatrix}. \tag{72}$$

In view of Eqs. (45) and (71), the actions of $\hat{\pi}_{12}$ and $\hat{\lambda}$ are the same for any scalar functions, such that $\hat{\lambda} \equiv \hat{\pi}_{12}$. The relations of $\mathbf{\Lambda}$ with the elements of the group $\mathcal{G}$ (37) are

$$[\mathbf{\Lambda}, \mathbf{C}_1]_+ = [\mathbf{\Lambda}, \mathbf{C}_2]_+ = [\mathbf{\Lambda}, \mathbf{C}_{12}] = 0, \tag{73}$$

where $[A, B]_+ = AB + BA$ is an anticommutator.

At this stage, it deserves to be pointed out that the above two-electron operators, including $\mathbf{B}_{12}$, $\mathbf{C}_{12}$, $\hat{\mathcal{P}}_{12}$, $\vec{j}_{12}$, and $\hat{\mathcal{I}}$, are independent of the convention (27), implying that the N-electron relativistic Hamiltonian (wave function) is symmetric (antisymmetric) in the usual sense, as dictated by the indistinguishability of the electrons. The change of the convention (27) can be achieved simply by the actions $\hat{\mathcal{P}}_{ij} O_j = O_i \hat{\mathcal{P}}_{ij}$ with $O_i \in \{\vec{\alpha}_i, \beta_i, \mathbf{C}_i, \vec{p}_i\}$.

Application to a Reduced 2-Electron System

Introducing the coordinate transformation

$$\vec{R}_{12} = \frac{1}{2}(\vec{r}_1 + \vec{r}_2), \quad \vec{r}_{12} = \vec{r}_1 - \vec{r}_2, \quad \vec{P}_{12} = \vec{p}_1 + \vec{p}_2, \quad \vec{p}_{12} = \frac{1}{2}(\vec{p}_1 - \vec{p}_2),$$

the Hamiltonian (39) can be expanded in powers of the inter-electronic distance r_{12}. The lowest order $\hat{h}_{12}$ reads

$$\hat{h}_{12} = c\vec{\sigma}_1 \cdot \vec{p}_{12}\mathbf{C}_1 - c\vec{\sigma}_2 \cdot \vec{p}_{12}\mathbf{C}_2 + V_C\mathbf{E} + V_B\mathbf{C}_{12}, \tag{74}$$

which determines the lowest power ν of the wave function Ψ in r_{12}. With this form of $\hat{h}_{12}$, we can readily arrive at the following commutation relations which play a key role in deriving the lowest power ν (for more details, see Ref. [5]).

Proposition 2. *The operators* $\{\hat{h}_{12}, \mathbf{C}_{12}, \hat{\mathcal{P}}_{12}, \hat{\mathcal{I}}, \hat{j}_{12}^2, \hat{j}_{12,z}\}$ *are mutually commutative and hence share the common eigenfunctions.*

Proof. Since all the operators have explicit forms, the proof can readily be accomplished by direct algebraic manipulations. Here we just take $[\hat{\mathcal{P}}_{12}, \hat{h}_{12}] = 0$ as an example. Substituting Eqs. (44) and (74) into $\hat{\mathcal{P}}_{12}\hat{h}_{12}$, we have

$$\begin{aligned}
\hat{\mathcal{P}}_{12}\hat{h}_{12} &= (\hat{\pi}_{12}\hat{\Pi}_{12}\mathbf{\Pi}_{12})(c\vec{\sigma}_1 \cdot \vec{p}_{12}\mathbf{C}_1 - c\vec{\sigma}_2 \cdot \vec{p}_{12}\mathbf{C}_2 + V_C\mathbf{E} + V_B\mathbf{C}_{12}) \\
&= \hat{\pi}_{12}(c\hat{\Pi}_{12}\vec{\sigma}_1 \cdot \vec{p}_{12}\mathbf{\Pi}_{12}\mathbf{C}_1 - c\hat{\Pi}_{12}\vec{\sigma}_2 \cdot \vec{p}_{12}\mathbf{\Pi}_{12}\mathbf{C}_2 \\
&\quad + V_C\hat{\Pi}_{12}\mathbf{\Pi}_{12}\mathbf{E} + \hat{\Pi}_{12}V_B\mathbf{\Pi}_{12}\mathbf{C}_{12}) \qquad (75) \\
&= \hat{\pi}_{12}(c\hat{\Pi}_{12}\vec{\sigma}_1 \cdot \vec{p}_{12}\mathbf{C}_2\mathbf{\Pi}_{12} - c\hat{\Pi}_{12}\vec{\sigma}_2 \cdot \vec{p}_{12}\mathbf{C}_1\mathbf{\Pi}_{12} \\
&\quad + V_C\hat{\Pi}_{12}\mathbf{E}\mathbf{\Pi}_{12} + \hat{\Pi}_{12}V_B\mathbf{C}_{12}\mathbf{\Pi}_{12}) \qquad (76) \\
&= \hat{\pi}_{12}(c\vec{\sigma}_2 \cdot \vec{p}_{12}\hat{\Pi}_{12}\mathbf{C}_2\mathbf{\Pi}_{12} - c\vec{\sigma}_1 \cdot \vec{p}_{12}\hat{\Pi}_{12}\mathbf{C}_1\mathbf{\Pi}_{12} \\
&\quad + V_C\hat{\Pi}_{12}\mathbf{E}\mathbf{\Pi}_{12} + V_B\hat{\Pi}_{12}\mathbf{C}_{12}\mathbf{\Pi}_{12}) \qquad (77) \\
&= \hat{\pi}_{12}(c\vec{\sigma}_2 \cdot \vec{p}_{12}\mathbf{C}_2\hat{\Pi}_{12}\mathbf{\Pi}_{12} - c\vec{\sigma}_1 \cdot \vec{p}_{12}\mathbf{C}_1\hat{\Pi}_{12}\mathbf{\Pi}_{12} \\
&\quad + V_C\mathbf{E}\hat{\Pi}_{12}\mathbf{\Pi}_{12} + V_B\mathbf{C}_{12}\hat{\Pi}_{12}\mathbf{\Pi}_{12}) \qquad (78) \\
&= (-c\vec{\sigma}_2 \cdot \vec{p}_{12}\mathbf{C}_2 + c\vec{\sigma}_1 \cdot \vec{p}_{12}\mathbf{C}_1 \\
&\quad + V_C\mathbf{E} + V_B\mathbf{C}_{12})(\hat{\pi}_{12}\hat{\Pi}_{12}\mathbf{\Pi}_{12}) = \hat{h}_{12}\hat{\mathcal{P}}_{12}, \qquad (79)
\end{aligned}$$

where we have applied Eqs. (53)–(55) in (76); Eqs. (48), (49), and (52) in (77); Eq. (36) in (78) as well as the relation $\hat{\pi}_{12}\vec{p}_{12} = -\vec{p}_{12}\hat{\pi}_{12}$ in (79). □

It follows that the eigenfunctions of the reduced Hamiltonian $\hat{h}_{12}$ (74) can be classified according to the quantum numbers of the operators $\mathbf{C}_{12}$, $\hat{\mathcal{P}}_{12}$, $\hat{\mathcal{I}}$, $\hat{j}_{12}^2$, and $\hat{j}_{12,z}$. For more details, see Ref. [5].

Summary

It has been shown that, in the language of matrix rings, a relativistic many-electron Hamiltonian can systematically be constructed from the one-electron Dirac operator and the electron-electron interaction through the Tracy-Singh product for partitioned matrices. A 4-spinor product "$\psi_i(1)\psi_j(2)\ldots\psi_k(N)$" should also be understood as a Tracy-Singh product $\psi_i(1) \circ \psi_j(2) \circ \cdots \circ \psi_k(N)$. Permuting the labeling of the 4-spinors rather than the spatial coordinates of the electrons then leads to a determinant that has the well-known block structure. Such algebraic representation provides a very useful and convenient mathematical tool for revealing the internal symmetries of relativistic wave functions. However, it must be emphasized that what have been discussed here are purely mathematical properties of the first quantized many-electron Dirac equation which has some inherent unphysical ingredients. To get the physics right, one has to go to a Fock space formulation that incorporates properly the charge conjugation symmetry [4].

Acknowledgements The research of this work was supported by grants from the National Natural Science Foundation of China (Project Nos. 21033001, 21273011, 21290192, and 11471025).

References

1. Dirac PAM (1928) The quantum theory of the electron. Proc R Soc Lond A 117:610
2. Liu W (2012) Perspectives of relativistic quantum chemistry: the negative energy cat smiles. Phys Chem Chem Phys 14:35
3. Liu W, Lindgren I (2013) Going beyond "no-pair relativistic quantum chemistry". J Chem Phys 139:014108
4. Liu W (2014) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59
5. Li Z, Shao S, Liu W (2012) Relativistic explicit correlation: coalescence conditions and practical suggestions. J Chem Phys 136:144117
6. Kato T (1957) On the eigenfunctions of many-particle systems in quantum mechanics. Commun Pure Appl Math 10:151
7. Kutzelnigg W (1984) Basis set expansion of the Dirac operator without variational collapse. Int J Quantum Chem 25:107
8. Lewin M, Séré E (2010) Spectral pollution and how to avoid it. Proc Lond Math Soc 100:864
9. Lin L, Shao S, E W (2013) Efficient iterative method for solving the Dirac-Kohn-Sham density functional theory. J Comput Phys 245:205
10. Thaller B (1992) The Dirac equation. Springer, Berlin

11. Dereziński J (2012) Open problems about many-body Dirac operators. IAMP News Bull Jan 2012:11
12. Tracy DS, Singh RP (1972) A new matrix product and its applications in partitioned matrix differentiation. Stat Neerl 26:143
13. Liu S (1999) Matrix results on the Khatri-Rao and Tracy-Singh products. Linear Algebra Appl 289:267

Coalescence Conditions of Relativistic Wave Functions

16

Sihong Shao, Zhendong Li, and Wenjian Liu

Contents

Introduction . . . 498
Relativistic Electron-Nucleus Coalescence Conditions . . . 499
Relativistic Electron-Electron Coalescence Conditions . . . 502
The Reduced Two-Electron Problem . . . 502
Symmetries of the Reduced Hamiltonian $\hat{h}_{12}$ . . . 506
Asymptotic Behaviors Determined by the Zeroth-Order Condition (41) . . . 512
Algebraic Equations for $\Psi^{A,e}_{(+,+)}$ . . . 516
Algebraic Equations for $\Psi^{A,e}_{(+,-)}$ . . . 519
Algebraic Equations for $\Psi^{A,o}_{(+,-)}$ . . . 519
Algebraic Equations for $\Psi^{A,o}_{(+,+)}$ . . . 521
Asymptotic Behaviors Determined by the First-Order Condition (42) . . . 522
$\Psi^{A,e}_{(+,+)}$ with $f_1^{LL(0)} = 0$ for the DCG Hamiltonian . . . 522
$\Psi^{A,o}_{(+,+)}$ for the DCG and DCB Hamiltonians . . . 524
Discussion . . . 525
Summary . . . 528
References . . . 529

S. Shao
LMAM and School of Mathematical Sciences, Peking University, Beijing, China
e-mail: sihong@math.pku.edu.cn

Z. Li • W. Liu
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: zhendongli2008@gmail.com; liuwjbdf@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_8

Abstract

The electron-electron coalescence conditions for the wave functions of the Dirac-Coulomb (DC), Dirac-Coulomb-Gaunt (DCG), and Dirac-Coulomb-Breit (DCB) Hamiltonians are analyzed by making use of the internal symmetries of the reduced two-electron systems. The results show that, at the coalescence point of two electrons, the wave functions of the DCG Hamiltonian are regular, while those of the DC and DCB Hamiltonians have r_{12}^{ν}-type weak singularities, with ν being negative and of $\mathcal{O}(\alpha^2)$. Yet, such asymptotic expansions of the relativistic wave functions are only valid within an extremely small convergence radius R_c of $\mathcal{O}(\alpha^2)$. Beyond this radius, the behaviors of the relativistic wave functions are still dominated by the nonrelativistic limit.

Keywords

Relativistic wave function • Coalescence condition • Internal symmetry

Introduction

Although finding explicit solutions to systems of interacting electrons is virtually impossible (in fact, finding accurate numerical solutions is already an arduous task), deriving some asymptotic behaviors (e.g., the electron-nucleus and electron-electron coalescence conditions) of the exact wave functions is well possible. Such information not only enriches our understandings of the exact wave functions but can also be employed to accelerate the convergence of orbital-based methods for constructing numerical wave functions. This was demonstrated already in 1929 by Hylleraas [1] for the ground state of helium. However, his ansatz was motivated by the observation that the helium ^1S state depends only on the shape of the electron-nucleus triangle, rather than by a consideration of the Coulomb singularity. In 1957, Kato [2] found that the ^1S state of the Schrödinger-Coulomb Hamiltonian must satisfy the following correlation cusp condition:

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

Here, the subscript "av" represents the average over the angular part of the relative coordinate $\vec{r}_{12}$. This condition arises directly from the requirement that the divergence of the Coulomb interaction at the electron-electron coalescence point ($r_{12} = 0$) should precisely be compensated by the local kinetic energy, so as to result in a finite local energy. The integrated form of Eq. (1) is

$$\Psi = \left(1+\frac{1}{2}r_{12}\right)\Psi(r_{12}=0)+\cdots. \tag{2}$$

It follows that, for singlet coupled electrons, the exact wave function is linear in r_{12} and hence has discontinuous first-order derivative around the coalescence point.

More general cusp conditions for molecular systems were derived by Pack and Brown [3] in 1966, viz.,

$$\left(\frac{\partial^{l+1}\Psi_S}{\partial r_{12}^{l+1}}\right)_{r_{12}=0} = \frac{1}{2}\left(\frac{\partial^{l}\Psi_S}{\partial r_{12}^{l}}\right)_{r_{12}=0}, \tag{3}$$

which covers Eq. (1) as a special case ($l = 0$). For triplet-coupled electrons, the exact wave function is antisymmetric with respect to the interchange of the spatial coordinates of the two coalescing electrons, such that only odd l enter Eq. (3), with the lowest l being $l = 1$. The derivative discontinuity then appears at second order, with the integrated form of Eq. (3) being [4]

$$\Psi = \vec{r}_{12} \cdot \left.\frac{\partial \Psi}{\partial \vec{r}_{12}}\right|_{r_{12}=0} \left(1 + \frac{1}{4} r_{12}\right) + \cdots . \tag{4}$$

Based on such analytic structures of the exact wave functions, a bunch of explicitly correlated wave function methods have been developed in the last decades [5–9]. In particular, by augmenting the conventional excitations into products of unoccupied one-electron orbitals by just a few explicitly correlated configurations (which depend explicitly on the interelectronic distance r_{ij}) and carefully factorizing the difficult many-electron integrals into products of one- and two-electron integrals through the resolution of the identity [10] with a complementary auxiliary basis set [11], the so-called R12/F12 methods [12, 13] have now evolved into practical tools for general molecules.

The request for relativistic explicitly correlated wave function methods for systems containing heavy atoms is even more imperative, as relativistic corrections converge more slowly with respect to the basis set size than nonrelativistic correlation energies [14–16]. However, at variance with the significant advances in nonrelativistic explicitly correlated methods, the development of the relativistic counterparts lags far behind. The increased complexities and the reduced symmetries certainly result in substantial technical difficulties but which are not really an issue. Rather, it is the lack of knowledge on the analytic structures of the relativistic wave functions that has been the major obstacle. The first analysis of the relativistic wave functions was made by Kutzelnigg [17]. Yet, a complete analysis was performed only rather recently [18]. The essential results are to be summarized here.

Relativistic Electron-Nucleus Coalescence Conditions

The Dirac equation for an electron subject to the attraction of a clamped point nucleus Z reads

$$\hat{h}^D \psi = E\psi, \tag{5}$$

$$\hat{h}^D = c\vec{\alpha} \cdot \vec{p} + \beta c^2 + \phi. \tag{6}$$

Here, $\vec{\alpha}$ and β are the usual 4-by-4 Dirac matrices, $c = \alpha^{-1}$ is the speed of light, $\vec{p} = -i\vec{\nabla}$ is the linear momentum operator, and $\phi(r) = -Z/r$ represents the nuclear attraction. Equation (5) can also be rewritten in block form:

$$\begin{pmatrix} \phi + c^2 & c\vec{\sigma}\cdot\vec{p} \\ c\vec{\sigma}\cdot\vec{p} & \phi - c^2 \end{pmatrix} \begin{pmatrix} \psi^L \\ \psi^S \end{pmatrix} = E \begin{pmatrix} \psi^L \\ \psi^S \end{pmatrix}. \tag{7}$$

Like the one-electron Schrödinger equation, Eq. (7) can be solved analytically [19] with the solution being of the form

$$\begin{pmatrix} \psi^L \\ \psi^S \end{pmatrix} = \begin{pmatrix} f^L(r)\mathcal{Y}_{\kappa,m}(\theta,\phi) \\ f^S(r)\mathcal{Y}_{-\kappa,m}(\theta,\phi) \end{pmatrix}, \tag{8}$$

where $\mathcal{Y}_{\kappa,m}(\theta,\phi)$ is a spherical spinor with angular momentum quantum numbers (κ, m), while $f^L(r)$ and $f^S(r)$ are the radial large and small components of the wave function. By using the relation

$$\vec{\sigma}\cdot\vec{p}(f^X(r)\mathcal{Y}_{\kappa,m}) = \mathbb{i}\left(\frac{\mathrm{d}}{\mathrm{d}r} + \frac{\kappa+1}{r}\right) f^X(r)\mathcal{Y}_{-\kappa,m}, \quad X \in \{L, S\}, \tag{9}$$

Eq. (7) can be recast into a pair of coupled radial equations:

$$\left(-\frac{Z}{r} - E + c^2\right) f^L(r) + \mathbb{i}c\left(\frac{\mathrm{d}}{\mathrm{d}r} + \frac{-\kappa+1}{r}\right) f^S(r) = 0, \tag{10}$$

$$\mathbb{i}c\left(\frac{\mathrm{d}}{\mathrm{d}r} + \frac{\kappa+1}{r}\right) f^L(r) + \left(-\frac{Z}{r} - E - c^2\right) f^S(r) = 0. \tag{11}$$

To derive the electron-nucleus coalescence conditions, we expand the radial functions in a power series of the electron-nucleus distance r

$$f^X(r) = r^\nu \left(f^{X(0)} + f^{X(1)}r + f^{X(2)}r^2 + \cdots\right), \quad X \in \{L, S\}. \tag{12}$$

Inserting these expressions into Eqs. (10) and (11) leads to a set of equations for the amplitudes $f^{X(k)}$. The lowest order in r is a matrix equation

$$\mathbf{M}\begin{pmatrix} f^{L(0)} \\ f^{S(0)} \end{pmatrix} = \begin{pmatrix} -Z & \mathbb{i}c(\nu-\kappa+1) \\ \mathbb{i}c(\nu+\kappa+1) & -Z \end{pmatrix} \begin{pmatrix} f^{L(0)} \\ f^{S(0)} \end{pmatrix} = 0, \tag{13}$$

which has nontrivial solutions if and only if $\det(\mathbf{M}) = 0$, from which the power ν can be determined as

$$\nu = \sqrt{\kappa^2 - (Z\alpha)^2} - 1. \tag{14}$$

The value of ν in Eq. (14) with a negative sign in front of the square root must be discarded, because otherwise the corresponding wave function would not be normalizable. The relation between $f^{S(0)}$ and $f^{L(0)}$ can then found to be

$$f^{S(0)} = \frac{-\mathrm{i}Z\alpha}{\nu + 1 - \kappa} f^{L(0)}. \tag{15}$$

It follows that the small and large radial components possess the same lowest-order ν (14) in r. Similarly, the equations for the first-order amplitudes $f^{X(1)}$ read

$$\left[-Zf^{L(1)} + (-E + c^2) f^{L(0)}\right] + \mathrm{i}c(\nu - \kappa + 2) f^{S(1)} = 0, \tag{16}$$

$$\mathrm{i}c(\nu + \kappa + 2) f^{L(1)} + \left[-Zf^{S(1)} - (E + c^2) f^{S(0)}\right] = 0, \tag{17}$$

which, together with Eq. (15), give rise to the following relation between $f^{L(1)}$ and $f^{L(0)}$:

$$f^{L(1)} = \frac{-Z[1 + \alpha^2 E(2\nu + 3 - 2\kappa)]}{(\nu + 1 - \kappa)(2\nu + 3)} f^{L(0)}, \tag{18}$$

which is no longer universal but energy dependent. In the nonrelativistic limit (nrl), i.e., $\nu \to |\kappa| - 1$ and $E = c^2 + E'$ with $E' \sim \mathcal{O}(\alpha^0)$, Eq. (18) reduces to

$$f^{L(1)} = \left(\frac{-2Z(1 - \kappa + |\kappa|)}{(|\kappa| - \kappa)(1 + 2|\kappa|)} + \mathcal{O}(\alpha^2)\right) f^{L(0)}. \tag{19}$$

For $\kappa = -(l + 1) < 0$, i.e., $s_{1/2}$, $p_{3/2}$, $d_{5/2}$, $f_{7/2}$ etc., Eq. (19) leads to

$$\frac{f^{L(1)}}{f^{L(0)}} = -\frac{Z}{l + 1} + \mathcal{O}(\alpha^2), \tag{20}$$

the first term of which is just the well-known nonrelativistic electron-nucleus cusp condition [3]. Note, however, that Eq. (19) does not apply to the case of $\kappa = l > 0$, i.e., $p_{1/2}$, $d_{3/2}$, $f_{5/2}$, etc., for the denominator $|\kappa| - \kappa$ vanishes. As a matter of fact, in this case, $f^{L(0)}$ is of $\mathcal{O}(\alpha^2)$ according to the exact solution [19] and hence vanishes in the nrl. The first nonvanishing term in the nrl will be $f^{L(1)}$, such that it is the ratio $f^{L(2)}/f^{L(1)}$ that should be explored. The derivation is similar, with the result also of form (20).

It is interesting to see that, for $\kappa < 0$, Eq. (15) leads to

$$\lim_{c\to\infty} \lim_{r\to 0} \frac{\psi^S}{\psi^L} = \lim_{c\to\infty} \frac{f^{S(0)}}{f^{L(0)}} = 0 = \lim_{r\to 0} \lim_{c\to\infty} \frac{\psi^S}{\psi^L}. \tag{21}$$

That is, the two limits of $r_{12} \to 0$ and $c \to \infty$ commute for the ratio ψ^S/ψ^L ($\kappa < 0$). However, this does not hold for the individual components. Consider the $1s_{\frac{1}{2}}$ state of a hydrogenic ion, whose radial components are

$$\psi^L = \sqrt{\frac{(\nu + 2)Z}{\Gamma(2\nu + 3)}} (2Z)^{\nu+1} r^\nu e^{-Zr}, \psi^S = -\mathrm{i}\sqrt{\frac{-\nu Z}{\Gamma(2\nu + 3)}} (2Z)^{\nu+1} r^\nu e^{-Zr}, \tag{22}$$

from which we obtain

$$\lim_{r\to 0}\lim_{c\to\infty}\psi^L = \lim_{r\to 0} 2Z^{\frac{3}{2}}e^{-Zr} = 2Z^{\frac{3}{2}}, \quad \lim_{c\to\infty}\lim_{r\to 0}\psi^L = \infty, \tag{23}$$

$$\lim_{r\to 0}\lim_{c\to\infty}\psi^S = 0, \quad \lim_{c\to\infty}\lim_{r\to 0}\psi^S = \infty. \tag{24}$$

It will be shown later on that the two limits do not commute even for the ratio Ψ^{SS}/Ψ^{LL} between the small-small and large-large components of two-electron wave functions, observed first by Kutzelnigg [17].

Relativistic Electron-Electron Coalescence Conditions

The Reduced Two-Electron Problem

The first-quantized N-electron Dirac equation reads

$$\hat{H}\Psi(1,2,\ldots,N) = E\Psi(1,2,\ldots,N), \tag{25}$$

$$\hat{H} = \sum_k \hat{h}_k^D + \sum_{k>l}\hat{g}_{kl}, \tag{26}$$

$$\hat{h}_k^D = c\vec{\alpha}_k\cdot\vec{p}_k + \beta_k c^2 + \phi_k, \tag{27}$$

$$\phi_k = -\sum_A \frac{Z_A}{r_{kA}}, \quad r_{kA} = |\vec{r}_k - \vec{r}_A|. \tag{28}$$

Here, $\hat{h}_k^D$ is the one-electron Dirac operator (6) for electron k subject to the nuclear attraction ϕ_k (28). The electron-electron interaction operator $\hat{g}_{kl}$ can be put into a generic form

$$\hat{g}_{kl} = d_C\frac{1}{r_{kl}} + d_G\frac{\vec{\alpha}_k\cdot\vec{\alpha}_l}{r_{kl}} + d_R\frac{(\vec{\alpha}_k\cdot\hat{r}_{kl})(\vec{\alpha}_l\cdot\hat{r}_{kl})}{r_{kl}} \tag{29}$$

with $\hat{r}_{kl} = \vec{r}_{kl}/r_{kl}$. The Coulomb, Coulomb-Gaunt, and Coulomb-Breit interactions are represented by $\hat{g}_{kl}$ with $(1,0,0)$, $(1,-1,0)$, and $(1,-1/2,-1/2)$ for the coefficients (d_C, d_G, d_R), respectively. The corresponding Hamiltonians $\hat{H}$ (26) are then called Dirac-Coulomb (DC), Dirac-Coulomb-Gaunt (DCG), and Dirac-Coulomb-Breit (DCB), respectively. To investigate the electron-electron coalescence conditions, suffice it to concentrate only on the relative motion of two electrons at small interelectronic distances. For this purpose, the coordinates $\vec{r}_1$ and $\vec{r}_2$ are transformed to the center of mass ($\vec{R}_{12}$) and the relative ($\vec{r}_{12}$) coordinates of two coalescing electrons, viz.,

$$\vec{R}_{12} = \frac{1}{2}(\vec{r}_1 + \vec{r}_2), \quad \vec{r}_{12} = \vec{r}_1 - \vec{r}_2. \tag{30}$$

The corresponding momenta read

$$\vec{P}_{12} = \vec{p}_1 + \vec{p}_2, \quad \vec{p}_{12} = \frac{1}{2}(\vec{p}_1 - \vec{p}_2). \tag{31}$$

In terms of such transformations, Eq. (25) can be rewritten as

$$\hat{h}_{12}\Psi = \hat{W}\Psi, \tag{32}$$

where

$$\hat{h}_{12} = \hat{t}_{12} + \hat{g}_{12}, \quad \hat{t}_{12} = c(\vec{\alpha}_1 - \vec{\alpha}_2) \cdot \vec{p}_{12}, \tag{33}$$

and $\hat{W}$ contains all the remaining terms:

$$\hat{W} = E - \left(\sum_{k\geq 3} \hat{h}_k^D + \sum_{k>l\geq 3} \hat{g}_{kl} \right) - \sum_{k=1,2} \left(\frac{1}{2} c\vec{\alpha}_k \cdot \vec{P}_{12} + \beta_k c^2 \right) - \sum_{k=1,2} \left(\phi_k + \sum_{l\geq 3} \hat{g}_{kl} \right). \tag{34}$$

The operators $\hat{t}_{12}$ (33) and $\hat{g}_{12}$ (29) describe the relative kinetic energy and the interaction energy of electrons 1 and 2, respectively, while $\hat{W}$ describes screenings on the interaction between the two electrons due to the rest of the system, including the electrostatic interaction of the two electrons with the rest of the system, the kinetic and potential terms of the other electrons, as well as the kinetic energy of the center of mass motion.

Consider the region of configuration space where electrons 1 and 2 are close together but all the other electrons and nuclei are well separated from these two electrons and from each other, viz., $0 \leq r_{12} \leq \epsilon$ and $r_{RA} = |\vec{R}_{12} - \vec{r}_A|$, $r_{Rk} = |\vec{R}_{12} - \vec{r}_k|$, $r_{kl} = |\vec{r}_k - \vec{r}_l| \gg \epsilon$ for $k, l \geq 3$. Here ϵ is an arbitrary small positive number. Within this region, the wave function Ψ can be expanded in powers of r_{12} as

$$\Psi = \Psi^{(\nu)} + \Psi^{(\nu+1)} + \cdots, \tag{35}$$

where ν is the lowest power of nonvanishing Ψ in r_{12}. The operators $\hat{h}_{12}$ (33) and $\hat{W}$ (34) can also be expanded in the same way. In particular, both $\hat{t}_{12}$ and $\hat{g}_{12}$ in $\hat{h}_{12}$

lower the power of r_{12} by one and can hence be labeled as $\hat{t}_{12}^{(-1)}$ and $\hat{g}_{12}^{(-1)}$, leading to $\hat{h}_{12}^{(-1)}$. For the potential terms in Eq. (34), the following partial wave expansions can be invoked:

$$\begin{aligned}
\frac{1}{r_{1A}} &= \frac{1}{|\vec{r}_{RA} + \frac{1}{2}\vec{r}_{12}|} = \frac{1}{r_{RA}} \sum_{l=0}^{+\infty} (-1)^l \left(\frac{r_{12}}{2r_{RA}}\right)^l P_l(\cos\theta_A), \\
\frac{1}{r_{2A}} &= \frac{1}{|\vec{r}_{RA} - \frac{1}{2}\vec{r}_{12}|} = \frac{1}{r_{RA}} \sum_{l=0}^{+\infty} \left(\frac{r_{12}}{2r_{RA}}\right)^l P_l(\cos\theta_A), \\
\frac{1}{r_{1k}} &= \frac{1}{|\vec{r}_{Rk} + \frac{1}{2}\vec{r}_{12}|} = \frac{1}{r_{Rk}} \sum_{l=0}^{+\infty} (-1)^l \left(\frac{r_{12}}{2r_{Rk}}\right)^l P_l(\cos\theta_k), \\
\frac{1}{r_{2k}} &= \frac{1}{|\vec{r}_{Rk} - \frac{1}{2}\vec{r}_{12}|} = \frac{1}{r_{Rk}} \sum_{l=0}^{+\infty} \left(\frac{r_{12}}{2r_{Rk}}\right)^l P_l(\cos\theta_k),
\end{aligned} \tag{36}$$

where θ_A is the angle between $\vec{r}_{RA}$ and $\vec{r}_{12}$, θ_k the angle between $\vec{r}_{Rk}$ and $\vec{r}_{12}$, and P_l the Legendre polynomials. The operator $\hat{W}$ then becomes

$$\hat{W} = \hat{W}^{(0)} + \mathcal{O}(\epsilon^2) + \cdots, \tag{37}$$

$$\begin{aligned}
\hat{W}^{(0)} = E &- \left(\sum_{k\geq 3} \hat{h}_k^D + \sum_{k>l\geq 3} \hat{g}_{kl}\right) - \sum_{k=1,2} \left(\frac{1}{2} c\vec{\alpha}_k \cdot \vec{P}_{12} + \beta_k c^2\right) \\
&- \sum_{k=1,2} \left(\phi_k^{(0)} + \sum_{l\geq 3} \hat{g}_{kl}^{(0)}\right),
\end{aligned} \tag{38}$$

where $\phi_k^{(0)}$ and $\hat{g}_{kl}^{(0)}$ for $k = 1, 2$ arise from the s-wave ($l = 0$) terms in Eq. (36) and can be summed up as

$$\phi_1^{(0)} + \phi_2^{(0)} = -2 \sum_A \frac{Z_A}{r_{RA}} := 2\phi^{(0)}, \tag{39}$$

$$\begin{aligned}
\hat{g}_{1l}^{(0)} + \hat{g}_{2l}^{(0)} = {} & d_C \frac{2}{r_{Rl}} + d_G \frac{(\vec{\alpha}_1 + \vec{\alpha}_2) \cdot \vec{\alpha}_l}{r_{Rl}} \\
& + d_R \frac{(\vec{\alpha}_1 \cdot \hat{r}_{Rl})(\vec{\alpha}_l \cdot \hat{r}_{Rl}) + (\vec{\alpha}_2 \cdot \hat{r}_{Rl})(\vec{\alpha}_l \cdot \hat{r}_{Rl})}{r_{Rl}}.
\end{aligned} \tag{40}$$

It is important to realize that all the odd order terms $\hat{W}^{(2n+1)}$ ($n \in \mathbb{N}$) in Eq. (37) vanish identically due to the cancelation of the odd l terms for electrons 1 and 2. By collecting the terms of the same order, Eq. (32) gives rise to a set of equations, with the lowest three orders being

$$\mathcal{O}\left(\epsilon^{\nu-1}\right): \qquad \left(\hat{t}_{12}^{(-1)}+\hat{g}_{12}^{(-1)}\right)\Psi^{(\nu)}=0, \tag{41}$$

$$\mathcal{O}\left(\epsilon^{\nu}\right): \quad \left(\hat{t}_{12}^{(-1)}+\hat{g}_{12}^{(-1)}\right)\Psi^{(\nu+1)}=\hat{W}^{(0)}\Psi^{(\nu)}, \tag{42}$$

$$\mathcal{O}\left(\epsilon^{\nu+1}\right): \left(\hat{t}_{12}^{(-1)}+\hat{g}_{12}^{(-1)}\right)\Psi^{(\nu+2)}=\hat{W}^{(0)}\Psi^{(\nu+1)}. \tag{43}$$

The zeroth-order coalescence condition (41) (NB: the ordering is relative to the leading term $\Psi^{(\nu)}$ in Eq. (35) and is determined by the largest m value of $\Psi^{(\nu+m)}$ in the considered equation) is essential for ensuring that the local energy

$$E_L=\frac{\hat{H}\Psi}{\Psi} \tag{44}$$

remains finite at the coalescence point. This can readily be understood as follows:

$$\lim_{r_{12}\to 0} E_L=\lim_{r_{12}\to 0}\frac{(\hat{h}_{12}-\hat{W}+E)\Psi}{\Psi} \tag{45}$$

$$=\lim_{r_{12}\to 0}\frac{\hat{h}_{12}^{(-1)}\Psi^{(\nu)}+\left[\hat{h}_{12}^{(-1)}\Psi^{(\nu+1)}+(-\hat{W}^{(0)}+E)\Psi^{(\nu)}\right]+\cdots}{\Psi^{(\nu)}+\Psi^{(\nu+1)}+\cdots} \tag{46}$$

$$=\lim_{r_{12}\to 0}\frac{\hat{h}_{12}^{(-1)}\Psi^{(\nu)}}{\Psi^{(\nu)}}+\lim_{r_{12}\to 0}\left(E+\frac{\hat{h}_{12}^{(-1)}\Psi^{(\nu+1)}-\hat{W}^{(0)}\Psi^{(\nu)}}{\Psi^{(\nu)}}\right)+\cdots, \tag{47}$$

where Eq. (45) arises from Eq. (44) and the relation $\hat{H}=\hat{h}_{12}-\hat{W}+E$. The first term of Eq. (47) is of $\mathcal{O}(r_{12}^{-1})$ and hence will diverge if the wave function does not satisfy condition (41) properly.

At this stage, it is instructive to compare Eqs. (41)–(43) with the nonrelativistic counterparts:

$$\mathcal{O}\left(\epsilon^{\nu-2}\right): \qquad \hat{t}_{12}^{S(-2)}\Psi_S^{(\nu)}=0, \tag{48}$$

$$\mathcal{O}\left(\epsilon^{\nu-1}\right): \qquad \hat{t}_{12}^{S(-2)}\Psi_S^{(\nu+1)}+\hat{g}_{12}^{S(-1)}\Psi_S^{(\nu)}=0, \tag{49}$$

$$\mathcal{O}\left(\epsilon^{\nu}\right): \quad \hat{t}_{12}^{S(-2)}\Psi_S^{(\nu+2)}+\hat{g}_{12}^{S(-1)}\Psi_S^{(\nu+1)}=\hat{W}^{S(0)}\Psi_S^{(\nu)}, \tag{50}$$

$$\mathcal{O}\left(\epsilon^{\nu+1}\right): \hat{t}_{12}^{S(-2)}\Psi_S^{(\nu+3)}+\hat{g}_{12}^{S(-1)}\Psi_S^{(\nu+2)}=\hat{W}^{S(0)}\Psi_S^{(\nu+1)}, \tag{51}$$

where

$$\hat{t}_{12}^{S(-2)}=\vec{p}_{12}^{\,2}, \quad \hat{g}_{12}^{S(-1)}=\frac{1}{r_{12}}, \tag{52}$$

$$\hat{W}^{S(0)}=E-\left(\sum_{k\geq 3}\hat{h}_k^S+\sum_{k>l\geq 3}\hat{g}_{kl}^S\right)-\frac{1}{4}\vec{P}_{12}^{\,2}-\sum_{k=1,2}\left(\phi_k^{(0)}+\sum_{l\geq 3}\hat{g}_{kl}^{S(0)}\right). \tag{53}$$

The most important difference in between lies in that the nonrelativistic kinetic energy operator $\hat{t}_{12}^{S(-2)}$ is a second-order differential operator, while the relativistic one, $\hat{t}_{12}^{(-1)}$, is only first order. Consequently, the lowest-order equation (48) for the Schrödinger equation is of $\mathcal{O}(\epsilon^{\nu-2})$, which determines the asymptotic behavior of the wave function $\Psi_S^{(\nu)}$ as $r_{12}^l Y_l^{m_l}$, with $\nu = l$ and $Y_l^{m_l}$ being the spherical harmonics. The next-order equation (49) involves $\hat{g}_{12}^{S(-1)}$, whose singularity results in discontinuous $(l+1)$-th-order derivatives characterized by the cusp condition (3). Both conditions (48) and (49) have to be satisfied for a finite local energy (44) of the Schödinger equation. The next two-order equations (50) and (51) can be employed to derive the second- and third-order coalescence conditions [4] that are no longer universal but system and state dependent and vary throughout configuration space due to the involvement of the $\hat{W}^{S(0)}$ operator. In contrast, in the relativistic case, the lowest-order equation (41) is only of $\mathcal{O}(\epsilon^{\nu-1})$ and already involves the singular term $\hat{g}_{12}^{(-1)}$. At variance with the universality of the first-order nonrelativistic cusp condition (3), the relativistic counterpart, i.e., the relation between $\Psi^{(\nu+1)}$ and $\Psi^{(\nu)}$ determined by Eq. (42), cannot be universal due to the appearance of $\hat{W}^{(0)}$ (38). Note also that the $\hat{W}^{(0)}$ operator (38) is more complicated than the nonrelativistic counterpart $\hat{W}^{S(0)}$ (53) for the former does not commute with all the symmetry operations of $\hat{h}_{12}$ (vide post). This will result in great difficulties in manipulating Eqs. (42) and (43).

Another significant difference in between lies in that the relativistic wave function Ψ in Eq. (32) has 16 components depending on the relative coordinate $\vec{r}_{12}$ and the spin degrees of freedom, while the nonrelativistic one Ψ_S is simply a scalar function. As such, the relativistic local energy (44) is also a multicomponent function, with each component being the ratio between the corresponding components of $\hat{H}\Psi$ and Ψ.

Given the great complexities, the asymptotic behaviors of the relativistic wave functions can still be obtained from conditions (41) and (42) by making use of the internal symmetries of the reduced Hamiltonian $\hat{h}_{12}$ (33).

Symmetries of the Reduced Hamiltonian $\hat{h}_{12}$

Equation (41) can be rewritten in block form

$$\begin{pmatrix} V_C & -c\vec{\sigma}_2\cdot\vec{p}_{12} & c\vec{\sigma}_1\cdot\vec{p}_{12} & V_B \\ -c\vec{\sigma}_2\cdot\vec{p}_{12} & V_C & V_B & c\vec{\sigma}_1\cdot\vec{p}_{12} \\ c\vec{\sigma}_1\cdot\vec{p}_{12} & V_B & V_C & -c\vec{\sigma}_2\cdot\vec{p}_{12} \\ V_B & c\vec{\sigma}_1\cdot\vec{p}_{12} & -c\vec{\sigma}_2\cdot\vec{p}_{12} & V_C \end{pmatrix} \begin{pmatrix} \Psi^{LL} \\ \Psi^{LS} \\ \Psi^{SL} \\ \Psi^{SS} \end{pmatrix} = 0, \tag{54}$$

where each block Ψ^{XY} $(X, Y \in \{L, S\})$ is a four-component function. For simplicity, the superscript (ν) of Ψ^{XY} has been dropped. The operators V_C and V_B read

$$V_C = d_C \frac{1}{r_{12}}, \quad V_B = d_G \frac{\vec{\sigma}_1 \cdot \vec{\sigma}_2}{r_{12}} + d_R \frac{(\vec{\sigma}_1 \cdot \hat{r}_{12})(\vec{\sigma}_2 \cdot \hat{r}_{12})}{r_{12}}. \tag{55}$$

To reveal the symmetry properties of $\hat{h}_{12}$, we rewrite it as [18, 20]

$$\hat{h}_{12} = c\vec{\sigma}_1 \cdot \vec{p}_{12}\mathbf{C}_1 - c\vec{\sigma}_2 \cdot \vec{p}_{12}\mathbf{C}_2 + V_C\mathbf{E} + V_B\mathbf{C}_{12} \tag{56}$$

in terms of the following "block operators" that merely interchange the blocks Ψ^{XY} of the wave function:

$$\mathbf{E} = \mathbf{I}_4 \circ \mathbf{I}_4 = \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} \circ \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} = \begin{pmatrix} I_4 & 0 & 0 & 0 \\ 0 & I_4 & 0 & 0 \\ 0 & 0 & I_4 & 0 \\ 0 & 0 & 0 & I_4 \end{pmatrix}, \tag{57}$$

$$\mathbf{C}_1 = \boldsymbol{\gamma}^5 \circ \mathbf{I}_4 = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix} \circ \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & I_4 & 0 \\ 0 & 0 & 0 & I_4 \\ I_4 & 0 & 0 & 0 \\ 0 & I_4 & 0 & 0 \end{pmatrix}, \tag{58}$$

$$\mathbf{C}_2 = \mathbf{I}_4 \circ \boldsymbol{\gamma}^5 = \begin{pmatrix} I_2 & 0 \\ 0 & I_2 \end{pmatrix} \circ \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & I_4 & 0 & 0 \\ I_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_4 \\ 0 & 0 & I_4 & 0 \end{pmatrix}, \tag{59}$$

$$\mathbf{C}_{12} = \boldsymbol{\gamma}^5 \circ \boldsymbol{\gamma}^5 = \mathbf{C}_1\mathbf{C}_2 = \mathbf{C}_2\mathbf{C}_1 = \begin{pmatrix} 0 & 0 & 0 & I_4 \\ 0 & 0 & I_4 & 0 \\ 0 & I_4 & 0 & 0 \\ I_4 & 0 & 0 & 0 \end{pmatrix}, \tag{60}$$

where the symbol $\circ$ represents the Tracy-Singh product [21, 22], which is a generalization of the standard Kronecker product ($\otimes$) for partitioned matrices. The multiplications between the "component operators" (e.g., $c\vec{\sigma}_1 \cdot \vec{p}_{12}$) and the "block operators" (e.g., $\mathbf{C}_1$) in Eq. (56) are similar to those between numbers and matrices (for more details, see Ref. [20]). Such a formulation is particularly advantageous in that the block structure of the wave function in Eq. (54) can always be retained and the symmetry properties of $\hat{h}_{12}$ can readily be deduced. To do so, we first introduce the following Abelian group [18]:

$$\mathcal{G} = \{\mathbf{E}, \mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_{12}\}, \tag{61}$$

which is just a direct product of groups $\{\mathbf{E}, \mathbf{C}_1\}$ and $\{\mathbf{E}, \mathbf{C}_2\}$. It can then be shown [18, 20] that the following operators

$$\left\{\hat{h}_{12}, \mathbf{C}_{12}, \hat{\mathcal{P}}_{12}, \hat{\mathcal{I}}, \hat{j}_{12}^2, \hat{j}_{12,z}\right\} \tag{62}$$

are mutually commutative and hence share the same eigenfunctions. Here $\hat{\mathcal{P}}_{12}$ is the permutation operator for electrons 1 and 2 (vide post), $\hat{\mathcal{I}}$ the space inversion for the relative coordinate $\vec{r}_{12}$, and $\vec{j}_{12}$ the angular momentum

$$\vec{j}_{12} = \vec{l}_{12} + \vec{s}, \quad \vec{s} = \vec{s}_1 + \vec{s}_2, \tag{63}$$

where $\vec{l}_{12} = \vec{r}_{12} \times \vec{p}_{12}$ is the orbital angular momentum of the relative motion and $\vec{s}_k$ the spin of electron k.

To classify the solutions of Eq. (54) according to the eigenvalues of the operators (62), we first construct the eigenfunctions of $\{\hat{j}_{12}^2, \hat{j}_{12,z}\}$ (denoted as $|(ls), jm_j\rangle$) via the *LS* coupling (63),

$$|(ls), jm_j\rangle = \sum_{m_s=-s}^{+s} |lm_l\rangle|sm_s\rangle\langle lm_l sm_s|jm_j\rangle, \tag{64}$$

where $\langle lm_l sm_s|jm_j\rangle$ are the Clebsch-Gordan coefficients. Note that the quantum number s of the total spin angular momentum $\vec{s}$ can only be 0 (singlet) or 1 (triplet) for two electrons. Given j, the quantum number l of $\vec{l}_{12}$ can only be j for $s = 0$ and can be $j+1$, j, or $j-1$ for $s = 1$. For simplicity, the four possible eigenfunctions are to be denoted as Ω_i:

$$\Omega_1 = |(l = j, s = 0), jm_j\rangle, \tag{65}$$

$$\Omega_2 = |(l = j, s = 1), jm_j\rangle, \tag{66}$$

$$\Omega_3 = |(l = j - 1, s = 1), jm_j\rangle, \tag{67}$$

$$\Omega_4 = |(l = j + 1, s = 1), jm_j\rangle, \tag{68}$$

which form an orthonormal basis set for the subspace of given j and m_j. As the parity of $|(ls), jm_j\rangle$ is $(-1)^l$, the four functions Ω_i can be classified into two groups, one with parity $(-1)^j$ (i.e., $l = j$) including Ω_1 and Ω_2 and the other with parity $(-1)^{j+1}$ (i.e., $l = j \pm 1$) including Ω_3 and Ω_4. Thus, for given j and m_j, the components Ψ^{XY} of Ψ can be expressed as

$$\Psi_+^{XY} = f_1^{XY}\Omega_1 + f_2^{XY}\Omega_2, \quad X, Y \in \{L, S\} \tag{69}$$

with parity $+(-1)^j$ or as

$$\Psi_-^{XY} = f_3^{XY}\Omega_3 + f_4^{XY}\Omega_4, \quad X, Y \in \{L, S\} \tag{70}$$

with parity $-(-1)^j$. While the amplitudes f_i^{XY} are dependent not only on the radial part of $\vec{r}_{12}$ but also on the center of mass coordinate $\vec{R}_{12}$ of the two electrons as

well as all the coordinates of the rest of the system, the Ω_i depend only on the spin-angular part of the relative coordinate $\vec{r}_{12}$. Therefore, the notation ${}^{2s+1}l_j$ ($s = 0, 1$; $l = s, p, d, \ldots$; $j = 0, 1, \ldots$) for the states to be used below should not be confused with the true spectroscopic terms that involve the total angular momenta of all the electrons. By noting that Ψ^{LL} and Ψ^{SS} have the same parity and Ψ^{LS} and Ψ^{SL} also have the same parity but different from that of Ψ^{LL} and Ψ^{SS} (see Ref. [20]), the wave function Ψ for given j and m_j can be constructed as

$$\Psi_+ = \begin{pmatrix} \Psi^{LL}_+ \\ \Psi^{LS}_- \\ \Psi^{SL}_- \\ \Psi^{SS}_+ \end{pmatrix}, \quad \Psi_- = \begin{pmatrix} \Psi^{LL}_- \\ \Psi^{LS}_+ \\ \Psi^{SL}_+ \\ \Psi^{SS}_- \end{pmatrix} \tag{71}$$

with parities $+(-1)^j$ and $-(-1)^j$, respectively.

The function Ψ_+ or Ψ_- still has eight unknowns but which can further be reduced by using $\mathbf{C}_{12}$ and $\hat{\mathcal{P}}_{12}$. Since the eigenvalues of $\mathbf{C}_{12}$ and $\hat{\mathcal{P}}_{12}$ can only be $+1$ or -1, the spaces for Ψ_+ and Ψ_- can respectively be decomposed as direct sums ($\oplus$) of the eigensubspaces

$$V_+ = V^A_{(+,+)} \oplus V^S_{(+,+)} \oplus V^A_{(+,-)} \oplus V^S_{(+,-)}, \tag{72}$$

$$V_- = V^A_{(-,+)} \oplus V^S_{(-,+)} \oplus V^A_{(-,-)} \oplus V^S_{(-,-)}, \tag{73}$$

where the second subscript $+$ ($-$) indicates the corresponding eigenvalue $+1$ (-1) of $\mathbf{C}_{12}$, while the superscript A (S) indicates antisymmetric (symmetric) under the permutation $\hat{\mathcal{P}}_{12}$. In addition, the following identities [18, 20]

$$\mathbf{C}_{12}\mathbf{C}_1 = \mathbf{C}_1\mathbf{C}_{12}, \tag{74}$$

$$\hat{\mathcal{I}}\mathbf{C}_1 = -\mathbf{C}_1\hat{\mathcal{I}}, \tag{75}$$

$$\hat{\mathcal{P}}_{12}\mathbf{C}_1 = \mathbf{C}_2\hat{\mathcal{P}}_{12} = \mathbf{C}_1\mathbf{C}_{12}\hat{\mathcal{P}}_{12} \tag{76}$$

imply that an arbitrary function Ψ with eigenvalues $\{\eta(\mathbf{C}_{12}), \eta(\hat{\mathcal{I}}), \eta(\hat{\mathcal{P}}_{12})\}$ will be transformed to a function $\mathbf{C}_1\Psi$ with eigenvalues $\{\eta(\mathbf{C}_{12}), -\eta(\hat{\mathcal{I}}), \eta(\mathbf{C}_{12})\eta(\hat{\mathcal{P}}_{12})\}$. Therefore, the following relations can be established for functions in spaces V_+ and V_-:

$$V^A_{(-,+)} = \mathbf{C}_1 V^A_{(+,+)}, \quad V^S_{(-,+)} = \mathbf{C}_1 V^S_{(+,+)}, \tag{77}$$

$$V^A_{(-,-)} = \mathbf{C}_1 V^S_{(+,-)}, \quad V^S_{(-,-)} = \mathbf{C}_1 V^A_{(+,-)}. \tag{78}$$

It can then immediately be deduced that the asymptotic behavior of the wave function Ψ_- constructed as

$$\Psi_- = \mathbf{C}_1\Psi_+ = \mathbf{C}_1 \begin{pmatrix} \Psi_+^{LL} \\ \Psi_-^{LS} \\ \Psi_-^{SL} \\ \Psi_+^{SS} \end{pmatrix} = \begin{pmatrix} \Psi_-^{SL} \\ \Psi_+^{SS} \\ \Psi_+^{LL} \\ \Psi_-^{LS} \end{pmatrix} \tag{79}$$

is exactly the same as that of Ψ_+. For instance, the ${}^1s_0$ ($= \Psi_+$) and ${}^3p_0$ ($= \mathbf{C}_1\Psi_+$) states will have the same asymptotic behaviors. Note that the presentations so far hold for both two identical fermions (electrons or positrons) and an electron-positron pair. For an electronic system, only the antisymmetric parts $V^A_{(+,+)}$ and $V^A_{(+,-)}$ of V_+ (72) and $V^A_{(-,+)}$ and $V^A_{(-,-)}$ of V_- (73) are relevant. Furthermore, because of the first equalities of Eqs. (77) and (78), we need to only consider the wave function Ψ_+ belonging to $V^A_{(+,+)}$, $V^A_{(+,-)}$, and $V^S_{(+,-)}$, for the asymptotic behaviors of Ψ_- belonging to $V^A_{(-,+)}$ and $V^A_{(-,-)}$ are the same as those of Ψ_+ in $V^A_{(+,+)}$ and $V^S_{(+,-)}$, respectively.

To construct explicitly the electronic wave functions Ψ_+, we first note that the eigenfunctions of $\mathbf{C}_{12}$ are simply

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_2 \\ \varphi_1 \end{pmatrix}, \quad \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ -\varphi_2 \\ -\varphi_1 \end{pmatrix} \tag{80}$$

with eigenvalues of $+1$ and -1, respectively. Additional restrictions on the amplitudes f_i^{XY} are further imposed by the antisymmetry principle. To see this, we write the permutation operator $\hat{\mathcal{P}}_{12}$ as

$$\hat{\mathcal{P}}_{12} = \hat{\pi}_{12}\hat{\Pi}_{12}\mathbf{\Pi}_{12}, \tag{81}$$

where $\hat{\pi}_{12}$ interchanges the spatial coordinates, viz.,

$$\hat{\pi}_{12} f(\vec{r}_1, \vec{r}_2) = f(\vec{r}_2, \vec{r}_1), \quad \hat{\pi}_{12} f(\vec{r}_{12}, \vec{R}_{12}) = f(-\vec{r}_{12}, \vec{R}_{12}), \tag{82}$$

while the "component operator" $\hat{\Pi}_{12}$ acts on the blocks $\Psi^{XY}(X, Y \in \{L, S\})$ of Ψ:

$$\hat{\Pi}_{12} = \frac{1}{2}(I_4 + \vec{\sigma}_1 \cdot \vec{\sigma}_2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{83}$$

The "block operator" $\mathbf{\Pi}_{12}$ in Eq. (81) is given as

$$\mathbf{\Pi}_{12} = \begin{pmatrix} I_4 & 0 & 0 & 0 \\ 0 & 0 & I_4 & 0 \\ 0 & I_4 & 0 & 0 \\ 0 & 0 & 0 & I_4 \end{pmatrix}. \tag{84}$$

Then the antisymmetry principle $\hat{\mathcal{P}}_{12}\Psi(1,2) = -\Psi(1,2)$ dictates that

$$\hat{\pi}_{12}\hat{\Pi}_{12}\Psi^{YX}(1,2) = -\Psi^{XY}(1,2), \quad X, Y \in \{L, S\}. \tag{85}$$

Straightforward manipulations further reveal that the respective actions of $\hat{\pi}_{12}$ and $\hat{\Pi}_{12}$ on Ω_i are

$$\hat{\pi}_{12}\Omega_i = (-1)^l \Omega_i, \tag{86}$$

$$\hat{\Pi}_{12}\Omega_i = [s(s+1) - 1]\Omega_i = (-1)^{s+1}\Omega_i, \quad s \in \{0, 1\}. \tag{87}$$

We therefore have

$$\hat{\pi}_{12}\hat{\Pi}_{12}\Omega_i = (-1)^{l+s+1}\Omega_i. \tag{88}$$

That is, Ω_i is the eigenfunction of $\hat{\pi}_{12}\hat{\Pi}_{12}$ with eigenvalue $(-1)^{l+s+1}$. For Ψ_+^{XY} (69), the action of $\hat{\pi}_{12}\hat{\Pi}_{12}$ leads to

$$\hat{\pi}_{12}\hat{\Pi}_{12}\Psi_+^{XY} = (-1)^{j+1} f_1^{XY}\Omega_1 + (-1)^j f_2^{XY}\Omega_2. \tag{89}$$

In view of Eq. (85), the coefficients must satisfy

$$f_1^{YX} = (-1)^j f_1^{XY}, \quad f_2^{YX} = (-1)^{j+1} f_2^{XY}. \tag{90}$$

Therefore, f_1^{XX} is nonzero only for even j, while f_2^{XX} is nonzero only for odd j. Similarly, for Ψ_-^{XY} (70), the action of $\hat{\pi}_{12}\hat{\Pi}_{12}$ leads to

$$\hat{\pi}_{12}\hat{\Pi}_{12}\Psi_-^{XY} = (-1)^{j+1} f_3^{XY}\Omega_3 + (-1)^{j+1} f_4^{XY}\Omega_4, \tag{91}$$

such that the coefficients are subject to

$$f_3^{YX} = (-1)^j f_3^{XY}, \quad f_4^{YX} = (-1)^j f_4^{XY}. \tag{92}$$

That is, both f_3^{XX} and f_4^{XX} are nonzero only for even j. These results together with Eq. (80) lead immediately to the following forms for functions Ψ_+ in $V^A_{(+,+)}$ and $V^A_{(+,-)}$

$$\Psi^{A,e}_{(+,+)} = \begin{pmatrix} f_1^{LL}\Omega_1 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_1^{LL}\Omega_1 \end{pmatrix}, \tag{93}$$

$$\Psi^{A,o}_{(+,+)} = \begin{pmatrix} f_2^{LL}\Omega_2 \\ 0 \\ 0 \\ f_2^{LL}\Omega_2 \end{pmatrix}, \tag{94}$$

$$\Psi^{A,e}_{(+,-)} = \begin{pmatrix} f_1^{LL}\Omega_1 \\ 0 \\ 0 \\ -f_1^{LL}\Omega_1 \end{pmatrix}, \tag{95}$$

$$\Psi^{A,o}_{(+,-)} = \begin{pmatrix} f_2^{LL}\Omega_2 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ -(f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4) \\ -f_2^{LL}\Omega_2 \end{pmatrix}, \tag{96}$$

for even j (denoted by a superscript e) and odd j (denoted by a superscript o), respectively. The forms for functions in $V^S_{(+,-)}$ are the same as those in $V^A_{(+,-)}$ if the parity of j is reversed, viz.,

$$\Psi^{S,e}_{(+,-)} \sim \Psi^{A,o}_{(+,-)}, \quad \Psi^{S,o}_{(+,-)} \sim \Psi^{A,e}_{(+,-)}. \tag{97}$$

Therefore, the asymptotic behaviors of the relativistic wave functions can simply be deduced from Eqs. (93) to (96). Noticeably, the number of unknowns in Ψ has been reduced from 16 to 1 for Eqs. (94) and (95) and to 3 for Eqs. (93) and (96). These results facilitate greatly the subsequent analysis of the asymptotic behaviors.

For completeness, all the eight types of functions in Eqs. (72) and (73), i.e., the common eigenfunctions of the operators (62), are given in Table 1. internal symmetry

Asymptotic Behaviors Determined by the Zeroth-Order Condition (41)

Having determined the structures of $\Psi^{(v)}$ (93)–(96), we can now insert $\Psi^{(v)}$ into Eq. (54) and integrate out the spin-angular part Ω_i to obtain equations for the radial part f_i^{XY}. To do so, the actions of $\vec{\sigma}_k \cdot \vec{p}_{12}$, $\vec{\sigma}_1 \cdot \vec{\sigma}_2$, and $\vec{\sigma}_k \cdot \hat{r}_{12}$ on functions $f_i^{XY}\Omega_i$ have to first be evaluated. The operator identity

Table 1 Classification of two-particle relativistic wave functions as common eigenfunctions of the operator set (62). See also Eqs. (72) and (73). Superscript A or S indicates antisymmetric or symmetric under the permutation $\hat{\mathcal{P}}_{12}$. Superscript e or o refers to even or odd j. The first subscript $+$ or $-$ refers to the parity $+(-1)^j$ or $-(-1)^j$, whereas the second subscript $+$ or $-$ indicates the eigenvalue $+1$ or -1 of $\mathbf{C}_{12}$ (cf. Eq. (80)). The spin-angular functions Ω_i are defined in Eqs. (65)–(68)

Eigensubspace	Wave function	
$V^{A}_{(+,+)}$	$\Psi^{A,e}_{(+,+)} = \begin{pmatrix} f_1^{LL}\Omega_1 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_1^{LL}\Omega_1 \end{pmatrix}$	$\Psi^{A,o}_{(+,+)} = \begin{pmatrix} f_2^{LL}\Omega_2 \\ 0 \\ 0 \\ f_2^{LL}\Omega_2 \end{pmatrix}$
$V^{S}_{(+,+)}$	$\Psi^{S,e}_{(+,+)} = \begin{pmatrix} f_2^{LL}\Omega_2 \\ 0 \\ 0 \\ f_2^{LL}\Omega_2 \end{pmatrix}$	$\Psi^{S,o}_{(+,+)} = \begin{pmatrix} f_1^{LL}\Omega_1 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_1^{LL}\Omega_1 \end{pmatrix}$
$V^{A}_{(+,-)}$	$\Psi^{A,e}_{(+,-)} = \begin{pmatrix} f_1^{LL}\Omega_1 \\ 0 \\ 0 \\ -f_1^{LL}\Omega_1 \end{pmatrix}$	$\Psi^{A,o}_{(+,-)} = \begin{pmatrix} f_2^{LL}\Omega_2 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ -f_3^{LS}\Omega_3 - f_4^{LS}\Omega_4 \\ -f_2^{LL}\Omega_2 \end{pmatrix}$
$V^{S}_{(+,-)}$	$\Psi^{S,e}_{(+,-)} = \begin{pmatrix} f_2^{LL}\Omega_2 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ -f_3^{LS}\Omega_3 - f_4^{LS}\Omega_4 \\ -f_2^{LL}\Omega_2 \end{pmatrix}$	$\Psi^{S,o}_{(+,-)} = \begin{pmatrix} f_1^{LL}\Omega_1 \\ 0 \\ 0 \\ -f_1^{LL}\Omega_1 \end{pmatrix}$
$V^{A}_{(-,+)}$	$\Psi^{A,e}_{(-,+)} = \begin{pmatrix} f_3^{LL}\Omega_3 + f_4^{LL}\Omega_4 \\ f_1^{LS}\Omega_1 \\ f_1^{LS}\Omega_1 \\ f_3^{LL}\Omega_3 + f_4^{LL}\Omega_4 \end{pmatrix}$	$\Psi^{A,o}_{(-,+)} = \begin{pmatrix} 0 \\ f_2^{LS}\Omega_2 \\ f_2^{LS}\Omega_2 \\ 0 \end{pmatrix}$
$V^{S}_{(-,+)}$	$\Psi^{S,e}_{(-,+)} = \begin{pmatrix} 0 \\ f_2^{LS}\Omega_2 \\ f_2^{LS}\Omega_2 \\ 0 \end{pmatrix}$	$\Psi^{S,o}_{(-,+)} = \begin{pmatrix} f_3^{LL}\Omega_3 + f_4^{LL}\Omega_4 \\ f_1^{LS}\Omega_1 \\ f_1^{LS}\Omega_1 \\ f_3^{LL}\Omega_3 + f_4^{LL}\Omega_4 \end{pmatrix}$
$V^{A}_{(-,-)}$	$\Psi^{A,e}_{(-,-)} = \begin{pmatrix} f_3^{LL}\Omega_3 + f_4^{LL}\Omega_4 \\ f_2^{LS}\Omega_2 \\ -f_2^{LS}\Omega_2 \\ -f_3^{LL}\Omega_3 - f_4^{LL}\Omega_4 \end{pmatrix}$	$\Psi^{A,o}_{(-,-)} = \begin{pmatrix} 0 \\ f_1^{LS}\Omega_1 \\ -f_1^{LS}\Omega_1 \\ 0 \end{pmatrix}$
$V^{S}_{(-,-)}$	$\Psi^{S,e}_{(-,-)} = \begin{pmatrix} 0 \\ f_1^{LS}\Omega_1 \\ -f_1^{LS}\Omega_1 \\ 0 \end{pmatrix}$	$\Psi^{S,o}_{(-,-)} = \begin{pmatrix} f_3^{LL}\Omega_3 + f_4^{LL}\Omega_4 \\ f_2^{LS}\Omega_2 \\ -f_2^{LS}\Omega_2 \\ -f_3^{LL}\Omega_3 - f_4^{LL}\Omega_4 \end{pmatrix}$

$$\vec{\sigma}_k \cdot \vec{p}_{12} = -\mathbb{i}(\vec{\sigma}_k \cdot \hat{r}_{12}) \left(\frac{\partial}{\partial r_{12}} - \frac{\vec{\sigma}_k \cdot \vec{l}_{12}}{r_{12}} \right) \tag{98}$$

reveals that only the formulas for $\vec{\sigma}_k \cdot \hat{r}_{12}$ and $\vec{\sigma}_k \cdot \vec{l}_{12}$ acting on Ω_i are needed for the evaluation of $\vec{\sigma}_k \cdot \vec{p}_{12}(f_i^{XY}\Omega_i)$. Being scalar operators, the actions of $\vec{\sigma}_k \cdot \hat{r}_{12}$, $\vec{\sigma}_k \cdot \vec{l}_{12}$, and $\vec{\sigma}_1 \cdot \vec{\sigma}_2$ on Ω_i can be expressed through the RI

$$\hat{Q}\Omega_i = \sum_{i'=1}^{4} \Omega_{i'} \langle \Omega_{i'} | \hat{Q} | \Omega_i \rangle, \quad \hat{Q} \in \left\{ \vec{\sigma}_k \cdot \hat{r}_{12}, \vec{\sigma}_k \cdot \vec{l}_{12}, \vec{\sigma}_1 \cdot \vec{\sigma}_2 \right\}, \tag{99}$$

where the matrix elements $\langle \Omega_{i'} | \hat{Q} | \Omega_i \rangle$ can systematically be evaluated using the Wigner-Eckart theorem for composite operators [23]. The resulting matrices can be summarized as follows:

$$[\vec{\sigma}_k \cdot \hat{r}_{12}] = \begin{pmatrix} 0 & 0 & \pm b & \mp a \\ 0 & 0 & -a & -b \\ \pm b & -a & 0 & 0 \\ \mp a & -b & 0 & 0 \end{pmatrix}, \quad a = \sqrt{\frac{j+1}{2j+1}}, \quad b = \sqrt{\frac{j}{2j+1}}, \tag{100}$$

$$[\vec{\sigma}_k \cdot \vec{l}_{12}] = \begin{pmatrix} 0 & \pm\sqrt{j(j+1)} & 0 & \\ \pm\sqrt{j(j+1)} & -1 & 0 & \\ 0 & 0 & j-1 & \\ 0 & 0 & 0 & -(j+2) \end{pmatrix} \tag{101}$$

$$[\vec{\sigma}_1 \cdot \vec{\sigma}_2] = \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{102}$$

where the upper and lower signs in the matrix elements of $[\vec{\sigma}_k \cdot \hat{r}_{12}]$ and $[\vec{\sigma}_k \cdot \vec{l}_{12}]$ correspond to $k = 1$ and 2, respectively.

To expedite subsequent manipulations, we combine the functions (93)–(96) with different eigenvalues of $\mathbf{C}_{12}$ to form eigenfunctions of $\{\hat{\mathcal{P}}_{12}, \hat{\mathcal{I}}, \hat{j}_{12}^2, \hat{j}_{12,z}\}$, i.e.,

$$\Psi_+^{A,e} = \begin{pmatrix} f_1^{LL}\Omega_1 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ f_1^{SS}\Omega_1 \end{pmatrix}, \tag{103}$$

$$\Psi_+^{A,o} = \begin{pmatrix} f_2^{LL}\Omega_2 \\ f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4 \\ -(f_3^{LS}\Omega_3 + f_4^{LS}\Omega_4) \\ f_2^{SS}\Omega_2 \end{pmatrix}, \tag{104}$$

each of which has 4 unknowns. These two expressions fully cover Eqs. (93)–(96). Substituting $\Psi_+^{A,e}$ (103) into Eq. (54) and integrating out the spin-angular part Ω_i give rise to four equations for f_1^{LL}, f_3^{LS}, f_4^{LS}, and f_1^{SS}:

$$2\mathbb{i}(-bF_3^{LS} + aF_4^{LS}) + d_C \frac{\alpha}{r_{12}} f_1^{LL} + (-3d_G - d_R)\frac{\alpha}{r_{12}} f_1^{SS} = 0, \tag{105}$$

$$-\mathbb{i}F_{1-}^{LLSS} + (d_C + d_G + qd_R)\frac{\alpha}{r_{12}} f_3^{LS} + (pd_R)\frac{\alpha}{r_{12}} f_4^{LS} = 0, \tag{106}$$

$$-\mathbb{i}F_{1+}^{LLSS} + (pd_R)\frac{\alpha}{r_{12}} f_3^{LS} + (d_C + d_G - qd_R)\frac{\alpha}{r_{12}} f_4^{LS} = 0, \tag{107}$$

$$2\mathbb{i}\left(-bF_3^{LS} + aF_4^{LS}\right) + d_C \frac{\alpha}{r_{12}} f_1^{SS} + (-3d_G - d_R)\frac{\alpha}{r_{12}} f_1^{LL} = 0, \tag{108}$$

while substituting $\Psi_+^{A,o}$ (104) into Eq. (54) leads to another four equations for f_2^{LL}, f_3^{LS}, f_4^{LS}, and f_2^{SS}:

$$-2\mathbb{i}\left(aF_3^{LS} + bF_4^{LS}\right) + d_C \frac{\alpha}{r_{12}} f_2^{LL} + (d_G + d_R)\frac{\alpha}{r_{12}} f_2^{SS} = 0, \tag{109}$$

$$\mathbb{i}F_{2-}^{LLSS} + (d_C - d_G - qd_R)\frac{\alpha}{r_{12}} f_3^{LS} - (pd_R)\frac{\alpha}{r_{12}} f_4^{LS} = 0, \tag{110}$$

$$\mathbb{i}F_{2+}^{LLSS} - (pd_R)\frac{\alpha}{r_{12}} f_3^{LS} + (d_C - d_G + qd_R)\frac{\alpha}{r_{12}} f_4^{LS} = 0, \tag{111}$$

$$2\mathbb{i}\left(aF_3^{LS} + bF_4^{LS}\right) + d_C \frac{\alpha}{r_{12}} f_2^{SS} + (d_G + d_R)\frac{\alpha}{r_{12}} f_2^{LL} = 0. \tag{112}$$

The intermediate quantities in the above equations are defined as

$$F_3^{LS} = \left(\frac{d}{dr} - \frac{j-1}{r}\right) f_3^{LS}, \tag{113}$$

$$F_4^{LS} = \left(\frac{d}{dr} + \frac{j+2}{r}\right) f_4^{LS}, \tag{114}$$

$$F_{1-}^{LLSS} = b\left(\frac{d}{dr} + \frac{j+1}{r}\right)\left(f_1^{LL} + f_1^{SS}\right), \tag{115}$$

$$F_{1+}^{LLSS} = a\left(-\frac{d}{dr} + \frac{j}{r}\right)\left(f_1^{LL} + f_1^{SS}\right), \tag{116}$$

$$F_{2-}^{LLSS} = -a\left(\frac{d}{dr} + \frac{j+1}{r}\right)\left(f_2^{LL} - f_2^{SS}\right), \tag{117}$$

$$F_{2+}^{LLSS} = b\left(-\frac{d}{dr} + \frac{j}{r}\right)\left(f_2^{LL} - f_2^{SS}\right), \tag{118}$$

$$p = \frac{2\sqrt{j(j+1)}}{2j+1} = 2ab, \tag{119}$$

$$q = \frac{1}{2j+1} = a^2 - b^2, \tag{120}$$

with a and b given in Eq. (100). The asymptotic behaviors of $\Psi_+^{A,e}$ and $\Psi_+^{A,o}$ can be obtained by inserting the expansions

$$f_i^{XY} = r_{12}^{\nu} f_i^{XY(0)} + \mathcal{O}\left(\epsilon^{\nu+1}\right), \quad X, Y \in \{L, S\} \tag{121}$$

into the corresponding equations (vide post). The value for ν and the mutual relations between $f_i^{XY(0)}$ are determined by the requirement that the algebraic equations for $f_i^{XY(0)}$ have nontrivial solutions. If the determinant of the coefficient matrix of the algebraic equations is zero for arbitrary ν, the desired ν has to be determined from the next-order equation (42). This particular situation will be discussed in section "Asymptotic Behaviors Determined by the First-Order Condition (42)."

Algebraic Equations for $\Psi_{(+,+)}^{A,e}$

The algebraic equations for $f_i^{XY(0)}$ (93) can be obtained from Eqs. (105) to (108) as

$$\mathbf{M}^e \begin{pmatrix} f_1^{LL(0)} \\ f_3^{LS(0)} \\ f_4^{LS(0)} \end{pmatrix} = 0 \tag{122}$$

with

$$\mathbf{M}^e = \begin{pmatrix} \alpha(d_C - 3d_G - d_R) & -2\mathbb{i}b(\nu - j + 1) & 2\mathbb{i}a(\nu + j + 2) \\ -2\mathbb{i}b(\nu + j + 1) & \alpha(d_C + d_G + qd_R) & \alpha p d_R \\ -2\mathbb{i}a(-\nu + j) & \alpha p d_R & \alpha(d_C + d_G - qd_R) \end{pmatrix}. \tag{123}$$

The determinant of the coefficient matrix $\mathbf{M}^e$ is

$$\det(\mathbf{M}^e) = 4\alpha\{(d_C + d_G + d_R)[(\nu + 1)^2 - 1] - d\} \tag{124}$$

with

$$d = (d_C + d_G - d_R)\left\{ j(j+1) - [(d_C - d_G)^2 - (2d_G + d_R)^2]\frac{\alpha^2}{4} \right\}. \tag{125}$$

The value of ν can then be determined by the conditions $\det(\mathbf{M}^e) = 0$ and $d_C + d_G + d_R \neq 0$. The situations for the DC, DCG, and DCB Hamiltonians are summarized below.

The DC Hamiltonian

Setting $d_C = 1$ and $d_G = d_R = 0$ in Eqs. (124) and (125) leads to

$$\det(\mathbf{M}^e) = 4\alpha \left\{ \nu^2 + 2\nu - j(j+1) + \frac{\alpha^2}{4} \right\} = 0, \tag{126}$$

$$\nu = \sqrt{j(j+1) + 1 - \frac{\alpha^2}{4}} - 1. \tag{127}$$

The value of ν in Eq. (127) with a negative sign in front of the square root must be discarded, because otherwise the corresponding wave functions would not be normalizable. The relations among $f_i^{XY(0)}$ read

$$f_3^{LS(0)} = \frac{2\mathrm{i}b}{\alpha}(\nu + j + 1) f_1^{LL(0)}, \tag{128}$$

$$f_4^{LS(0)} = \frac{2\mathrm{i}a}{\alpha}(-\nu + j) f_1^{LL(0)}. \tag{129}$$

For $j = 0$, Eq. (127) reduces to

$$\nu = \sqrt{1 - \frac{\alpha^2}{4}} - 1 = -\frac{\alpha^2}{8} + \mathcal{O}(\alpha^4), \tag{130}$$

which indicates that the wave function for the ${}^1s_0$ state has a weak singularity at $r_{12} = 0$, as already noticed by Kutzelnigg [17]. As $a = 1$ and $b = 0$ for $j = 0$ (see Eq. (100)), Eqs. (128) and (129) reduce to

$$f_3^{LS(0)} = 0, \tag{131}$$

$$f_4^{LS(0)} = \frac{\mathrm{i}\alpha}{4} f_1^{LL(0)} + \mathcal{O}(\alpha^3). \tag{132}$$

Further in view of the action (79) of $\mathbf{C}_1$ on $\Psi^{A,e}_{(+,+)}$ (93) for $j = 0$, one sees that the ${}^3p_0$ state has the same singularity (130) as ${}^1s_0$.

The DCG Hamiltonian

Setting $d_C = 1$, $d_G = -1$, and $d_R = 0$ in Eq. (123) leads to

$$\mathbf{M}^e = \begin{pmatrix} 4\alpha & -2\mathrm{i}b(\nu - j + 1) & 2\mathrm{i}a(\nu + j + 2) \\ -2\mathrm{i}b(\nu + j + 1) & 0 & 0 \\ -2\mathrm{i}a(-\nu + j) & 0 & 0 \end{pmatrix}. \tag{133}$$

As $\det(\mathbf{M}^e)$ is identically zero, there always exist nontrivial solutions for $f_i^{XY(0)}$. Actually, there can be two cases, $j = 0$ or $j \neq 0$.

If $f_1^{LL(0)} \neq 0$, it can be deduced from the second and third rows of Eqs. (122) together with (133) that

$$j = \nu = b = 0, \tag{134}$$

with the relations among amplitudes being

$$f_3^{LS(0)} = 0, \tag{135}$$

$$f_4^{LS(0)} = -\mathrm{i}\alpha f_1^{LL(0)}. \tag{136}$$

Equation (135) follows directly from the fact that the function Ω_3 (67) does not exist for $j = 0$. Note that the present zero value of ν is different from that $(\sqrt{1+\alpha^2}-1)$ obtained by Kutzelnigg [17].

On the other hand, if $f_1^{LL(0)} = 0$, only the relation

$$b(\nu - j + 1) f_3^{LS(0)} = a(\nu + j + 2) f_4^{LS(0)} \tag{137}$$

can be obtained. Note that $j \neq 0$ in this case, because otherwise $f_3^{LS(0)}$ and hence $f_4^{LS(0)}$ would also vanish, contradicting the requirement for nontrivial solutions. To determine the power ν, the algebraic equations of $\mathcal{O}(\epsilon^{\nu})$ must be considered, as the Coulomb and Gaunt singularities happen to cancel out at $\mathcal{O}(\epsilon^{\nu-1})$; see Eqs. (106) and (107) for $d_C = 1$, $d_G = -1$, and $d_R = 0$. We postpone the discussion of this situation to section "Asymptotic Behaviors Determined by the First-Order Condition (42)."

The DCB Hamiltonian

Setting $d_C = 1$ and $d_G = d_R = -\frac{1}{2}$ in Eqs. (124) and (125) leads to

$$\det(\mathbf{M}^e) = -4\alpha j(j+1). \tag{138}$$

Obviously, a nontrivial solution can only be obtained for $j = 0$, for which $\mathbf{M}^e$ (123) becomes

$$\mathbf{M}^e = \begin{pmatrix} 3\alpha & 0 & 2\mathrm{i}(\nu+2) \\ 0 & 0 & 0 \\ 2\mathrm{i}\nu & 0 & \alpha \end{pmatrix}. \tag{139}$$

Since $f_3^{LS(0)} = 0$ again because of the nonexistence of Ω_3 (67) for $j = 0$, the requirement of nontrivial solution is only fulfilled if the minor M_{22}^e

$$M_{22}^e = \det \begin{pmatrix} 3\alpha & 2\mathrm{i}(\nu+2) \\ 2\mathrm{i}\nu & \alpha \end{pmatrix} = 3\alpha^2 + 4\nu(\nu+2) \tag{140}$$

vanishes, leading to

$$\nu = \sqrt{1 - \frac{3\alpha^2}{4}} - 1 = -\frac{3\alpha^2}{8} + \mathcal{O}(\alpha^4). \tag{141}$$

The relation between the amplitudes is then

$$f_4^{LS(0)} = -\frac{2\mathrm{i}\nu}{\alpha} f_1^{LL(0)} = \frac{3\mathrm{i}\alpha}{4} f_1^{LL(0)} + \mathcal{O}(\alpha^3). \tag{142}$$

It is seen from Eq. (141) that the wave function of the DCB Hamiltonian is also singular at $r_{12} = 0$, somewhat worse than that of the DC Hamiltonian.

Curiously, had $d_C = 1$ and $d_G = -d_R = -1/2$ been chosen in Eqs. (124) and (125), corresponding to an artificial interaction consisting of the Coulomb potential minus the gauge part of the Breit term, we would obtain $\nu = 0$ independently of j as well as the simple relation $f_4^{LS(0)} = \frac{\mathrm{i}\alpha}{2} f_1^{LL(0)}$ for the ${}^1s_0$ state. This is the result actually obtained by Kutzelnigg [17], originally claimed for the DCB Hamiltonian. Of course, this unfortunate mistake was already noticed by himself [24], two decades after the work though.

Algebraic Equations for $\Psi^{A,e}_{(+,-)}$

In the case of $\Psi^{A,e}_{(+,-)}$ (95), the only nontrivial algebraic equation that can be obtained from Eqs. (105) to (108) is

$$\alpha(d_C + 3d_G + d_R) f_1^{LL(0)} = 0. \tag{143}$$

Since the prefactor is different from zero for the DC, DCG, and DCB Hamiltonians, we have $f_1^{LL(0)} = 0$. That is, there exist no nontrivial solutions for all the three Hamiltonians. This occurs also to $\Psi^{S,o}_{(+,-)}$ and $\Psi^{A,o}_{(-,-)}$ for $\Psi^{S,o}_{(+,-)}$ has the same form as $\Psi^{A,e}_{(+,-)}$ (see Eq. (97)) and $\Psi^{A,o}_{(-,-)} = \mathbf{C}_1 \Psi^{S,o}_{(+,-)}$ (see Eq. (78)).

Algebraic Equations for $\Psi^{A,o}_{(+,-)}$

The algebraic equations for $f_i^{XY(0)}$ (96) are obtained from Eqs. (109) to (112) as

$$\mathbf{M}^o \begin{pmatrix} f_2^{LL(0)} \\ f_3^{LS(0)} \\ f_4^{LS(0)} \end{pmatrix} = 0, \tag{144}$$

where

$$\mathbf{M}^o = \begin{pmatrix} \alpha(d_C - d_G - d_R) & -2\mathrm{i}a(\nu - j + 1) & -2\mathrm{i}b(\nu + j + 2) \\ -2\mathrm{i}a(\nu + j + 1) & \alpha(d_C - d_G - qd_R) & -\alpha p d_R \\ 2\mathrm{i}b(-\nu + j) & -\alpha p d_R & \alpha(d_C - d_G + qd_R) \end{pmatrix}. \tag{145}$$

The determinant of the coefficient matrix $\mathbf{M}^o$ is

$$\det(\mathbf{M}^o) = 4\alpha[(d_C - d_G + d_R)(\nu + 1)^2 - d] \tag{146}$$

with

$$d = (d_C - d_G - d_R)\left\{ j(j+1) - [(d_C - d_G)^2 - d_R^2]\frac{\alpha^2}{4} \right\}. \tag{147}$$

The situations for the DC, DCG, and DCB Hamiltonians are summarized as follows.

The DC Hamiltonian

Setting $d_C = 1$ and $d_G = d_R = 0$ in Eq. (146) leads to

$$\det(\mathbf{M}^o) = 4\alpha\left\{ (\nu + 1)^2 - j(j+1) + \frac{\alpha^2}{4} \right\} = 0, \tag{148}$$

$$\nu = \sqrt{j(j+1) - \frac{\alpha^2}{4}} - 1. \tag{149}$$

The relations between the amplitudes read

$$f_3^{LS(0)} = \frac{2\mathrm{i}a}{\alpha}(\nu + j + 1) f_2^{LL(0)}, \tag{150}$$

$$f_4^{LS(0)} = \frac{2\mathrm{i}b}{\alpha}(\nu - j) f_2^{LL(0)}. \tag{151}$$

The DCG Hamiltonian

Setting $d_C = 1$, $d_G = -1$, and $d_R = 0$ in Eq. (146) leads to

$$\det(\mathbf{M}^o) = 8\alpha\{(\nu + 1)^2 - j(j+1) + \alpha^2\} = 0, \tag{152}$$

$$\nu = \sqrt{j(j+1) - \alpha^2} - 1. \tag{153}$$

The relations between the amplitudes read

$$f_3^{LS(0)} = \frac{\mathbb{i}a}{\alpha}(\nu + j + 1) f_2^{LL(0)}, \tag{154}$$

$$f_4^{LS(0)} = \frac{\mathbb{i}b}{\alpha}(\nu - j) f_2^{LL(0)}. \tag{155}$$

The DCB Hamiltonian

Setting $d_C = 1$ and $d_G = d_R = -\frac{1}{2}$ in Eq. (146) leads to

$$\det(\mathbf{M}^o) = 4\alpha[(\nu + 1)^2 - 2j(j + 1) + \alpha^2] = 0, \tag{156}$$

$$\nu = \sqrt{2j(j + 1) - \alpha^2} - 1. \tag{157}$$

The relations between the amplitudes read

$$f_3^{LS(0)} = \frac{\mathbb{i}a}{\alpha}(\nu + 2j + 1) f_2^{LL(0)}, \tag{158}$$

$$f_4^{LS(0)} = \frac{\mathbb{i}b}{\alpha}(\nu - 2j - 1) f_2^{LL(0)}. \tag{159}$$

In sum, as j is odd in Eqs. (149), (153), and (157), the wave functions $\Psi^{A,o}_{(+,-)}$ are all regular at the coalescence point for all the three Hamiltonians. These results hold also for $\Psi^{S,e}_{(+,-)}$ and $\Psi^{A,e}_{(-,-)}$ for $\Psi^{S,e}_{(+,-)}$ has the same form as $\Psi^{A,o}_{(+,-)}$ (see Eq. (97)) and $\Psi^{A,e}_{(-,-)} = \mathbf{C}_1 \Psi^{S,e}_{(+,-)}$ (see Eq. (78)). It is just that, for $\Psi^{A,e}_{(-,-)}$ with even j, the restriction $j \geq 2$ should be imposed, because $\det(\mathbf{M}^o) > 0$ for $j = 0$, which implies no nontrivial solutions.

Algebraic Equations for $\Psi^{A,o}_{(+,+)}$

In the case of $\Psi^{A,o}_{(+,+)}$ (94), the only nontrivial algebraic equation that can be obtained from Eqs. (109) to (112) is

$$\alpha(d_C + d_G + d_R) f_2^{LL(0)} = 0. \tag{160}$$

The prefactor $d_C + d_G + d_R$ equals one for the DC Hamiltonian and hence $f_2^{LL(0)} = 0$. Therefore, no nontrivial solutions exist for $\Psi^{A,o}_{(+,+)}$ of the DC Hamiltonian. In contrast, the prefactor is zero for the DCG and DCB Hamiltonians, such that $f_2^{LL(0)}$ cannot be determined. This situation will further be discussed below.

Asymptotic Behaviors Determined by the First-Order Condition (42)

As discussed before, there are two situations where the solutions cannot be determined by the zeroth-order equation (41). One is $\Psi^{A,e}_{(+,+)}$ (93) with $f_1^{LL(0)} = 0$ for the DCG Hamiltonian, and the other is $\Psi^{A,o}_{(+,+)}$ (94) for the DCG and DCB Hamiltonians. The desired solutions can only be found by resorting to Eq. (42) of $\mathcal{O}(\epsilon^\nu)$. To do so, we first rewrite the operator $\hat{W}^{(0)}$ (38) in block form [18,20]:

$$\hat{W}^{(0)} = \hat{w}_0^{(0)}\mathbf{E} - c^2\mathbf{B}_{12} + \hat{w}_1^{(0)}\mathbf{C}_1 + \hat{w}_2^{(0)}\mathbf{C}_2, \tag{161}$$

where

$$\mathbf{B}_{12} = \boldsymbol{\beta}\circ\mathbf{I}_4 + \mathbf{I}_4\circ\boldsymbol{\beta} = \begin{pmatrix} 2I_4 & 0\,0 & 0 \\ 0 & 0\,0 & 0 \\ 0 & 0\,0 & 0 \\ 0 & 0\,0 & -2I_4 \end{pmatrix}, \tag{162}$$

$$\hat{w}_0^{(0)} = E - \left(\sum_{k\geq 3}\hat{h}_k^D + \sum_{k>l\geq 3}\hat{g}_{kl}\right) - 2\phi^{(0)} - \sum_{l\geq 3} d_C\frac{2}{r_{Rl}}, \tag{163}$$

$$\hat{w}_k^{(0)} = -\sum_{l\geq 3}\left[d_G\frac{\vec{\sigma}_k\cdot\vec{\alpha}_l}{r_{Rl}} + d_R\frac{(\vec{\sigma}_k\cdot\hat{r}_{Rl})(\vec{\alpha}_l\cdot\hat{r}_{Rl})}{r_{Rl}}\right] - \frac{1}{2}c\vec{\sigma}_k\cdot\vec{P}_{12}, \quad k = 1,2. \tag{164}$$

Since the spin-angular functions Ω_i of given j and m_j are not eigenfunctions of $\vec{\sigma}_k$ ($k = 1,2$) in $\hat{w}_k^{(0)}$ (164), the amplitudes $f_i^{XY(0)}$ with different j and m_j become coupled. In addition, both the spins and orbital angular momenta of electrons 1 and 2 will be entangled individually with those of the other electrons due to the presence of $\vec{\alpha}_l$ and $\hat{r}_{Rl}$ in $\hat{w}_k^{(0)}$. Therefore, the reduced two-electron problem becomes truly a many-body problem and no simple solutions can be found from Eq. (42). Yet, the situation can be simplified by neglecting the couplings between the $f_i^{XY(0)}$ with different j and m_j. That is, $\Psi^{(\nu)}$ are assumed to be eigenfunctions of $\{\hat{j}_{12}^2, \hat{j}_{12,z}\}$.

$\Psi^{A,e}_{(+,+)}$ with $f_1^{LL(0)} = 0$ for the DCG Hamiltonian

In order to determine the asymptotic behavior of $\Psi^{A,e}_{(+,+)}$ (93) with $f_1^{LL(0)} = 0$ for the DCG Hamiltonian, it is assumed that $\Psi^{(\nu)}$ is an eigenfunction of $\{\hat{j}_{12}^2, \hat{j}_{12,z}\}$, viz.,

$$\Psi^{(\nu)} = \Psi^{A,e}_{(+,+)} = r_{12}^{\nu} \begin{pmatrix} 0 \\ f_3^{LS(0)}\Omega_3 + f_4^{LS(0)}\Omega_4 \\ f_3^{LS(0)}\Omega_3 + f_4^{LS(0)}\Omega_4 \\ 0 \end{pmatrix}, \tag{165}$$

with the relation between $f_3^{LS(0)}$ and $f_4^{LS(0)}$ given by Eq. (137). The action of $\hat{W}^{(0)}$ on $\Psi^{(\nu)}$ is

$$\begin{aligned} \hat{W}^{(0)}\Psi^{(\nu)} = r_{12}^{\nu} & \begin{pmatrix} 0 \\ \hat{w}_0^{(0)}(f_3^{LS(0)}\Omega_3 + f_4^{LS(0)}\Omega_4) \\ \hat{w}_0^{(0)}(f_3^{LS(0)}\Omega_3 + f_4^{LS(0)}\Omega_4) \\ 0 \end{pmatrix} \\ + r_{12}^{\nu} & \begin{pmatrix} (\hat{w}_1^{(0)} + \hat{w}_2^{(0)})(f_3^{LS(0)}\Omega_3 + f_4^{LS(0)}\Omega_4) \\ 0 \\ 0 \\ (\hat{w}_1^{(0)} + \hat{w}_2^{(0)})(f_3^{LS(0)}\Omega_3 + f_4^{LS(0)}\Omega_4) \end{pmatrix}. \end{aligned} \tag{166}$$

The first and second parts of $\hat{W}^{(0)}\Psi^{(\nu)}$ correspond to different angular momenta since $\hat{w}_1^{(0)}$ and $\hat{w}_2^{(0)}$ do not commute with $\{\hat{j}_{12}^2, \hat{j}_{12,z}\}$. Therefore, to determine the power ν, suffice it to only consider the first part of $\hat{W}^{(0)}\Psi^{(\nu)}$ (166), which shares the same symmetry as $\Psi^{A,e}_{(+,+)}$. In this case, $\Psi^{(\nu+1)}$ can still be chosen as the form of $\Psi^{A,e}_{(+,+)}$ (93). Substituting the expression for $\Psi^{(\nu+1)}$ into Eq. (42) and integrating out the angular parts Ω_3 and Ω_4 yield two algebraic equations:

$$-2\mathbb{i}b(\nu + j + 2) f_1^{LL(1)} = \alpha \hat{w}_0^{(0)} f_3^{LS(0)}, \tag{167}$$

$$-2\mathbb{i}a(-\nu + j - 1) f_1^{LL(1)} = \alpha \hat{w}_0^{(0)} f_4^{LS(0)}, \tag{168}$$

where $f_1^{LL(1)}$ represents the first-order unknown in $\Psi^{(\nu+1)}$. Equations (167), (168), and (137) together give rise to

$$\nu = j - 1, \tag{169}$$

$$f_4^{LS(0)} = 0. \tag{170}$$

Note that the value of j here cannot be 0, because otherwise $f_3^{LS(0)}$ would also vanish, contradicting the requirement that ν should be the lowest power of Ψ with at least one nonvanishing amplitude $f_i^{XY(0)}$.

Finally, it is interesting to see from Eqs. (134) and (169) that the wave functions $\Psi^{A,e}_{(+,+)}$ of the DCG Hamiltonian are of integral powers of r_{12} and hence free of

singularities. In the case of $f_1^{LL(0)} = 0$, Ψ^{LL} and Ψ^{SS} are of order r_{12}^{v+1}, one order higher than that of Ψ^{LS} and Ψ^{SL}. This is quite different from all the other cases where all the components share the same power.

$\Psi^{A,o}_{(+,+)}$ for the DCG and DCB Hamiltonians

The yet undetermined $\Psi^{A,o}_{(+,+)}$ (94) for the DCG and DCB Hamiltonians can be analyzed in the same way as before. Again assume that $\Psi^{A,o}_{(+,+)}$ is an eigenfunction of $\{\hat{j}_{12}^2, \hat{j}_{12,z}\}$, viz.,

$$\Psi^{(v)} = \Psi^{A,o}_{(+,+)} = r_{12}^v \begin{pmatrix} f_2^{LL(0)}\Omega_2 \\ 0 \\ 0 \\ f_2^{LL(0)}\Omega_2 \end{pmatrix}. \tag{171}$$

The action of $\hat{W}^{(0)}$ on $\Psi^{(v)}$ leads to

$$\begin{aligned} \hat{W}^{(0)}\Psi^{(v)} = r_{12}^v & \begin{pmatrix} (\hat{w}_0^{(0)} - 2c^2) f_2^{LL(0)}\Omega_2 \\ 0 \\ 0 \\ (\hat{w}_0^{(0)} + 2c^2) f_2^{LL(0)}\Omega_2 \end{pmatrix} \\ + r_{12}^v & \begin{pmatrix} 0 \\ (\hat{w}_1^{(0)} + \hat{w}_2^{(0)}) f_2^{LL(0)}\Omega_2 \\ (\hat{w}_1^{(0)} + \hat{w}_2^{(0)}) f_2^{LL(0)}\Omega_2 \\ 0 \end{pmatrix}. \end{aligned} \tag{172}$$

To determine the power v, only the first part of $\hat{W}^{(0)}\Psi^{(v)}$ (172) is to be considered, which conserves antisymmetry and has the same parity as $\Psi^{A,o}_{(+,+)}$, but breaks the $\mathbf{C}_{12}$ symmetry. Therefore, $\Psi^{(v+1)}$ has to take the form (104). Substituting $\Psi^{(v+1)}$ into Eq. (42) and integrating out Ω_2 gives rise to

$$\begin{aligned} & \alpha f_2^{LL(1)} - 2\mathbb{i}a(v - j + 2) f_3^{LS(1)} - 2\mathbb{i}b(v + j + 3) f_4^{LS(1)} - \alpha f_2^{SS(1)} \\ = & \alpha \left(\hat{w}_0^{(0)} - 2c^2\right) f_2^{LL(0)}, \end{aligned} \tag{173}$$

$$\begin{aligned} & -\alpha f_2^{LL(1)} + 2\mathbb{i}a(v - j + 2) f_3^{LS(1)} + 2\mathbb{i}b(v + j + 3) f_4^{LS(1)} + \alpha f_2^{SS(1)} \\ = & \alpha \left(\hat{w}_0^{(0)} + 2c^2\right) f_2^{LL(0)}. \end{aligned} \tag{174}$$

Table 2 Asymptotic behaviors (r_{12}^{ν}) of the wave functions of the Dirac-Coulomb (DC), Dirac-Coulomb-Gaunt (DCG), and Dirac-Coulomb-Breit (DCB) Hamiltonians. $j = l$ for $s = 0$ and $j = l + 1, l, l - 1$ for $s = 1$. For other explanations, see Table 1 (Source: Reprinted with permission from J. Chem. Phys. 136, 144117 (2012). © 2012, American Institute of Physics)

Hamiltonian	Wave function	$\nu(j)$	j
DC	$\Psi^{A,e}_{(+,+)}, \Psi^{A,e}_{(-,+)}$	$\sqrt{j(j+1)+1-\frac{\alpha^2}{4}}-1$	$0, 2, 4, \ldots$
	$\Psi^{A,o}_{(+,-)}, \Psi^{A,e}_{(-,-)}$	$\sqrt{j(j+1)-\frac{\alpha^2}{4}}-1$	$1, 2, 3, \ldots$
DCG	$\Psi^{A,e}_{(+,+)}, \Psi^{A,e}_{(-,+)}$	0	0
	$\Psi^{A,e}_{(+,+)}, \Psi^{A,e}_{(-,+)}$	$j-1$	$2, 4, 6, \ldots$
	$\Psi^{A,o}_{(+,-)}, \Psi^{A,e}_{(-,-)}$	$\sqrt{j(j+1)-\alpha^2}-1$	$1, 2, 3, \ldots$
DCB	$\Psi^{A,e}_{(+,+)}, \Psi^{A,e}_{(-,+)}$	$\sqrt{1-\frac{3\alpha^2}{4}}-1$	0
	$\Psi^{A,o}_{(+,-)}, \Psi^{A,e}_{(-,-)}$	$\sqrt{2j(j+1)-\alpha^2}-1$	$1, 2, 3, \ldots$

Note that the left-hand side of Eq. (173) is just the negative of that of Eq. (174), which is also evident from Eqs. (109) and (112). The sum of Eqs. (173) and (174) then leads to

$$2\alpha \hat{w}_0^{(0)} f_2^{LL(0)} = 0. \tag{175}$$

Since $\hat{w}_0^{(0)}$ cannot always be zero, $f_2^{LL(0)}$ must be zero. Therefore, there exist no nontrivial solutions for $\Psi^{A,o}_{(+,+)}$ of the DCG and DCB Hamiltonians.

For clarity, the so-obtained asymptotic behaviors of the various relativistic wave functions are summarized in Table 2.

Discussion

It can be deduced from Table 2 that, except for $\nu(j)$ with $j = 0 = l$ (DC, DCG, DCB) and $\nu(j) = j - 1$ with $j = l + 1$ (DCG), the powers $\nu(j)$ do not have the correct nrl ($\nu = l$). That is, the two limits of $r_{12} \to 0$ and $c \to \infty$ generally do not commute. To further scrutinize this peculiarity, we consider the exact wave function of the DC Hamiltonian for a two-electron system

$$\begin{pmatrix} V_C & c\vec{\sigma}_2 \cdot \vec{p}_2 & c\vec{\sigma}_1 \cdot \vec{p}_1 & 0 \\ c\vec{\sigma}_2 \cdot \vec{p}_2 & V_C & 0 & c\vec{\sigma}_1 \cdot \vec{p}_1 \\ c\vec{\sigma}_1 \cdot \vec{p}_1 & 0 & V_C & c\vec{\sigma}_2 \cdot \vec{p}_2 \\ 0 & c\vec{\sigma}_1 \cdot \vec{p}_1 & c\vec{\sigma}_2 \cdot \vec{p}_2 & V_C \end{pmatrix} \begin{pmatrix} \Psi^{LL} \\ \Psi^{LS} \\ \Psi^{SL} \\ \Psi^{SS} \end{pmatrix} = \begin{pmatrix} W\Psi^{LL} \\ (W + 2c^2)\Psi^{LS} \\ (W + 2c^2)\Psi^{SL} \\ (W + 4c^2)\Psi^{SS} \end{pmatrix}, \tag{176}$$

where

$$W = E - 2c^2 - \phi_1 - \phi_2. \tag{177}$$

Equation (176) gives rise to the following relations between the components Ψ^{XY}

$$\frac{\Psi^{SL}}{\Psi^{LS}} = \frac{\vec{\sigma}_1 \cdot \vec{p}_1 \Psi^{LL} + \vec{\sigma}_2 \cdot \vec{p}_2 \Psi^{SS}}{\vec{\sigma}_2 \cdot \vec{p}_2 \Psi^{LL} + \vec{\sigma}_1 \cdot \vec{p}_1 \Psi^{SS}}, \tag{178}$$

$$\frac{\Psi^{SS}}{\Psi^{LL}} = \frac{\vec{\sigma}_1 \cdot \vec{p}_1 \Psi^{LS} + \vec{\sigma}_2 \cdot \vec{p}_2 \Psi^{SL}}{\vec{\sigma}_2 \cdot \vec{p}_2 \Psi^{LS} + \vec{\sigma}_1 \cdot \vec{p}_1 \Psi^{SL}} R(c, r_{12}), \tag{179}$$

$$R(c, r_{12}) = \frac{W - V_C}{W - V_C + 4c^2}. \tag{180}$$

Equation (178) implies that Ψ^{SL} and Ψ^{LS} are of the same orders in c and r_{12}. Therefore, the ratio $\frac{\Psi^{SS}}{\Psi^{LL}}$ (179) at the two limits is determined mainly by the function $R(c, r_{12})$ (180), which behaves as

$$\lim_{c\to\infty} R(c, r_{12})$$

$$= \lim_{c\to\infty} \left\{ c^{-2} \left(-\frac{1}{4r_{12}} + \frac{W}{4} \right) + c^{-4} \left(-\frac{1}{16 r_{12}^2} + \frac{W}{8 r_{12}} - \frac{W^2}{16} \right) + \mathcal{O}(c^{-6}) \right\} \tag{181}$$

$$= 0, \tag{182}$$

$$\lim_{r_{12}\to 0} R(c, r_{12})$$

$$= \lim_{r_{12}\to 0} \left\{ 1 + 4c^2 r_{12} + \mathcal{O}\left(r_{12}^2\right) \right\} \tag{183}$$

$$= 1. \tag{184}$$

That is,

$$\lim_{r_{12}\to 0} \lim_{c\to\infty} R(c, r_{12}) = 0, \tag{185}$$

$$\lim_{c\to\infty} \lim_{r_{12}\to 0} R(c, r_{12}) = 1. \tag{186}$$

The following limits can then readily be deduced:

$$\lim_{c\to\infty} \frac{\Psi^{SS}}{\Psi^{LL}} = 0, \tag{187}$$

$$\lim_{r_{12}\to 0} \frac{\Psi^{SS}}{\Psi^{LL}} = \lim_{r_{12}\to 0} \frac{\vec{\sigma}_1 \cdot \vec{p}_1 \Psi^{LS} + \vec{\sigma}_2 \cdot \vec{p}_2 \Psi^{SL}}{\vec{\sigma}_2 \cdot \vec{p}_2 \Psi^{LS} + \vec{\sigma}_1 \cdot \vec{p}_1 \Psi^{SL}} \cdot \lim_{r_{12}\to 0} R(c, r_{12}) \tag{188}$$

$$= \lim_{r_{12}\to 0} \frac{\vec{\sigma}_1 \cdot \vec{p}_{12} \Psi^{LS(\nu)} - \vec{\sigma}_2 \cdot \vec{p}_{12} \Psi^{SL(\nu)}}{\vec{\sigma}_1 \cdot \vec{p}_{12} \Psi^{SL(\nu)} - \vec{\sigma}_2 \cdot \vec{p}_{12} \Psi^{LS(\nu)}} \tag{189}$$

$$= \lim_{r_{12}\to 0} \frac{-V_C \Psi^{SS(\nu)}}{-V_C \Psi^{LL(\nu)}} \tag{190}$$

$$= 1. \tag{191}$$

Equation (189) arises from the coordinate transformation (30) followed by the expansions (35) of Ψ^{XY} in r_{12}, while Eq. (190) follows directly from the first and fourth rows of the homogeneous equation (54) determining the asymptotic behaviors of Ψ. Finally, Eq. (191) stems from the relation $\lim_{r_{12}\to 0} \frac{\Psi^{SS(v)}}{\Psi^{LL(v)}} = \lim_{r_{12}\to 0} \frac{r_{12}^v f_{j=0}^{SS(0)}}{r_{12}^v f_{j=0}^{LL(0)}} = 1$, where the first equality holds because the $j = 0$ state is singular while other states vanish at $r_{12} = 0$ and the second equality is implied directly by the structure (93) arising from the $\mathbf{C}_{12}$ operation. It is therefore clear that the two limits do not commute, viz.,

$$0 = \lim_{r_{12}\to 0} \lim_{c\to\infty} \frac{\Psi^{SS}}{\Psi^{LL}} \neq \lim_{c\to\infty} \lim_{r_{12}\to 0} \frac{\Psi^{SS}}{\Psi^{LL}} = 1. \tag{192}$$

This is quite different from the electron-nucleus coalescence; see Eq. (21).

It is to be noted that the expansion (183) of $R(c, r_{12})$ around $r_{12} = 0$ only has a finite convergence radius $R_c = \frac{1}{|4c^2+W|}$, which is about $\frac{\alpha^2}{4}$ for the usual bound states of interest ($W \sim \mathcal{O}(c^0)$). Therefore, the previously obtained asymptotic behaviors of the relativistic wave functions are only valid for $r_{12} < R_c$. As for the behaviors of the wave functions at $r_{12} > R_c$, the following expansion of $R(c, r_{12})$ around $r_{12} = +\infty$ should be adopted:

$$R(c, r_{12}) = \frac{W}{4c^2 + W} - \frac{4c^2}{(4c^2 + W)^2 r_{12}} - \frac{4c^2}{(4c^2 + W)^3 r_{12}^2} + \cdots \tag{193}$$

For the situation $|W| < 4c^2$, the right-hand side can further be expanded around $c = \infty$, leading formally to Eq. (181). However, it should be noted that $r_{12} > R_c$ and $|W| < 4c^2$ together span only a subdomain of the expansion (181), viz., $|\frac{W-V_C}{4c^2}| < 1$ or equivalently $W - 4c^2 < \frac{1}{r_{12}} < W + 4c^2$. That is, the expansions (193) and (181) agree with each other only if both $|W| < 4c^2$ and $r_{12} > R_c$ hold.

The expansion (181) of $R(c, r_{12})$ around $c = \infty$ is intimately linked to the explicitly correlated direct perturbation theory (DPT) [25]. Because of the change of the metric, DPT assumes the expansion

$$\bar{\Psi}^{SS} = c^2 \Psi^{SS} = \bar{\Psi}_0^{SS} + c^{-2} \bar{\Psi}_2^{SS} + \mathcal{O}(c^{-4}), \tag{194}$$

which amounts to using (181) for the scaled quantity $\bar{R}(c, r_{12}) = c^2 R(c, r_{12}) = \bar{R}_0(r_{12}) + c^{-2}\bar{R}_2(r_{12}) + \mathcal{O}(c^{-4})$. As already mentioned, for $|W| < 4c^2$, the usual energy range of interest, $\bar{R}(c, r_{12})$, only converges for $r_{12} > R_c$. Therefore, the extension of $\bar{R}(c, r_{12})$ to the domain $r_{12} < R_c$ is doomed to fail. Such a failure occurs already to the second-order relativistic correction $c^{-4}E_4$, which involves the term $\langle \bar{\Psi}_0^{SS} | \frac{1}{r_{12}} | \bar{\Psi}_0^{SS} \rangle$. In view of the relations $\bar{\Psi}_0^{SS} \sim \Psi_0^{LL} \bar{R}_0(r_{12})$, $\Psi_0^{LL} \sim r_{12}^0$, and $\bar{R}_0(r_{12}) = -\frac{1}{4r_{12}} + \frac{W}{4}$, it can immediately be seen that $\bar{\Psi}_0^{SS}$ goes as r_{12}^{-1}, in agreement with the previous result [25] deduced from the nonrelativistic correlation cusp condition (1) as well as the kinetic balance conditions. Therefore, the integrand

of $\langle\bar{\Psi}_0^{SS}|\frac{1}{r_{12}}|\bar{\Psi}_0^{SS}\rangle$ goes as r_{12}^{-3} at small r_{12}, resulting in divergence. For higher orders of DPT, even more severe divergences would arise, as can be seen from the more singular term $\bar{R}_2(r_{12})$ going as r_{12}^{-2} (see Eq. (181)). In contrast, the exact Ψ^{SS} is much better behaved than $\bar{\Psi}_0^{SS}$ at $r_{12} < R_c$; see Eq. (183). More specifically, the only singular term of Ψ^{SS} goes as $f^{SS(0)}r_{12}^{\nu}$ with ν given in Eq. (130) for $j = 0$. Besides, as dictated by the $\mathbf{C}_{12}$ symmetry, $f^{SS(0)}$ has the same magnitude as $f^{LL(0)}$, implying that the leading term of $\bar{\Psi}^{SS} = c^2\Psi^{SS}$ should be of $\mathcal{O}(c^2)$, larger by two orders of magnitude than the presumed $\mathcal{O}(c^0)$ starting point of DPT in Eq. (194). This is quite different from the one-electron case, where the leading term of the ratio ψ^S/ψ^L is of $\mathcal{O}(c^{-1})$, such that the leading term of the scaled small component $\bar{\psi}^S = c\psi^S$ is indeed of $\mathcal{O}(c^0)$. It is therefore not surprising that the α-morphology of one-electron wave functions is guaranteed, but that of many-electron DC wave functions holds only for some bounded interaction [26].

Another point that deserves to be mentioned here is that the denominator of $R(c, r_{12})$ (180) becomes zero at $r_{12} = R_c$, such that the ratio $\frac{\Psi^{SS}}{\Psi^{LL}}$ is divergent if the factor $\frac{\vec{\sigma}_1\cdot\vec{p}_1\Psi^{LS}+\vec{\sigma}_2\cdot\vec{p}_2\Psi^{SL}}{\vec{\sigma}_2\cdot\vec{p}_2\Psi^{LS}+\vec{\sigma}_1\cdot\vec{p}_1\Psi^{SL}}$ cannot offset this singularity. As there is no obvious reason for this factor as well as Ψ^{LL} to always be zero on the $3N-1$ dimensional hypersurface determined by the constraint $r_{12} = R_c$, Ψ^{SS} might be divergent at $r_{12} = R_c$. It can hence be conjectured that the DC wave functions cannot be normalized. In other words, the first-quantized DC Hamiltonian has no bound electronic states. This formal argument for the exact DC wave functions is in line with the previous findings based on second-order perturbation analysis [27], all-order numerical calculations [28], or simplified quasi-solvable models [29].

Summary

The asymptotic behaviors of relativistic wave functions at the electron-electron coalescence points have been analyzed in depth by making full use of the internal symmetries of the reduced two-electron systems. In addition, formal evidence has been found to show that the first-quantized many-electron Dirac equation has no bound electronic states. The results enrich our understandings of relativistic wave functions. However, it must be realized that all the results here are purely mathematical properties of the first-quantized many-electron Dirac equation. Unfortunately, the equation has some inherent unphysical ingredients and hence cannot directly be solved [30–32]. This essentially means that the term "exact relativistic wave function" is physically meaningless. Still, however, the present results are very useful for designing practical relativistic explicitly correlated methods, a topic that will be discussed elsewhere.

Acknowledgements The research of this work was supported by grants from the National Natural Science Foundation of China (Project Nos. 21033001, 21273011, 21290192, and 11471025).

References

1. Hylleraas EA (1929) Neue berechnung der energie des Heliums im grundzustande, sowie des tiefsten terms von ortho-Helium. Z Phys 54:347
2. Kato T (1957) On the eigenfunctions of many-particle systems in quantum mechanics. Commun Pure Appl Math 10:151
3. Pack RT, Brown WB (1966) Cusp conditions for molecular wavefunctions. J Chem Phys 45:556
4. Tew DP (2008) Second order coalescence conditions of molecular wave functions. J Chem Phys 129:014104
5. Klopper W, Manby FR, Ten-no S, Valeev EF (2006) R12 methods in explicitly correlated molecular electronic structure theory. Int Rev Phys Chem 25:427
6. Shiozaki T, Valeev EF, Hirata S (2009) Explicitly correlated coupled-cluster methods. Annu Rev Comput Chem 5:131
7. Hättig C, Klopper W, Köhn A, Tew DP (2012) Explicitly correlated electrons in molecules. Chem Rev 112:4
8. Kong L, Bischoff FA, Valeev EF (2012) Explicitly correlated R12/F12 methods for electronic structure. Chem Rev 112:75
9. Ten-no S (2012) Explicitly correlated wave functions: Summary and perspective. Theor Chem Acc 131:1070
10. Kutzelnigg W (1985) r_{12}-dependent terms in the wave function as closed sums of partial wave amplitudes for large l. Theor Chim Acta 68:445
11. Valeev EF (2004) Improving on the resolution of the identity in linear R12 ab initio theories. Chem Phys Lett 395:190
12. Kutzelnigg W, Klopper W (1991) Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. general theory. J Chem Phys 94:1985
13. Ten-no S (2004) Initiation of explicitly correlated Slater-type geminal theory. Chem Phys Lett 398:56
14. Salomonson S, Öster P (1989) Relativistic all-order pair functions from a discretized single-particle Dirac Hamiltonian. Phys Rev A 40:5548
15. Ottschofski E, Kutzelnigg W (1997) Direct perturbation theory of relativistic effects for explicitly correlated wave functions: the He isoelectronic series. J Chem Phys 106:6634
16. Halkier A, Helgaker T, Klopper W, Olsen J (2000) Basis-set convergence of the two-electron Darwin term. Chem Phys Lett 319:287
17. Kutzelnigg W (1989) Generalization of Kato's cusp conditions to the relativistic case. In: Mukherjee D (ed) Aspects of many-body effects in molecules and extended systems. Springer, Berlin, p 353
18. Li Z, Shao S, Liu WJ (2012) Relativistic explicit correlation: coalescence conditions and practical suggestions. Chem Phys 136:144117
19. Bethe HA, Salpheter EE (1977) Quantum mechanics of one- and two-electron atoms. Plenum Publishing, New York
20. Shao S, Li Z, Liu W, Basic structures of relativistic wave functions. In: Liu W (ed) Handbook of relativistic quantum chemistry. Springer
21. Tracy DS, Singh RP (1972) A new matrix product and its applications in partitioned matrix differentiation. Stat Neerl 26:143
22. Liu S (1999) Matrix results on the Khatri-Rao and Tracy-Singh products. Linear Algebra Appl 289:267
23. Edmonds AR (1957) Angular momentum in quantum mechanics. Princeton University Press, Princeton
24. Kutzelnigg W (2008) Relativistic corrections to the partial wave expansion of two-electron atoms. Int J Quantum Chem 108:2280

25. Kutzelnigg W (2002) Perturbation theory of relativistic effects. In: Schwerdtfeger P (ed) Relativistic electronic structure theory. Part 1. Fundamentals. Elsevier, Amsterdam, p 664
26. Gesztesy F, Grosse H, Thaller B (1984) Relativistic corrections to bound-state energies for two-fermion systems. Phys Rev D 30:2189
27. Brown RE, Ravenhall DG (1951) On the interaction of two electrons. Proc R Soc A 208:552
28. Pestka G, Bylicki M, Karwowski J (2006) Application of the complex-coordinate rotation to the relativistic Hylleraas-CI method: a case study. J Phys B At Mol Opt Phys 39:2979
29. Sucher J (1985) Continuum dissolution and the relativistic many-body problem: a solvable model. Phys Rev Lett 55:1033
30. Liu W (2012) Perspectives of relativistic quantum chemistry: the negative energy cat smiles. Phys Chem Chem Phys 14:35
31. Liu W, Lindgren I (2013) Going beyond "no-pair relativistic quantum chemistry". J Chem Phys 139:014108
32. Liu W (2014) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59

Relativistic Explicit Correlation: Problems and Solutions

17

Wenjian Liu, Sihong Shao, and Zhendong Li

Contents

Introduction 531
R-MP2-F12 536
Summary 542
References 543

Abstract

The fundamental problems inherent in relativistic explicit correlation are highlighted, with practical suggestions for guiding future development of relativistic explicitly correlated wave function methods.

Keywords

Relativistic wave function • Coalescence condition • Explicit correlation • Extended no-pair projection

Introduction

It has been well recognized that, being smooth functions, orbital products (Slater determinants) fail to model the Coulomb cusp behaviors [1, 2] of the exact wave functions of the Schrödinger equation. A direct consequence is that electron

W. Liu (✉) • Z. Li
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: liuwjbdf@gmail.com; zhendongli2008@gmail.com

S. Shao
LMAM and School of Mathematical Sciences, Peking University, Beijing, China
e-mail: sihong@math.pku.edu.cn

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_9

correlation energies calculated by orbital-based methods converge extremely slowly with respect to the basis set size. The situation can only be improved by using trial wave functions that depend explicitly on the interelectronic distances r_{ij}. This was demonstrated by Hylleraas [3], already in 1929, for the ground state of helium. However, the development of explicitly correlated wave function methods for general molecular systems has long been hampered by the concomitant complicated many-electron integrals. The problem was solved eventually by Kutzelnigg [4] in 1985, who observed that the three- and four-electron integrals involving the correlation factor f_{12} can be factorized into products of one- and two-electron integrals by careful use of the resolution of the identity (RI). Since then, a bunch of explicitly correlated wave function methods have been developed in the last decades (for recent reviews, see Refs. [5–8]). In particular, by augmenting the conventional excitations into products of unoccupied one-electron orbitals by just a small set of explicitly correlated configurations, the so-called R12/F12 methods [9,10] have now evolved into practical tools for general molecules.

The request for relativistic explicitly correlated wave function methods for systems containing heavy atoms is even more imperative, as relativistic corrections converge more slowly with respect to the basis set size than nonrelativistic correlation energies [11–13]: The leading partial wave increment (PWI) of the first-order relativistic correction to the ground state of helium-like ions goes [14] as $(l+\frac{1}{2})^{-2}$, which is to be compared with $(l+\frac{1}{2})^{-4}$ for the nonrelativistic correlation energy [15]. More insights can be gained [16] by means of double perturbation theory that treats both relativity and electron-electron interaction as perturbations, with the nonrelativistic bare-nuclear Hamiltonian as zeroth order, viz.,

$$\begin{aligned}\Psi &= \Psi^{(0,0)} + \lambda\Psi^{(0,1)} + \alpha^2\Psi^{(2,0)} + \lambda^2\Psi^{(0,2)} + \alpha^2\lambda\Psi^{(2,1)} \\ &\quad + \alpha^4\Psi^{(4,0)} + \cdots, \quad \alpha = c^{-1}, \qquad (1)\end{aligned}$$

$$E = E^{(0,0)} + E^{\mathrm{corr}} + E^{\mathrm{rel}} + E^{\mathrm{rel/corr}}, \qquad (2)$$

$$E^{\mathrm{corr}} = \sum_{n=1}^{\infty} \lambda^n E^{(0,n)}, \qquad (3)$$

$$E^{\mathrm{rel}} = \sum_{m=1}^{\infty} \alpha^{2m} E^{(2m,0)}, \qquad (4)$$

$$E^{\mathrm{X}} = \sum_{m,n=1}^{\infty} \alpha^{2m}\lambda^n E^{(2m,n)}. \qquad (5)$$

Here, the first and second superscripts denote the respective orders in relativity (α) and electron-electron interaction (λ). In particular, $\Psi^{(0,0)}$ is simply the antisymmetrized product of nonrelativistic hydrogenic orbitals. For the ground state of helium-like ions, the $(l+\frac{1}{2})^{-4}$ type of leading PWI of the nonrelativistic

correlation energy $E^{\rm corr}$ (3) is known [17, 18] to arise from the second-order term $E^{(0,2)} = \langle \Psi^{(0,0)} | \frac{1}{r_{12}} | \Psi^{(0,1)} \rangle$, where $\Psi^{(0,1)}$ is of $\mathcal{O}(r_{12}^1)$. The configuration interaction type of expansion of $\Psi^{(0,1)}$, with one-electron functions saturated up to angular momentum l, would hence amount to approximating r_{12} as

$$r_{12} \approx \sum_{k=0}^{l} (r_{12})_k, \quad (r_{12})_k = \{r_{12}\}_k P_k(\cos\theta_{12}),$$
$$\{r_{12}\}_k = \frac{1}{2k+3} \frac{r_<^{k+2}}{r_>^{k+1}} - \frac{1}{2k-1} \frac{r_<^k}{r_>^{k-1}}, \tag{6}$$

which converges very slowly in modeling the identity $1 = \frac{1}{r_{12}} r_{12}$. The first-order uncorrelated relativistic correction $E^{(2,0)}$ in $E^{\rm rel}$ (4) can directly be evaluated and hence does not contribute to the PWI. It is also straightforward to show that only the s-wave term contributes to $E^{(2,1)}$ (5). The leading PWI of the cross term $E^{(2,2)}$ goes as $(l + \frac{1}{2})^{-2}$ due to the mass-velocity term $\langle \Psi^{(0,1)} | T_1 T_2 | \Psi^{(0,1)} \rangle$. It is therefore clear that the observed leading PWI [14] of $E_2 = \sum_{n=1}^{\infty} E^{(2,n)}$, i.e., first order in relativity but infinite order in correlation, is actually due to the lowest-order interplay $E^{(2,2)}$ between relativity and correlation. Note in passing that the two-electron Darwin term, going also as $(l + \frac{1}{2})^{-2}$, only appears in the Breit-Pauli Hamiltonian [13] but not in direct perturbation theory (DPT) [19].

The central question is: *How to do relativistic explicit correlation*?

More specifically, the following issues must be addressed before going to a specific *Ansatz*.

1. To achieve fast basis set convergence furnished by the correlation factor f_{12}, a first quantized kinetic energy operator, whether nonrelativistic or relativistic, has to be adopted. This essentially means that any four-component relativistic explicitly correlated wave function method must adopt the first quantized Dirac-Coulomb (DC), Dirac-Coulomb-Gaunt (DCG), or Dirac-Coulomb-Breit (DCB) Hamiltonian. However, the resulting configuration space DC/DCG/DCB equation has unphysical ingredients and cannot directly be solved even if it had bound states [20]. To show this, we compare the expressions for the second-order correlation energy $\mathcal{E}^{(2)}$ (which is not expanded in relativity) by quantum electrodynamics (QED) [21, 22]

$$\mathcal{E}^{(2)}_{\rm QED} = \mathcal{E}^{(2)}_{\rm QED,1} + \mathcal{E}^{(2)}_{\rm QED,2}, \tag{7}$$
$$\mathcal{E}^{(2)}_{\rm QED,1} = \mathcal{E}^{(2)}_{\rm QED,1a} + \mathcal{E}^{(2)}_{\rm QED,1b}, \tag{8}$$
$$\mathcal{E}^{(2)}_{\rm QED,1a} = -\left[\frac{(V_{\rm HF})^a_{\tilde{i}} (V_{\rm HF})^{\tilde{i}}_a}{\varepsilon_{\tilde{i}} - \varepsilon_a} + \frac{(V_{\rm HF})^i_{\tilde{i}} (V_{\rm HF})^{\tilde{i}}_i}{\varepsilon_{\tilde{i}} - \varepsilon_i} \right], \tag{9}$$

$$\mathcal{E}^{(2)}_{\mathrm{QED},1b} = \frac{Q^a_i Q^i_a}{\varepsilon_i - \varepsilon_a} + \frac{(V_{\mathrm{HF}})^a_{\tilde{i}} Q^{\tilde{i}}_a + Q^a_{\tilde{i}} (V_{\mathrm{HF}})^{\tilde{i}}_a}{\varepsilon_{\tilde{i}} - \varepsilon_a} - \frac{Q^i_{\tilde{i}} Q^{\tilde{i}}_i - (V_{\mathrm{HF}})^i_{\tilde{i}} Q^{\tilde{i}}_i - Q^i_{\tilde{i}} (V_{\mathrm{HF}})^{\tilde{i}}_i}{\varepsilon_i - \varepsilon_{\tilde{i}}}, \tag{10}$$

$$\mathcal{E}^{(2)}_{\mathrm{QED},2} = \frac{1}{4} \frac{\bar{g}^{ab}_{ij} \bar{g}^{ij}_{ab}}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} + \left[\frac{1}{2} \frac{\bar{g}^{ab}_{i\tilde{j}} \bar{g}^{i\tilde{j}}_{ab}}{\varepsilon_i + \varepsilon_{\tilde{j}} - \varepsilon_a - \varepsilon_b} - \frac{1}{4} \frac{\bar{g}^{ij}_{\tilde{i}\tilde{j}} \bar{g}^{\tilde{i}\tilde{j}}_{ij}}{\varepsilon_{\tilde{i}} + \varepsilon_{\tilde{j}} - \varepsilon_i - \varepsilon_j} - \frac{1}{2} \frac{\bar{g}^{ia}_{\tilde{i}\tilde{j}} \bar{g}^{\tilde{i}\tilde{j}}_{ia}}{\varepsilon_{\tilde{i}} + \varepsilon_{\tilde{j}} - \varepsilon_i - \varepsilon_a} \right], \tag{11}$$

$$(V_{\mathrm{HF}})^q_p = \bar{g}^{qj}_{pj}, \quad \bar{g}^{rs}_{pq} = g^{rs}_{pq} - g^{rs}_{qp}, \quad g^{rs}_{pq} = \langle pq|g(1,2)|rs\rangle, \tag{12}$$

$$Q^q_p = -\frac{1}{2} \bar{g}^{qs}_{ps} \mathrm{sgn}(\varepsilon_s), \tag{13}$$

and by the configuration space approach (CSA) [23]

$$\mathcal{E}^{(2)}_{\mathrm{CS}} = \mathcal{E}^{(2)}_{\mathrm{CS},1} + \mathcal{E}^{(2)}_{\mathrm{CS},2}, \tag{14}$$

$$\mathcal{E}^{(2)}_{\mathrm{CS},1} = 0, \tag{15}$$

$$\mathcal{E}^{(2)}_{\mathrm{CS},2} = \frac{1}{4} \frac{\bar{g}^{ab}_{ij} \bar{g}^{ij}_{ab}}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} + \left[\frac{1}{4} \frac{\bar{g}^{\tilde{i}\tilde{j}}_{ij} \bar{g}^{ij}_{\tilde{i}\tilde{j}}}{\varepsilon_i + \varepsilon_j - \varepsilon_{\tilde{i}} - \varepsilon_{\tilde{j}}} + \frac{1}{2} \frac{\bar{g}^{a\tilde{j}}_{ij} \bar{g}^{ij}_{a\tilde{j}}}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_{\tilde{j}}} \right]. \tag{16}$$

Here, the orbital indices $\{i, j\}$ and $\{a, b\}$ refer to the occupied and unoccupied Dirac-Hartree-Fock (DHF) positive energy states (PES), respectively, while $\{\tilde{i}, \tilde{j}\}$ to the DHF negative energy states (NES). The Einstein summation convention over repeated indices has been assumed. The HF-like potential Q (13) describes the electron vacuum polarization (EVP) and self-energy (ESE) and stems from the "charge-conjugated contraction" [21, 22] of fermion operators when normal ordering the second quantized two-body interaction $\frac{1}{2} g^{rs}_{pq} a^\dagger_p a^\dagger_q a_s a_r$ with respect to the filled Dirac sea (which regards the NES as holes). By contrast, the CSA [23] has adopted an empty Dirac picture (which regards the NES as particles) and hence breaks charge conjugation symmetry.

The comparison reveals two points: (a) The CSA misses totally the leading QED EVP-ESE term (10). (b) Even after neglecting this term, $\mathcal{E}^{(2)}_{\mathrm{CS}}$ (14) agrees with $\mathcal{E}^{(2)}_{\mathrm{QED}}$ (7) only in the terms involving solely the PES, but differs from the latter in all the terms involving the NES (see the bracketed terms in Eqs. (9), (11),

and (16)). It can hence be concluded that the no-pair approximation (NPA) is a conceptual must for the CSA, rather than merely an effective means for avoiding the (in)famous Brown-Ravenhall disease [23] arising from the last, "continuum dissolution" term of Eq. (16). This gives rise to a first paradox: Four-component relativistic correlation must work with the first quantized DC/DCG/DCB Hamiltonian *operator* for fast convergence and the NPA for right physics, but these two are mutually exclusive. The PES-projected Hamiltonian is effectively a second quantized Hamiltonian and hence does not have the required operator form for the kinetic energy. Moreover, the finite spectrum of the PES-projected Hamiltonian, limited by the given basis set (GBS), implies that the effect of the correlation factor f_{12} is just null: The r_{12}-dependent geminal basis functions try to simulate a correlation space that is complementary to the orbital product space but that is not part of the Hamiltonian. In essence, any second quantized Hamiltonian is incompatible with explicit correlation. It will be shown below that this incompatibility problem can be resolved by the "extended no-pair projection" [20].

2. The must NPA of CSA also implies that the term "exact relativistic wave function" is physically meaningless (NB: QED works exclusively with electron and photon propagators and even has no Hamiltonian and wave function in the sense of eigvenvalue problem). Even if it were physical, the "exact relativistic wave function" is not normalizable [16]. On the other, it is hardly possible to get the analytical structures of the projected relativistic wave functions, simply because a unique and exact projector does not exist and different projectors may lead to different analytical structures for the wave functions. This represents again a paradox: The analytical structures of the wave functions are needed for guiding explicitly correlated wave function *Ansätze*, but they are not obtainable. It will be shown below that the direct knowledge on the analytical structures of the projected relativistic wave functions is not really needed. Rather, the knowledge on the "exact relativistic wave functions" can directly be transplanted to the NPA. Note in particular that the mathematical form of the asymptotic behavior of a wave function at the coalescence point of two electrons is independent of the physical nature of the state, bound, or scattering.
3. The recent analysis [16] reveals that, at the coalescence point of two electrons, the wave functions of the DCG Hamiltonian are regular, while those of the DC and DCB Hamiltonians have weak singularities of the type r_{12}^{ν}, with ν being negative and of $\mathcal{O}(\alpha^2)$. Moreover, there does not exist a simple relativistic analog of the nonrelativistic cusp condition [1, 2]. That is, the relation between the amplitudes $\Psi^{(\nu+1)}$ of $\mathcal{O}(r_{12}^{\nu+1})$ and $\Psi^{(\nu)}$ of $\mathcal{O}(r_{12}^{\nu})$ is not universal but is dependent on the energy and state. This warrants in principle a rational correlation factor f_{12}^{XY} for each block Ψ^{XY} of the wave function Ψ,

$$f_{12}^{XY} = r_{12}^{\nu}\left(\frac{a_{11}^{XY} + a_{12}^{XY} r_{12}}{a_{21}^{XY} + a_{22}^{XY} r_{12}}\right), \quad X, Y = L, S, \tag{17}$$

where the parameters a_{ij}^{XY} are to be optimized. However, this is hardly practical. Given that the asymptotic behaviors r_{12}^{ν} of the DC/DCG/DCB wave functions

affect only an extremely small neighborhood (with radius $R_c \sim \mathcal{O}(\alpha^2)$) of the coalescence point, it can be expected that they are only of minor importance for the calculation of the electronic energy, thanks to the suppression by the volume element $4\pi r_{12}^2$ for very small r_{12}. Anyhow, it is the extended region away from the coalescence point and the overall shape of the correlation hole that are really important [24]. In this extended region, the asymptotic behaviors of the "exact relativistic wave functions" are still dominated by the nonrelativistic limit (nrl). We are then led to a scenario of using a common nonrelativistic correlation factor f_{12}, e.g., the Slater-type [10]

$$f_{12} = -\frac{1}{\gamma}\exp(-\gamma r_{12}) = -\frac{1}{\gamma} + r_{12} + \mathcal{O}\left(r_{12}^2\right), \tag{18}$$

for all the blocks Ψ^{XY} of the relativistic wave function. Still, however, how to apply this factor remains uncertain due to the fact [25] that the two limits $\alpha \to 0$ and $r_{12} \to 0$ do not commute (vide post).

Given the above formal issues, there have been attempts toward relativistic explicitly correlated wave function methods, based on either the DC Hamiltonian [26–29] or its approximate variants [30–33]. As said above, the former type of calculations suffer from unphysical contributions from the NES, whereas the latter work well only for not too heavy atoms. For heavy and super-heavy atoms, four-component relativistic explicitly correlated methods based on the no-pair DC/DCG/DCB Hamiltonian should be developed.

R-MP2-F12

To show how to resolve the issues raised previously for relativistic explicit correlation, we consider the explicitly correlated second-order Møller-Plesset perturbation theory (R-MP2-F12). First of all, the incompatibility between the PES-projected Hamiltonian and explicit correlation can be resolved by introducing an *extended* no-pair Hamiltonian [20], viz.,

$$\hat{H}_+ = \begin{pmatrix} \hat{P}\hat{H}_{\mathrm{CS}}\hat{P} & \hat{P}\hat{H}_{\mathrm{CS}}\hat{Q} \\ \hat{Q}\hat{H}_{\mathrm{CS}}\hat{P} & \hat{Q}\hat{H}_{\mathrm{CS}}\hat{Q} \end{pmatrix}, \tag{19}$$

$$\hat{H}_{\mathrm{CS}} = \sum_k^N \hat{D}_k + \sum_{k>l}^N \hat{g}_{kl}, \tag{20}$$

$$\hat{D} = c\vec{\alpha}\cdot\vec{p} + \beta c^2 + V_{\mathrm{ext}} I_4, \quad V_{\mathrm{ext}}(\vec{r}) = -\sum_A^{N_A} \frac{Z_A}{|\vec{r} - \vec{R}_A|}, \tag{21}$$

$$\hat{g}_{kl} = d_C \frac{1}{r_{kl}} + d_G \frac{\vec{\alpha}_k \cdot \vec{\alpha}_l}{r_{kl}} + d_R \frac{(\vec{\alpha}_k \cdot \hat{r}_{kl})(\vec{\alpha}_l \cdot \hat{r}_{kl})}{r_{kl}}, \quad \hat{r}_{kl} = \frac{\vec{r}_{kl}}{r_{kl}}, \tag{22}$$

where D_k is the Dirac operator (21) for electron k. The Coulomb, Coulomb-Gaunt, and Coulomb-Breit interactions are represented by $\hat{g}_{kl}$ with $(1, 0, 0)$, $(1, -1, 0)$, and $(1, -1/2, -1/2)$ for the coefficients (d_C, d_G, d_R), respectively. The projectors $\hat{P}$ and $\hat{Q}$ in $\hat{H}_+$ (19) are to act on the conventional $(\hat{C}|0\rangle)$ and explicit $(\hat{X}|0\rangle)$ correlation spaces, respectively. To construct such projectors, we first decompose the identity operator as

$$\hat{1} = \hat{\Lambda}^+ + \hat{\Lambda}^-, \quad \hat{\Lambda}^+ = \hat{O}^+ + \hat{V}^+ + \hat{\tilde{V}}^+, \quad \hat{\Lambda}^- = \hat{V}^- + \hat{\tilde{V}}^-, \tag{23}$$

where $\hat{O}^+$, $\hat{V}^+$, and $\hat{V}^-$ are the respective projectors for the occupied PES, unoccupied PES, and NES defined by the GBS, whereas $\hat{\tilde{V}}^+$ and $\hat{\tilde{V}}^-$ are the corresponding orthogonal complements that vanish in the complete basis set (CBS) limit. We then have

$$\hat{P}_{12} = (\hat{O}_1^+ + \hat{V}_1^+)(\hat{O}_2^+ + \hat{V}_2^+), \tag{24}$$

$$\hat{Q}_{12} = (\hat{\Lambda}_1^+ - \hat{O}_1^+)(\hat{\Lambda}_2^+ - \hat{O}_2^+) - \hat{V}_1^+ \hat{V}_2^+. \tag{25}$$

The particular form of $\hat{Q}_{12}$ (25) is to ensure strong orthogonality to the vacuum and orthogonality to the conventional products of virtual PES and will vanish in the CBS limit. Since the orthogonal complement $\hat{\tilde{V}}^+$ is unknown, a suitable approximation has to be introduced. In view of the definitions (23), we consider the following option [16]:

$$\hat{\tilde{V}}^+ = \hat{1} - (\hat{O}^+ + \hat{V}^+) - (\hat{V}^- + \hat{\tilde{V}}^-) \simeq \hat{1} - (\hat{O}^+ + \hat{V}^+) - \hat{V}^-, \tag{26}$$

which amounts to neglecting the negative energy complement $\hat{\tilde{V}}^-$. This leads to

$$\hat{\Lambda}^+ = \hat{1} - \hat{\Lambda}^- \simeq \hat{1} - \hat{V}^-, \tag{27}$$

$$\hat{Q}_{12} \simeq (\hat{1}_1 - \hat{O}_1^+ - \hat{V}_1^-)(\hat{1}_2 - \hat{O}_2^+ - \hat{V}_2^-) - \hat{V}_1^+ \hat{V}_2^+ \tag{28}$$

$$= (\hat{1}_1 - \hat{O}_1^+ - \hat{V}_1^-)(\hat{1}_2 - \hat{O}_2^+ - \hat{V}_2^-)(\hat{1}_1 \hat{1}_2 - \hat{V}_1^+ \hat{V}_2^+). \tag{29}$$

It is interesting to see that the so-defined $\hat{Q}_{12}$ is formally in line with the filled Dirac picture: Both $\hat{O}^+$ and $\hat{V}^-$ can be viewed as occupied. As a whole, $\hat{Q}_{12}$ ensures strong orthogonality to the "vacuum" $\hat{O}^+ + \hat{V}^-$ as well as orthogonality to the conventional correlation subspace $\hat{V}_1^+ \hat{V}_2^+$. In the nrl, the $\hat{Q}_{12}$ operator (28) becomes

$$Q_{12} \overset{c\to\infty}{\to} (\hat{1}_1^+ - \hat{O}_1^+)(\hat{1}_2^+ - \hat{O}_2^+) - \hat{V}_1^+ \hat{V}_2^+, \tag{30}$$

which is known as *Ansatz* 3 in nonrelativistic F12 methods [34]. For comparison, Ten-no and Yamaki [35] proposed a different approximation for $\hat{\tilde{V}}^+$

$$\hat{\bar{V}}^+ \simeq \hat{1}^L - \hat{P}^L, \tag{31}$$

where $\hat{1}^L$ and P^L are the projectors onto the large-component bases of the CBS and GBS, respectively. They then invoked the following strong orthogonality projector (*Ansatz* 2β):

$$\hat{\mathcal{Q}}_{12} = (\hat{1}_1^L - \hat{P}_1^L + \hat{V}_1^+)(\hat{1}_2^L - \hat{P}_2^L + \hat{V}_2^+), \tag{32}$$

which reduces to $\hat{V}_1^+ \hat{V}_2^+$ in the CBS limit and $(\hat{1}_1^+ - \hat{O}_1^+)(\hat{1}_2^+ - \hat{O}_2^+)$ in the nrl. Since the large-component space $\hat{P}^L$ and hence its orthogonal complement $\hat{1}^L - \hat{P}^L$ are composites of the PES and NES, the so-defined $\hat{\mathcal{Q}}_{12}$ (32) suffers from severe unphysical contamination from the NES [20]. By contrast, the present $\hat{Q}_{12}$ (28) is essentially immune from such contamination because of the huge gap (which is in the order of $2N^S mc^2$, with N^S being the number of NES) between the occupied PES and the energetically lowest NES described by the GBS.

The Hamiltonian H_+ (19) alongside with the projectors $\hat{P}_{12}$ (24) and $\hat{Q}_{12}$ (28) can now act on the unprojected many-electron basis functions $(\hat{C}|0\rangle, \hat{X}|0\rangle)^T$, *independently* of the wave function *Ansatz*. Taking MP2 as an example, we have the following Hylleraas functional for a given occupied electron pair $|ij\rangle$:

$$H^{(ij)} = \langle u_{ij}|\hat{F}_{12} - \varepsilon_i - \varepsilon_j|u_{ij}\rangle + \langle u_{ij}|\hat{g}_{12}|ij\rangle + \langle ij|\hat{g}_{12}|u_{ij}\rangle, \tag{33}$$

$$\hat{F}_{12} = \hat{F}_1 + \hat{F}_2, \tag{34}$$

$$|u_{ij}\rangle = |ab\rangle t_{ab}^{ij} + \hat{Q}_{12}\hat{\mathcal{R}}_{12}|kl\rangle c_{kl}^{ij}, \tag{35}$$

$$\hat{\mathcal{R}}_{12} = f_{12}\left(\frac{3}{8} + \frac{1}{8}\hat{p}_{12}\right), \tag{36}$$

where $\hat{F}_p$ is the Fock operator and ε_p the corresponding eigenvalue, while $\hat{p}_{12}$ in $\hat{\mathcal{R}}_{12}$ (36) permutes merely the spatial coordinates [36]. The functional (33) can be decoupled into a conventional (MP2) part and an F12 correction term by omitting their couplings. The amplitudes c_{kl}^{ij} are then obtained by solving the linear system of equations

$$B_{mn}^{kl(ij)} C_{kl}^{ij} = -V_{mn}^{ij}, \tag{37}$$

$$B_{mn}^{kl(ij)} = \langle mn|\hat{\mathcal{R}}_{12}\hat{Q}_{12}(\hat{F}_{12} - \varepsilon_i - \varepsilon_j)\hat{Q}_{12}\hat{\mathcal{R}}_{12}|kl\rangle, \tag{38}$$

$$V_{mn}^{ij} = \langle mn|\hat{\mathcal{R}}_{12}\hat{Q}_{12}g_{12}|ij\rangle. \tag{39}$$

Alternatively, one can start from the first-order equation

$$(\hat{F}_{12} - \varepsilon_i - \varepsilon_j)|\omega_{ij}\rangle + \hat{g}_{12}|ij\rangle = 0 \tag{40}$$

for the *unprojected* pair function $|\omega_{ij}\rangle$ that is strongly orthogonal to $|ij\rangle$ but contaminated by NES. Thanks to the relation $[\hat{F}_{12}, \hat{\Lambda}_{12}^{++}] = 0$, multiplying Eq. (40) from the left by $\hat{\Lambda}_{12}^{++} = \hat{\Lambda}_1^+ \hat{\Lambda}_2^+$ leads to

$$\hat{\Lambda}_{12}^{++}(\hat{F}_{12} - \varepsilon_i - \varepsilon_j)|u_{ij}\rangle + \hat{\Lambda}_{12}^{++}\hat{g}_{12}\hat{\Lambda}_{12}^{++}|ij\rangle = 0, \quad |u_{ij}\rangle = \hat{\Lambda}_{12}^{++}|\omega_{ij}\rangle, \tag{41}$$

which is just the stationarity condition for the *projected* functional (33), further approximated to Eq. (37) though. This simple manipulation shows that one can first construct the unprojected pair function $|\omega_{ij}\rangle$ and then from it the projected one $|u_{ij}\rangle$. This trivial result has an important implication: The knowledge on the analytic structures of the projected wave functions is not really needed. Rather, that of the "exact relativistic wave functions" can directly be transplanted to the no-pair approximation. This holds at least to first order in the electron-electron interaction.

Apart from the usual two-electron integrals, the following kinds of integrals

$$f_{12}, \ \frac{f_{12}}{r_{12}}, \ f_{12}^2, \ [\hat{T}_1, f_{12}], \ [[\hat{T}_1 + \hat{T}_2, f_{12}], f_{12}] \tag{42}$$

are also required. They can all be evaluated analytically for Gaussian-type spinors. Plugging the Dirac kinetic operator in $[\hat{T}_1, f_{12}]$ yields

$$[\vec{\alpha}_1 \cdot \vec{p}_1, f_{12}] = [\vec{\alpha}_1 \cdot \vec{p}_{12}, f_{12}] = -if'_{12}(\vec{\alpha}_1 \cdot \hat{r}_{12}), \tag{43}$$

such that

$$[[\hat{T}_1 + \hat{T}_2, f_{12}], f_{12}] = 0. \tag{44}$$

This strongly suggests the use of "approximation C" [37] when evaluating the integral $\langle ij|f_{12}\hat{F}_{12}f_{12}|kl\rangle$ involving the exchange operator $\hat{K}_{12}$, viz.,

$$\begin{aligned}
\langle ij|f_{12}\hat{F}_{12}f_{12}|kl\rangle &= \langle ij|f_{12}(\hat{F}_{12} + \hat{K}_{12})f_{12}|kl\rangle - \langle ij|f_{12}\hat{K}_{12}f_{12}|kl\rangle \\
&= \frac{1}{2}\langle ij|[f_{12}, [\hat{F}_{12} + \hat{K}_{12}, f_{12}]]|kl\rangle \\
&\quad + \frac{1}{2}\langle ij|[\hat{F}_{12} + \hat{K}_{12}, f_{12}^2]_+|kl\rangle - \langle ij|f_{12}\hat{K}_{12}f_{12}|kl\rangle \\
&= \frac{1}{2}\langle ij|(\hat{F}_{12} + \hat{K}_{12})f_{12}^2|kl\rangle + \frac{1}{2}\langle ij|f_{12}^2(\hat{F}_{12} + \hat{K}_{12})|kl\rangle \\
&\quad - \langle ij|f_{12}\hat{K}_{12}f_{12}|kl\rangle.
\end{aligned} \tag{45}$$

All the three terms can then be approximated by the RI with a kinetically balanced complementary auxiliary basis set [34].

In sum, the use of the extended no-pair projection (19) with Q_{12} (28), the *Ansatz* (35), and a nonrelativistic correlation factor f_{12} renders no-pair four-component relativistic explicitly correlated methods completely parallel to the nonrelativistic counterparts.

Yet, additional remarks are still necessary. The *Ansatz* (35) implies that a common correlation factor f_{12} is applied to all the components $|k^X l^Y\rangle$ $(X, Y = L, S)$ of the determinant $|kl\rangle$, in close analogy with the nonrelativistic counterpart. Since $|kl\rangle$ reduces to $Z_{12}|k^L l^L\rangle$ in the nrl, with Z_{12} being the so-called geminal kinetic balance (GKB)

$$Z_{12} = \begin{pmatrix} 1 \\ \frac{\vec{\sigma}_2\cdot\vec{p}_2}{2c} \\ \frac{\vec{\sigma}_1\cdot\vec{p}_1}{2c} \\ \frac{(\vec{\sigma}_1\cdot\vec{p}_1)(\vec{\sigma}_2\cdot\vec{p}_2)}{4c^2} \end{pmatrix}, \tag{46}$$

this *Ansatz* (35) can be characterized as "$f_{12} * Z_{12}$." A possible drawback of this *Ansatz* resides in that the nonrelativistic cusp condition [1, 2] cannot strictly be guaranteed [35]. To achieve this, one may adopt the "$Z_{12} * f_{12}$" *Ansatz* [26–28, 35]:

$$|u_{ij}\rangle = |ab\rangle t_{ab}^{ij} + Q_{12} Z_{12} \hat{\mathcal{R}}_{12} |k^L l^L\rangle c_{kl}^{ij}, \tag{47}$$

which originates from the limiting process of first taking $\alpha \to 0$ and then $r_{12} \to 0$. In contrast, the previous *Ansatz* (35) amounts to taking the reverse process. Since the two limits do not commute [25] and, in particular, a relativistic analog of the universal nonrelativistic cusp condition does not exist [16], there seem no a priori arguments to favor which of the two options. Yet, the following points might fell on the "$f_{12} * Z_{12}$" *Ansatz* (35):

(a) Due to symmetry reason [16], the Ψ^{SS} and Ψ^{LL} components of the "relativistic exact wave functions" are of the same magnitude at the coalescence point of two electrons. A direct deduction is that the leading term $\bar{\Psi}_0^{SS}$ of the expansion, $\bar{\Psi}^{SS} = c^2\Psi^{SS} = \bar{\Psi}_0^{SS} + \alpha^2\bar{\Psi}_2^{SS} + \mathcal{O}(\alpha^4)$, is $\mathcal{O}(c^2)$ at small r_{12}, which *invalidates* the $\mathcal{O}(c^0)$ assumption of $\bar{\Psi}_0^{SS}$ in relativistic perturbation theory. Note that this observation results from the limiting process of first taking $r_{12} \to 0$ and then $\alpha \to 0$. When the limiting process is reversed, it can be deduced [19] from the nonrelativistic cusp condition [1] as well as the geminal kinetic balance (46) that $\bar{\Psi}_0^{SS}$ behaves as r_{12}^{-1} at the coalescence point of two electrons, thereby far more singular than Ψ^{SS} that goes as r_{12}^{ν}, with ν being negative but only of order $\mathcal{O}(\alpha^2) \sim 10^{-5}$. As a result, the second-order relativistic correction to correlation, $\alpha^4 E^{(4,2)}$ in Eq. (5), is already divergent due to the involvement of $\langle\bar{\Psi}_0^{SS}|\frac{1}{r_{12}}|\bar{\Psi}_0^{SS}\rangle$, whose integrand goes as r_{12}^{-3} at small r_{12}. Even more severe divergence will occur to higher-order relativistic corrections [16]. It can therefore be argued whether the combination of nonrelativistic cusp and kinetic balance conditions is really valid for the DC Hamiltonian, nonexpanded though.

(b) The one-electron counterpart of $Z_{12}|k^L l^L\rangle$ is

$$|k^S\rangle = Z_1|k^L\rangle, \quad Z_i = \frac{\vec{\sigma}_i \cdot \vec{p}_i}{2c}, \tag{48}$$

but which is not the kinetic balance per se. Rather, it is its realization at the basis set level, viz.,

$$|\chi_\mu^S\rangle = Z_1|\chi_\mu^L\rangle, \tag{49}$$

that is the restricted kinetic balance (RKB) [38]. While the prescription (49) for generating the small-component basis is numerically stable, the direct use of relation (48) results in variational collapse. Therefore, it is likely that the second term of Eq. (47) will lead to "two-electron prolapse," to a lesser extent than the one-electron collapse though. To remedy this defect, one can replace Eq. (47) by

$$|u_{ij}\rangle = |ab\rangle t_{ab}^{ij} + Q_{12}|\Phi_{kl}\rangle c_{kl}^{ij}, \tag{50}$$

$$|\Phi_{kl}^{LL}\rangle = \hat{\mathcal{R}}_{12}|k^L l^L\rangle, \tag{51}$$

$$|\Phi_{kl}^{LS}\rangle = Z_2\hat{\mathcal{R}}_{12}|k^L \tilde{l}^L\rangle, \quad |\tilde{l}^L\rangle = |\chi_\mu^L\rangle C_{\mu l}^S, \tag{52}$$

$$|\Phi_{kl}^{SL}\rangle = Z_1\hat{\mathcal{R}}_{12}|\tilde{k}^L l^L\rangle, \quad |\tilde{k}^L\rangle = |\chi_\mu^L\rangle C_{\mu k}^S, \tag{53}$$

$$|\Phi_{kl}^{SS}\rangle = Z_1 Z_2\hat{\mathcal{R}}_{12}|\tilde{k}^L \tilde{l}^L\rangle, \tag{54}$$

$$|k^X l^Y\rangle = \frac{1}{\sqrt{2}}\left(\psi_k^X(1)\otimes\psi_l^Y(2) - \psi_l^X(1)\otimes\psi_k^Y(2)\right), \quad X, Y = L, S, \tag{55}$$

where $|\tilde{k}^L\rangle$ and $|\tilde{l}^L\rangle$ are the so-called pseudo-large components [39], which have the same symmetry properties and nrl as the large components $|p^L\rangle$ and are represented by linear combinations of the large-component basis functions but with the small-component coefficients. Equations (51)–(54) represent hence the true GKB, just like the one-electron RKB (49). Note also that the antisymmetrization (55) permutes the spinor indices rather than the coordinates. For valence electrons, the pseudo-large and large components are very similar, such that Eqs. (50) and (47) do not differ much. However, this is not the case for core or semi-core electrons.

As a matter of fact, the GKB realized by Eqs. (51)–(54) can be deduced naturally from the modified Dirac and Dirac-Coulomb equations [39] or simply from the parities of the small and large components [40]. It is *not* an approximation per se (apart from the incompleteness error in the large-large-component basis) and ensures the correct nonrelativistic cusp conditions [1, 2]. This is a strong point to favor the "$Z_{12} * f_{12}$" *Ansatz* (50) over the "$f_{12} * Z_{12}$" one (35) (which cannot ensure the nonrelativistic cusp conditions). However, it should be noted that, because of

the use of common correlation factor, the *Ansatz* (50) does not respect the fact [16] that the Ψ^{XY} $(X, Y = L, S)$ components of the DC/DCG/DCB wave functions Ψ have the same asymptotic coalescence behaviors r_{12}^{v}. In principle, the powers of r_{12} for the $|\Phi_{kl}^{LS}\rangle$, $|\Phi_{kl}^{SL}\rangle$, and $|\Phi_{kl}^{SS}\rangle$ components should be higher than that for $|\Phi_{kl}^{LL}\rangle$ by one, one, and two, respectively (for more details, see Ref. [16]). However, the implementation becomes then more involved. Given that the large-large components have predominant contributions to correlation energy, the use of common correlation factor should be acceptable.

Summary

The formal aspects of relativistic explicit correlation have been scrutinized in depth, with the main take-home messages being:

1. The perturbative treatment of relativistic corrections to correlation makes sense only at lowest order, viz., $\alpha^2\lambda^n E^{(2,n)}$ with $n \geq 2$, because higher-order terms, $\alpha^{2m}\lambda^n E^{(2m,n)}$ with $m > 1$ and $n \geq 2$, suffer from unsurmountable divergence. This should apply also to standard orbital-based correlation methods for basis sets of increasing quality may soon approach the divergence.
2. Any four-component relativistic explicitly correlated method must work with the *extended* no-pair projected DC/DCG/DCB Hamiltonian [20]. Here, the DCG Hamiltonian is particularly recommended for two reasons: (a) The asymptotic behaviors r_{12}^{v} of its wave functions are regular at the coalescence points of two electrons and have the correct nrl $(v = l)$ [16]. (b) It describes correctly both the spin-same-orbit and spin-other-orbit couplings, in contrast with the DC Hamiltonian that describes only the spin-same-orbit coupling.
3. The $\hat{Q}_{12}$ projector (28) is clearly advantageous over $\hat{\mathcal{Q}}_{12}$ (32) insofar as the contamination of NES is concerned.
4. Due to noncommutation of the $\alpha \to 0$ and $r_{12} \to 0$ limits, a decisive choice of the "$f_{12} * Z_{12}$" (35) and "$Z_{12} * f_{12}$" (50) *Ansätze* for incorporating the explicit correlation factor f_{12} cannot be made a priori. Both have pros and cons. Nonetheless, more weights can be put on the latter for three reasons: (a) The two-electron functions are truly kinetically balanced. (b) The nonrelativistic cusp conditions can be ensured. (c) The large-large components dominate the contributions to correlation energy, such that the violation of the same asymptotic coalescence behaviors for all the components of the wave function does not really matter.
5. Given the one-electron and short-range nature of spin-orbit couplings, it is more practical to do only spin-free relativistic explicit correlation. Here, the *extended* no-pair projected spin-free part of the modified DC (sf-MDC) Hamiltonian can be adopted. The (spin-free) rational generator [36] $\hat{\mathcal{R}}_{12}$ (36) then recovers its original meaning. Note also that the asymptotic behaviors of the wave functions of the sf-MDC Hamiltonian are less singular than those of the original DC ones

[16] [NB: This cannot be taken as a point to turn down the previous arguments against the "$Z_{12} * f_{12}$" *Ansatz* (47)/(50), for the same problems still persist. See however point (4) above]. The interplay between spin-orbit coupling and correlation can then be accounted for by using standard orbital-based methods in conjunction with the newly obtained two-component spin-orbit Hamiltonians [41, 42]. Even the spin-free PES and NES required by the sf-MDC explicit correlation can also be generated by the spin-free exact two-component (X2C) Hamiltonians [40]. That is, only the explicit correlation step involves the sf-MDC Hamiltonian *operator*. It is likely that the increment between the explicit and implicit (orbital-based) correlations can be accounted for at some low level of methodology, e.g., MP2 or CEPA (coupled electron pair approximation). It is very unfortunate that the X2C Hamiltonians cannot be adapted to explicit correlation for they are defined only algebraically.

6. When the treatment of no-pair correlation has reached an accuracy of approximately 10^{-2} eV, the $\mathcal{O}(Z\alpha)^3$ error inherent in the NPA can be corrected by adding in the terms of $\mathcal{E}^{(2)}_{\mathrm{QED}}$ (7) involving the NES.

Acknowledgements The research of this work was supported by grants from the National Natural Science Foundation of China (Project Nos. 21033001, 21273011, 21290192, and 11471025).

References

1. Kato T (1957) On the eigenfunctions of many-particle systems in quantum mechanics. Commun Pure Appl Math 10:151
2. Pack RT, Brown WB (1966) Cusp conditions for molecular wavefunctions. J Chem Phys 45:556
3. Hylleraas EA (1929) Neue berechnung der energie des heliums im grundzustande, sowie des tiefsten terms von ortho-helium. Z Physik 54:347
4. Kutzelnigg W (1985) r_{12}-dependent terms in the wave function as closed sums of partial wave amplitudes for large l. Theor Chim Acta 68:445
5. Klopper W, Manby FR, Ten-no S, Valeev EF (2006) R12 methods in explicitly correlated molecular electronic structure theory. Int Rev Phys Chem 25:427
6. Shiozaki T, Hirata S, Valeev EF (2010) Explicitly correlated coupled-cluster methods. Ann Rev Comput Chem 5:131
7. Hättig C, Klopper W, Köhn A, Tew DP (2012) Explicitly correlated electrons in molecules. Chem Rev 112:4
8. Kong L, Bischoff FA, Valeev EF (2012) Explicitly correlated R12/F12 methods for electronic structure. Chem Rev 112:75
9. Kutzelnigg W, Klopper W (1991) Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory. J Chem Phys 94:1985
10. Ten-no S (2004) Initiation of explicitly correlated Slater-type geminal theory. Chem Phys Lett 398:56
11. Salomonson S, Öster P (1989) Relativistic all-order pair functions from a discretized single-particle Dirac Hamiltonian. Phys Rev A 40:5548
12. Ottschofski E, Kutzelnigg W (1997) Direct perturbation theory of relativistic effects for explicitly correlated wave functions: the He isoelectronic series. J Chem Phys 106:6634

13. Halkier A, Helgaker T, Klopper W, Olsen J (2000) Basis-set convergence of the two-electron Darwin term. Chem Phys Lett 319:287
14. Kutzelnigg W (2008) Relativistic corrections to the partial wave expansion of two-electron atoms. Int J Quantum Chem 108:2280
15. Hill RN (1985) Rates of convergence and error estimation formulas for the Rayleigh–Ritz variational method. J Chem Phys 83:1173
16. Li Z, Shao S, Liu W (2012) Relativistic explicit correlation: coalescence conditions and practical suggestions. J Chem Phys 136:144117
17. Schwartz C (1962) Importance of angular correlations between atomic electrons. Phys Rev 126:1015
18. Schwartz C (1963) Estimating convergence rates of variational calculations. Methods Comput Phys 2:241
19. Kutzelnigg W (2002) Perturbation theory of relativistic effects. In: Schwerdtfeger P (ed) Relativistic electronic structure theory. Part 1. Fundamentals. Elsevier, Amsterdam, p 664
20. Liu W (2012) Perspectives of relativistic quantum chemistry: the negative energy cat smiles. Phys Chem Chem Phys 14:35
21. Liu W, Lindgren I (2013) Going beyond "no-pair relativistic quantum chemistry". J Chem Phys 139:014108
22. Liu W (2014) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59
23. Brown GE, Ravenhall DG (1951) On the interaction of two electrons. Proc R Soc A 208:552
24. Gilbert TL (1963) Interpretation of the rapid convergence of correlated wave functions. Rev Mod Phys 35:491
25. Kutzelnigg W (1989) Generalization of Kato's cusp conditions to the relativistic case. In: Mukherjee D (ed) Aspects of many-body effects in molecules and extended systems. Lecture notes in chemistry, vol 50. Springer, Berlin, p. 353
26. Kałokowska A (1997) Explicitly correlated trial functions in relativistic variational calculations. J Phys B: At Mol Opt Phys 30:2773
27. Pestka G, Bylicki M, Karwowski J (2007) Complex coordinate rotation and relativistic Hylleraas-CI: helium isoelectronic series. J Phys B: At Mol Opt Phys 40:2249
28. Pestka G, Bylicki M, Karwowski J (2012) Geminals in Dirac–Coulomb Hamiltonian eigenvalue problem. J Math Chem 50:510
29. Nakatsuji H, Nakashima H (2005) Analytically solving the relativistic Dirac-Coulomb equation for atoms and molecules. Phys Rev Lett 95:050407
30. Nakatsuka Y, Nakajima T, Nakata M, Hirao K (2010) Relativistic quantum Monte Carlo method using zeroth-order regular approximation Hamiltonian. J Chem Phys 132:054102
31. Nakatsuka Y, Nakajima T, Hirao K (2010) Electron-nucleus cusp correction scheme for the relativistic zeroth-order regular approximation quantum Monte Carlo method. J Chem Phys 132:174108
32. Bischoff FA, Klopper W (2010) Second-order electron-correlation and self-consistent spin-orbit treatment of heavy molecules at the basis-set limit. J Chem Phys 132:094108
33. Bischoff FA, Valeev EF, Klopper W, Janssen CL (2010) Scalar relativistic explicitly correlated R12 methods. J Chem Phys 132:214104
34. Valeev EF (2004) Improving on the resolution of the identity in linear R12 ab initio theories. Chem Phys Lett 395:190
35. Ten-no S, Yamaki D (2012) Communication: explicitly correlated four-component relativistic second-order Møller-Plesset perturbation theory. J Chem Phys 137:131101
36. Ten-no S (2004) Explicitly correlated second order perturbation theory: introduction of a rational generator and numerical quadratures. J Chem Phys 121:117
37. Kedžuch S, Milko M, Noga J (2005) Alternative formulation of the matrix elements in MP2-R12 theory. Int J Quantum Chem 105:929
38. Stanton RE, Havriliak S (1984) Kinetic balance: a partial solution to the problem of variational safety in Dirac calculations. J Chem Phys 81:1910

39. Dyall KG (1994) Second-order Møller-Plesset perturbation theory for molecular Dirac-Hartree-Fock wavefunctions. Theory for up to two open-shell electrons. J Chem Phys 100:2118
40. Liu W (2010) Ideas of relativistic quantum chemistry. Mol Phys 108:1679
41. Li Z, Xiao Y, Liu W (2012) On the spin separation of algebraic two-component relativistic Hamiltonians. J Chem Phys 137:154114
42. Li Z, Xiao Y, Liu W (2014) On the spin separation of algebraic two-component relativistic Hamiltonians: molecular properties. J Chem Phys 141:054111

Relativistic Density Functional Theory

18

Eberhard Engel

Contents

Introduction 548
Existence Theorem 550
Relativistic Kohn–Sham Equations 552
Relativistic Spin-Density Functional Theory 554
Relativistic Exchange–Correlation Functional 556
Orbital-Dependent Exchange–Correlation Functionals 557
Relativistic Optimized Potential Method 559
Relativistic Corrections in E_{xc}: I. Transverse Exchange 563
Relativistic Corrections in E_{xc}: II. Coulomb Exchange 567
Relativistic Corrections in E_{xc}: III. Coulomb Correlation 569
Relativistic Local Density Approximation 570
Relativistic Generalized Gradient Approximation 575
Relativistic Corrections in E_{xc}: IV. Bonding Properties 575
Summary 576
References 577

Abstract

This chapter gives an overview of relativistic density functional theory. Both its foundations, the existence theorem and the Kohn–Sham equations, and its core quantity, the exchange–correlation (xc) energy functional, are discussed. It is first outlined how a workable relativistic Kohn–Sham scheme can be obtained within the framework of quantum electrodynamics and which alternatives for its implementation are available. Particular emphasis is then placed on the relativistic corrections to the xc-functional. The modification of its functional form due to the relativistic motion of the electrons and their retarded interaction

E. Engel (✉)
Center for Scientific Computing, J.W. Goethe-Universität Frankfurt, Frankfurt/Main, Germany
e-mail: engel@th.physik.uni-frankfurt.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_18

via exchange of photons is distinguished from the effect resulting from insertion of a relativistic density into the functional. The difference between the relativistic xc-functional and its nonrelativistic form is studied in detail for the case of the exchange functional (which can be handled exactly via the optimized effective potential method). This analysis is complemented by some first-principles results for the correlation functional, relying on a perturbative approach. Finally, the accuracy of approximate relativistic xc-functionals, the local density and the generalized gradient approximation, is assessed on the basis of the exact results.

Keywords

Breit interaction • Current-current response function • Exact exchange • Exchange-correlation energy functional • Exchange-correlation magnetic field • Ground state energy functional • Ground state four current • Hartree energy • Interacting v-representability • Kinetic energy functional • Kohn-Sham equations • No-pair approximation • Optimized effective potential method • Relativistic generalized gradient • Approximation • Relativistic local density approximation • Relativistic spin-density functional theory • Transverse interaction • Variational equation

Introduction

Given the success of nonrelativistic DFT, the question for a relativistic generalization arises quite naturally. The appropriate framework for this generalization is provided by quantum electrodynamics (QED), as the most fundamental approach to the relativistic many-electron problem. As a result, relativistic density functional theory (RDFT) necessarily inherits the full complexity of QED (▸ Chap. 6, "QED Effects and Challenges"). Features such as the need for renormalization and the gauge freedom do not only surface on the formal level, i.e., in the foundations of RDFT, but also in the effective single-particle equations, making them extremely difficult to implement. The final goal of the RDFT formalism, however, is an efficient description of molecules and solids with truly heavy constituents, without the ambition to reach QED accuracy. The most important step towards this goal is the no-pair approximation, in which the contributions of virtual electron–positron pairs to all relevant quantities are neglected ▸ Chap. 11, "No-Pair Relativistic Hamiltonians: Q4C and X2C". The resulting relativistic Kohn–Sham (RKS) equations differ from their nonrelativistic counterpart by the relativistic form of the kinetic energy operator and the relativistic coupling between the particles and the effective RKS potential. As in the nonrelativistic situation, the total KS potential consists of an external, a Hartree, and an exchange–correlation (xc) component, which is obtained as functional derivative of the core quantity of RDFT, the xc-energy functional E_{xc}. Both the Hartree term and E_{xc} reflect the fact that in QED the electrons interact by the exchange of photons rather than via the instantaneous Coulomb interaction. Together with the relativistic kinematics of the electrons, this

retarded interaction leads to a modification of $E_{\rm xc}$, compared to its nonrelativistic form. It is the primary intention of this chapter to discuss the xc-functional of RDFT, with an emphasis on the role of the relativistic corrections in this functional.

The chapter starts with an overview of the foundations of QED-based RDFT, the existence theorem and the resulting Kohn–Sham (KS) formalism. In this approach, the four-current density j^μ is the fundamental variable which is used to represent observables such as the ground state energy. In practice, however, RDFT variants working with the charge and magnetization densities or the relativistic extensions of the nonrelativistic spin densities are utilized. The corresponding KS equations are therefore also summarized. For a rigorous assessment of the relevance of relativistic corrections in $E_{\rm xc}$, one has to resort to first-principles expressions for $E_{\rm xc}$. A first-principles treatment is in particular possible for the RDFT exchange $E_{\rm x}$. This functional, while known as an implicit functional of j^μ, is explicitly known only in terms of the KS four spinors. Its self-consistent application in the KS formalism relies on the relativistic optimized potential method, which is outlined next. On this basis, the properties of $E_{\rm xc}$ are analyzed more closely, addressing in particular the transverse interaction. The analysis of the exact $E_{\rm x}$ is complemented by results obtained with an MP2-type correlation functional. Finally, the relativistic extension of the local density approximation (LDA) as well as the generalized gradient approximation (GGA) are discussed. The resulting functionals are used to examine the importance of the relativistic corrections in $E_{\rm xc}$ for bonding/cohesive properties.

[The present text is based on Chapter 8 of [1].]

Notation

In this chapter, space-time points are denoted by four vectors:

$$x \equiv (x^\mu) = (ct, \boldsymbol{r}) = (ct, r^1, r^2, r^3) = (ct, r^i). \tag{1}$$

Greek (Minkowski) indices always extend from 0 to 3, Roman indices always denote spatial components, $i, j, \ldots = 1, 2, 3$. The associated metric tensor reads

$$g_{\mu\nu} \equiv \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -\delta_{ij} \end{pmatrix}. \tag{2}$$

Contraction of any four vector with the metric tensor transforms covariant into contravariant components and vice versa,

$$a_\mu = g_{\mu\nu}\, a^\nu \; ; \qquad a^\mu = g^{\mu\nu}\, a_\nu. \tag{3}$$

The spatial vector $\boldsymbol{a}$, which characterizes the actual physical three vector, consists of the contravariant components $\boldsymbol{a} = (a^1, a^2, a^3)$, from which the covariant components a_i differ by a minus sign, $a_i = -a^i$. The four gradient is abbreviated by

$$\frac{\partial}{\partial x^\mu} = \left(\frac{1}{c}\frac{\partial}{\partial t}, \boldsymbol{\nabla}\right). \tag{4}$$

The summation convention is used throughout,

$$a^\mu b_\mu \equiv \sum_{\mu=0}^{3} a^\mu b_\mu \; ; \qquad a^i b_i \equiv \sum_{i=1}^{3} a^i b_i = -\sum_{i=1}^{3} a^i b^i = -\boldsymbol{a}\cdot\boldsymbol{b}. \tag{5}$$

The four vector α^μ is defined in terms of the standard Dirac matrices γ^μ, $\boldsymbol{\alpha}$ and β as

$$\alpha^\mu = \gamma^0\gamma^\mu = (1, \boldsymbol{\alpha}) \qquad \beta = \gamma^0. \tag{6}$$

$e = |e|$ is used throughout this chapter.

Existence Theorem

The existence theorem of RDFT is based on bound-state quantum electrodynamics (▸ Chaps. 6, "QED Effects and Challenges" and ▸ 9, "Unifying Many-Body Perturbation Theory with Quantum Electrodynamics"). In this approach, the nuclei (including their magnetic moments) and all other external sources are represented in terms of a stationary four potential $V^\mu(\boldsymbol{x})$, while the electrons interact by the exchange of photons. Using the QED-Hamiltonian, it was shown [2–5] that there exists a one-to-one correspondence between the class of all ground states $|\Psi_0\rangle$ which just differ by gauge transformations and the associated ground state four-current density $j_0^\mu(\boldsymbol{x})$, provided the ground state is nondegenerate (up to gauge transformations, of course). Mathematically speaking, the ground state is a unique functional of the ground state four-current density, $|\Psi[j]\rangle$, once the gauge has been fixed universally by some suitable requirement on $V^\mu(\boldsymbol{x})$. If a particular ground state four-current density j_0 is inserted into this functional, one obtains exactly the associated ground state, $|\Psi_0\rangle = |\Psi[j_0]\rangle$ (with the gauge chosen).

As a result, the ground state energy can be expressed as a (four- current) density functional $E[j]$,

$$E[j] = \langle\Psi_0[j]|\hat{H}_{\mathrm{QED}}|\Psi_0[j]\rangle, \tag{7}$$

where $\hat{H}_{\mathrm{QED}}$ denotes the Hamiltonian of bound-state QED (in the following the short form "density functional" will also be used for functionals of j^μ). For this functional, one has a minimum principle [2–5],

$$E[j_0] < E[j] \qquad \text{for all four-current densities} \quad j^\mu \neq j_0^\mu, \tag{8}$$

with j_0^μ denoting the four-current density resulting from the actual external potential of the system.

Unlike in the case of nonrelativistic DFT, there is an ambiguity resulting from the gauge freedom of QED. Since a mere gauge transformation of V^μ does not affect the ground state energy, a unique relation between individual ground state wave functions and the gauge invariant four-current density of QED cannot exist, as long as the gauge is not fixed in some unique fashion. As a consequence, the issue of gauge invariance has to be addressed in the derivation of explicit functionals for RDFT.

Key to the proof of the existence theorem is the minimum principle for the ground state energy, in complete analogy to the nonrelativistic Hohenberg–Kohn argument [6]. In the case of QED, this minimum principle is a nontrivial issue, since expectation values such as the ground state energy and the four-current density diverge unless they are renormalized. The compatibility of the standard renormalization scheme of QED with the logic of the Hohenberg–Kohn argument has been demonstrated in [7, 8]. The necessity for renormalization also becomes apparent in all explicit expressions for (current) density functionals, both on the exact level and in the derivation of approximations. For a detailed discussion of the existence theorem and its field theoretical background, the reader is referred to [1].

The minimum principle (8) leads to the variational equation

$$\frac{\delta}{\delta j^\nu(\boldsymbol{r})} \left\{ E[j] - \mu \int d^3x \, j^0(\boldsymbol{x}) \right\} = 0, \tag{9}$$

where the subsidiary condition reflects current conservation,

$$\frac{\partial j^\mu}{\partial x^\mu} = \boldsymbol{\nabla} \cdot \boldsymbol{j}(\boldsymbol{r}) = 0 \qquad \Longrightarrow \qquad \int d^3r \, j^0(\boldsymbol{r}) = N. \tag{10}$$

Equation (9) allows one to recast the relativistic many-body problem as a minimization procedure for $E[j]$.

The transition from the minimum principle (8) to the variational equation (9) is based on the functional differentiability of $E[j]$, which, in turn, requires $E[j]$ to be defined on a sufficiently dense set of j^μ. This leads to the question whether, for any given function j^μ, one can find some V^μ, so that j^μ is obtained as ground state four-current density by solving the stationary relativistic many-body problem with V^μ (termed *interacting v-representability*). Given the counterexamples to interacting v-representability known for nonrelativistic DFT, one should not expect the domain of (7) to be sufficiently dense. As demonstrated by Eschrig and collaborators [9, 10], interacting v-representability (and thus differentiability) can, however, be ensured by a redefinition of $E[j]$ via the functional Legendre transform approach, in analogy to the Lieb energy functional of nonrelativistic DFT [11].

Relativistic Kohn–Sham Equations

The basis for the relativistic KS scheme is the assumption that there exists a relativistic system of noninteracting particles – the RKS system – with exactly the same ground state four-current density j^μ as that of the interacting system one is actually interested in. The particles of the RKS system experience a multiplicative effective four-potential $v_{\rm s}^\mu(\boldsymbol{r})$, including the radiative effects induced by this potential: the RKS system is necessarily governed by a QED-Hamiltonian, since it has to cover the noninteracting limit of the full Hamiltonian of bound state QED.

In practice, however, mapping the interacting QED system onto a noninteracting one is neither feasible nor of interest. On the one hand, the evaluation of the full four-current density and energy of such an RKS system would require summation over all negative and positive energy solutions of the associated Dirac-type RKS equations as well as appropriate renormalization in each step of the iterative RKS procedure (since the RKS vacuum depends on $v_{\rm s}^\mu$ and thus changes during the iteration process). This self-consistent process including non-perturbative renormalization is extremely difficult to implement. On the other hand, DFT typically aims at an efficient but only moderately accurate description of the electronic structure of molecules or solids, just capturing the essential bonding and excitation mechanisms. In this context, radiative corrections are of limited interest, since they primarily affect the innermost electrons (for a perturbative evaluation of radiative corrections on the basis of RDFT spinors, see, e.g., [12, 13]). Electronic structure calculations are therefore usually relying on the *no-pair* approximation in which all effects resulting from the creation of virtual particle–antiparticle pairs are neglected (for a more detailed discussion see [1]). This neglect automatically implies that charge conservation, Eq. (10), reduces to the more familiar particle number conservation.

Within this approximation, the ground state four-current density j^μ of the RKS system and, hence, by assumption, of the interacting system can be written as

$$j^\mu(\boldsymbol{r}) = \sum_k \Theta_k \phi_k^\dagger(\boldsymbol{r}) \alpha^\mu \phi_k(\boldsymbol{r}) \tag{11}$$

$$\Theta_k = \begin{cases} 0 \text{ for } \epsilon_k \leq -2mc^2 \\ 1 \text{ for } -2mc^2 < \epsilon_k \leq \epsilon_{\rm F} \\ 0 \text{ for } \epsilon_{\rm F} < \epsilon_k \end{cases} . \tag{12}$$

Here ϕ_k represents the single-particle four spinors of the RKS system, ϵ_k the corresponding eigenenergies. The *no-pair* approximation is reflected by the occupation numbers Θ_k: all RKS states below $-2mc^2$ are suppressed (with the understanding that ϵ_k does not contain the rest mass of the RKS particles). The Fermi level $\epsilon_{\rm F}$ defines the highest occupied RKS state.

The existence theorem of RDFT is equally valid for noninteracting systems. Consequently, the total energy of the RKS system is a functional of its ground state

four-current density (11). The same then applies to the kinetic energy T_s of the RKS system, which, in the *no-pair* approximation, is given by

$$T_s = \sum_k \Theta_k \int d^3r\, \phi_k^\dagger(\boldsymbol{r})\big[-i\hbar c\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + (\beta-1)mc^2\big]\phi_k(\boldsymbol{r}) \tag{13}$$

(since the total energy of the RKS system is $T_s + \int d^3r\, j_\mu v_s^\mu$). Using T_s, one can write the ground state energy functional of the interacting system as

$$E = T_s + E_{\text{ext}} + E_{\text{H}} + E_{\text{xc}}. \tag{14}$$

Here E_{ext} denotes the interaction of the electrons with the external sources,

$$E_{\text{ext}}[j] = \int d^3r\, j_\mu(\boldsymbol{r})\, V^\mu(\boldsymbol{r}). \tag{15}$$

E_{H} is the Hartree energy, which, in the context of RDFT, is usually defined to include the transverse (retarded Breit) interaction (i.e., to include all direct matrix elements of the electron–electron interaction of the order e^2),

$$E_{\text{H}}[j] = \frac{e^2}{2}\int d^3x \int d^3y\, \frac{j_\mu(\boldsymbol{x})\, j^\mu(\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|} \;=\; E_{\text{H}}^{\text{C}} + E_{\text{H}}^{\text{T}} \tag{16}$$

$$E_{\text{H}}^{\text{C}}[j^0] = \frac{e^2}{2}\int d^3r \int d^3r'\, \frac{n(\boldsymbol{r})\, n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} \tag{17}$$

$$E_{\text{H}}^{\text{T}}[\boldsymbol{j}] = -\frac{e^2}{2c^2}\int d^3r \int d^3r'\, \frac{\boldsymbol{j}(\boldsymbol{r})\cdot\boldsymbol{j}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}, \tag{18}$$

where the four-current density (11) has been split into the charge density $j^0 \equiv n$ and the spatial components $\boldsymbol{j}$ defined as

$$j^\mu(\boldsymbol{r}) \;=\; \big(n(\boldsymbol{r}), \boldsymbol{j}(\boldsymbol{r})/c\big). \tag{19}$$

The exchange–correlation (xc) energy E_{xc} comes up for all remaining energy contributions, with the exception of the radiative corrections, which have to be dropped from (14) for consistency with (11)–(13). Since all other quantities in (14) are density functionals, so is the xc-energy, $E_{\text{xc}}[j]$.

The minimization of the total energy functional (14) then leads to the RKS equations,

$$\big\{-i\hbar c\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + (\beta-1)mc^2 + \alpha_\mu v_s^\mu(\boldsymbol{r})\big\}\,\phi_k(\boldsymbol{r}) = \epsilon_k\phi_k(\boldsymbol{r}). \tag{20}$$

In addition to the external potential V^μ, the potential of the RKS system v_s^μ includes the functional derivatives of E_{H}, the Hartree potential v_{H}^μ, and of E_{xc}, the xc-potential v_{xc}^μ,

$$v_{\rm s}^{\mu}(\boldsymbol{r}) = V^{\mu}(\boldsymbol{r}) + v_{\rm H}^{\mu}(\boldsymbol{r}) + v_{\rm xc}^{\mu}(\boldsymbol{r}) \tag{21}$$

$$v_{\rm H}^{\mu}(\boldsymbol{r}) = e^2 \int d^3r' \, \frac{j^{\mu}(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \tag{22}$$

$$v_{\rm xc}^{\mu}(\boldsymbol{r}) = \frac{\delta E_{\rm xc}[j]}{\delta j_{\mu}(\boldsymbol{r})}. \tag{23}$$

As usual, self-consistent solution of Eqs. (11), (12), and (20)–(23) is required.

Relativistic Spin-Density Functional Theory

Several variants of RDFT have been introduced, primarily with the aim to avoid the gauge arbitrariness. This problem ceases to exist as soon as there is no external magnetic field, so that the electrons only experience a scalar potential, $V^{\mu} = (v_{\rm ext}, \mathbf{0})$. In this case, there is a one-to-one mapping between $v_{\rm ext}$ (up to global constants), the ground state, and the ground state density [4]. The ground state can thus be interpreted as a functional of the density only, $|\Psi_0[n]\rangle$. The same then applies to the spatial components of the current density $\boldsymbol{j}$,

$$\boldsymbol{j}[n] = \langle \Psi_0[n] | \hat{\boldsymbol{j}} | \Psi_0[n] \rangle \ .$$

Consequently, the minimization of the energy (14) with respect to n leads to

$$\left\{ -i\hbar c \boldsymbol{\alpha} \cdot \boldsymbol{\nabla} + (\beta - 1)mc^2 + v_{\rm s}(\boldsymbol{r}) \right\} \phi_k(\boldsymbol{r}) \ = \ \epsilon_k \phi_k(\boldsymbol{r}), \tag{24}$$

with the density n given by the time-like component of (11) and

$$\begin{aligned} v_{\rm s}(\boldsymbol{r}) = v_{\rm ext}(\boldsymbol{r}) + v_{\rm H}(\boldsymbol{r}) + \frac{\delta E_{\rm xc}[n, \boldsymbol{j}]}{\delta n(\boldsymbol{r})} \\ + \int d^3r' \left[\frac{\delta E_{\rm xc}[n, \boldsymbol{j}]}{\delta \boldsymbol{j}(\boldsymbol{r}')} + \frac{\delta E_{\rm H}^{\rm T}[\boldsymbol{j}]}{\delta \boldsymbol{j}(\boldsymbol{r}')} \right] \cdot \frac{\delta \boldsymbol{j}[n](\boldsymbol{r}')}{\delta n(\boldsymbol{r})}. \end{aligned} \tag{25}$$

Since $\boldsymbol{j}[n]$ is not known, however, one usually neglects the $\boldsymbol{j}$-dependence of $E_{\rm H}$ and $E_{\rm xc}$ in (25). Corresponding results were published long before this form of RDFT was formally established (see, e.g., [14]).

In the nonrelativistic limit, the full coupling of the magnetic field in the external four potential

$$V^{\mu} = (v_{\rm ext}, -e\boldsymbol{A}_{\rm ext}) \qquad\qquad (\text{with } \boldsymbol{B}_{\rm ext} = \boldsymbol{\nabla} \times \boldsymbol{A}_{\rm ext})$$

to the electrons via $\hat{\boldsymbol{j}} \cdot \boldsymbol{A}_{\rm ext}$ reduces to a coupling of the magnetization density $\boldsymbol{m}$ to the magnetic field, $\hat{\boldsymbol{m}} \cdot \boldsymbol{B}_{\rm ext}$. This transition is most easily established by the Gordon

decomposition, in which the total $\boldsymbol{j}$ is decomposed into an orbital current and the curl of $\boldsymbol{m}$. If one combines the standard $\hat{n}v_{\text{ext}}$-coupling with the $\hat{\boldsymbol{m}} \cdot \boldsymbol{B}_{\text{ext}}$-coupling (rather than the $\hat{\boldsymbol{j}} \cdot \boldsymbol{A}_{\text{ext}}$-coupling) and ignores the issue of renormalization, one can establish [4] a unique correspondence between the ground state $|\Psi_0\rangle$ and the ground state charge and magnetization densities $n, \boldsymbol{m}$, $|\Psi_0[n, \boldsymbol{m}]\rangle$. For $\boldsymbol{B}_{\text{ext}} \neq \boldsymbol{0}$, the resulting relativistic "spin" density functional formalism (RSDFT) represents an approximation to full RDFT, which neglects the coupling between $\boldsymbol{A}_{\text{ext}}$ and the orbital current. For $\boldsymbol{B}_{\text{ext}} = \boldsymbol{0}$, however, RSDFT simply covers a wider class of Hamiltonians than physically required (i.e., all Hamiltonians including an $\hat{\boldsymbol{m}} \cdot \boldsymbol{B}_{\text{ext}}$-coupling), so that the RSDFT formalism becomes exact.

Together with the ground state, the current density is a functional of n and $\boldsymbol{m}$,

$$\boldsymbol{j}[n, \boldsymbol{m}] = \langle \Psi_0[n, \boldsymbol{m}] | \hat{\boldsymbol{j}} | \Psi_0[n, \boldsymbol{m}] \rangle \ ,$$

and the same applies to the ground state energy and its components,

$$E_{\text{H}}^{\text{T}}[\boldsymbol{j}] = E_{\text{H}}^{\text{T}}[\boldsymbol{j}[n, \boldsymbol{m}]] \qquad E_{\text{xc}}[n, \boldsymbol{j}] = E_{\text{xc}}[n, \boldsymbol{j}[n, \boldsymbol{m}]] \equiv E_{\text{xc}}[n, \boldsymbol{m}] \ .$$

Assuming the noninteracting RKS system to reproduce the ground state densities n and $\boldsymbol{m}$ of the interacting system,

$$n(\boldsymbol{r}) = \sum_k \Theta_k \phi_k^\dagger(\boldsymbol{r}) \phi_k(\boldsymbol{r}) \tag{26}$$

$$\boldsymbol{m}(\boldsymbol{r}) = \mu_B \sum_k \Theta_k \phi_k^\dagger(\boldsymbol{r}) \beta \boldsymbol{\Sigma} \phi_k(\boldsymbol{r}) \tag{27}$$

$$\text{with} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\sigma} \end{pmatrix} \qquad \text{and} \qquad \mu_B = \frac{e\hbar}{2mc}, \tag{28}$$

the minimum principle for the total energy then yields the equations [4, 5, 15, 16]

$$\left\{ -i\hbar c \boldsymbol{\alpha} \cdot \boldsymbol{\nabla} + (\beta - 1)mc^2 + v_{\text{s}} + \mu_B \beta \boldsymbol{\Sigma} \cdot \boldsymbol{B}_{\text{s}} \right\} \phi_k = \epsilon_k \phi_k \tag{29}$$

$$v_{\text{s}}(\boldsymbol{r}) = v_{\text{ext}}(\boldsymbol{r}) + v_{\text{H}}(\boldsymbol{r}) + \frac{\delta E_{\text{xc}}[n, \boldsymbol{m}]}{\delta n(\boldsymbol{r})} + \frac{\delta E_{\text{H}}^{\text{T}}[\boldsymbol{j}[n, \boldsymbol{m}]]}{\delta n(\boldsymbol{r})} \tag{30}$$

$$\boldsymbol{B}_{\text{s}}(\boldsymbol{r}) = \boldsymbol{B}_{\text{ext}}(\boldsymbol{r}) + \frac{\delta E_{\text{xc}}[n, \boldsymbol{m}]}{\delta \boldsymbol{m}(\boldsymbol{r})} + \frac{\delta E_{\text{H}}^{\text{T}}[\boldsymbol{j}[n, \boldsymbol{m}]]}{\delta \boldsymbol{m}(\boldsymbol{r})}. \tag{31}$$

As in Eq. (25), E_{H}^{T} is usually neglected in (30) and (31), relying on its proportionality to $1/c^2$. Equations (26)–(31) have, for instance, been used to study open-shell atoms and molecules as well as the magnetic anisotropy of solids [9, 17–20].

The direct counterpart of conventional nonrelativistic spin-density functional theory is obtained from the general RSDFT formalism, if the spatial variation of

the orientation of $\boldsymbol{m}$ is ignored (the legitimacy of this step in the case of open-shell atoms and molecules is discussed in [9,18,19]). As soon as the coupling between the electrons and the magnetic field is restricted to $\hat{m}_z B_{\mathrm{ext},z}$, the ground state is uniquely determined by n and m_z only or, alternatively, by the generalized spin densities $n_\pm$,

$$n_\pm(\boldsymbol{r}) = \frac{1}{2}\left[n(\boldsymbol{r}) \pm \frac{1}{\mu_B} m_z(\boldsymbol{r})\right]. \tag{32}$$

The corresponding RKS equations are given by

$$\left\{-i\hbar c\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + (\beta-1)mc^2 + \sum_{\sigma=\pm} P_\sigma v_{\mathrm{s},\sigma}\right\}\phi_k = \epsilon_k \phi_k \tag{33}$$

$$n_\pm(\boldsymbol{r}) = \sum_k \Theta_k \phi_k^\dagger(\boldsymbol{r}) P_\pm \phi_k(\boldsymbol{r}) \tag{34}$$

$$v_{\mathrm{s},\sigma}(\boldsymbol{r}) = v_{\mathrm{ext}}(\boldsymbol{r}) + v_{\mathrm{H}}(\boldsymbol{r}) + v_{\mathrm{xc,ff}}(\boldsymbol{r}) \tag{35}$$

$$v_{\mathrm{xc,ff}}(\boldsymbol{r}) = \frac{\delta E_{\mathrm{xc}}[n_+,n_-]}{\delta n_\sigma(\boldsymbol{r})} + \frac{\delta E_{\mathrm{H}}^{\mathrm{T}}[\boldsymbol{j}\,[n_+,n_-]]}{\delta n_\sigma(\boldsymbol{r})} \tag{36}$$

$$P_+ = \begin{pmatrix} 1&0&0&0\\ 0&0&0&0\\ 0&0&0&0\\ 0&0&0&1 \end{pmatrix}, \qquad P_- = \begin{pmatrix} 0&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&0 \end{pmatrix}. \tag{37}$$

Here $B_{\mathrm{ext},z}$ has already been set to zero. Equations (33)–(37) represent the standard RDFT approach to magnetic systems.

Relativistic Exchange–Correlation Functional

The close relationship of RSDFT and in particular of Eqs. (33)–(37) with nonrelativistic spin-density functional theory has triggered the use of nonrelativistic spin-density functionals in R(S)DFT. This approach is easily implemented in Eq. (36): one simply inserts a nonrelativistic functional $E_{\mathrm{xc}}[n_\uparrow, n_\downarrow]$ with the spin densities $n_\uparrow, n_\downarrow$ replaced by n_+, n_-. This approximation neglects all relativistic corrections to the functional dependence of E_{xc} on n_+, n_- but retains the relativistic corrections inherent in the densities themselves. The obvious question then is: are the relativistic corrections to the functional dependence of any relevance? To answer this, one has to analyze the available relativistic forms for E_{xc}.

Exact representations of the relativistic E_{xc} can be derived both via the relativistic variant of the adiabatic connection [21] and by means of KS-based many-body theory [22]. The adiabatic connection of RDFT is based on the assumption that for any scaled electron–electron coupling strength λe^2, with $0 \le \lambda \le 1$, one can find an external four potential $u_\lambda^\mu(\boldsymbol{x})$, so that the ground state four-current

density resulting from the corresponding λ-dependent QED Hamiltonian is identical with that obtained for the true coupling strength, $\lambda = 1$, and true potential $V^{\mu}(\boldsymbol{x})$. Applying the coupling constant integration technique to the λ-dependent components of the Hamiltonian, one finds a representation of E_{xc} in terms of the current–current response function of the λ-dependent system. The adiabatic connection of RDFT has primarily been used to derive the relativistic LDA (see section "Relativistic Local Density Approximation").

In KS-based many-body theory, the total QED Hamiltonian is split into the noninteracting RKS Hamiltonian (assuming the RKS four-potential to be known) and a remainder $\hat{W}$, for which again the coupling constant integration technique can be utilized. The resulting expression for E_{xc} is a power series in $\hat{W}$, from which explicit approximations can be obtained with the standard (diagrammatic) methods of QED.

Orbital-Dependent Exchange–Correlation Functionals

KS-based many-body theory is particularly useful for the derivation of orbital-dependent xc-functionals, such as the exact exchange of RDFT. Using Feynman gauge and the *no-pair* approximation, one obtains [22]

$$E_{\mathrm{x}} = -\frac{e^2}{2}\sum_{k,l}\Theta_k\Theta_l\int d^3r\int d^3r'\,\frac{\cos(\omega_{kl}|\boldsymbol{r}-\boldsymbol{r}'|)}{|\boldsymbol{r}-\boldsymbol{r}'|} \times \phi_k^{\dagger}(\boldsymbol{r})\alpha_{\mu}\phi_l(\boldsymbol{r})\;\phi_l^{\dagger}(\boldsymbol{r}')\alpha^{\mu}\phi_k(\boldsymbol{r}'), \tag{38}$$

where $\omega_{kl} = |\epsilon_k - \epsilon_l|/(\hbar c)$. Consistent with the Hartree energy (16), the exchange functional (38) contains a transverse contribution, which may be extracted explicitly by subtraction of the standard Coulomb exchange,

$$E_{\mathrm{x}}^{\mathrm{C}} = -\frac{e^2}{2}\sum_{k,l}\Theta_k\Theta_l\int d^3r\int d^3r'\,\frac{\phi_k^{\dagger}(\boldsymbol{r})\phi_l(\boldsymbol{r})\;\phi_l^{\dagger}(\boldsymbol{r}')\phi_k(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} \tag{39}$$

$$E_{\mathrm{x}}^{\mathrm{T}} = E_{\mathrm{x}} - E_{\mathrm{x}}^{\mathrm{C}}. \tag{40}$$

The expressions (38)–(40) are functionals of j^{μ} in the same sense as T_{s} is a functional of j^{μ}: the RKS-spinors are unique functionals of j^{μ}, as the ground state Slater determinant of the RKS system is a unique functional of j^{μ} by virtue of the RDFT existence theorem for noninteracting particles.

It is worthwhile to emphasize that the expression (38) is gauge invariant. In general, gauge invariance requires the inclusion of the negative energy states in all intermediate sums over states, which show up in a perturbative treatment of the electron–electron interaction. Since the *no-pair* approximation systematically neglects the negative energy states, a gauge dependence is usually introduced.

As an exception from this rule, the *no-pair* exchange (38) turns out to be gauge invariant due to the multiplicative nature of the RKS potential [22], justifying the use of the simplest gauge, the Feynman gauge, in (38). The gauge invariance of (38) also emphasizes the difference between the exchange of RDFT and relativistic Hartree–Fock (RHF) exchange: unlike (38), the transverse RHF exchange is gauge-dependent, since the RHF spinors experience a nonlocal potential.

Insertion of the four-current density (11) into (16) reveals that the Hartree energy contains some self-interaction:

$$E_{\mathrm{H}} = \frac{e^2}{2} \sum_{kl} \Theta_k \Theta_l \int d^3r \int d^3r' \, \frac{\phi_k^\dagger(\boldsymbol{r}) \alpha_\mu \phi_k(\boldsymbol{r}) \phi_l^\dagger(\boldsymbol{r}') \alpha^\mu \phi_l(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}.$$

The self-interaction terms with $k = l$ are, however, exactly cancelled by the exact exchange (38). In fact, this is not only true for the Coulomb component of both energies but also for the transverse interaction.

Orbital-dependent functionals can also be derived for the relativistic correlation functional

$$E_{\mathrm{c}} = E_{\mathrm{xc}} - E_{\mathrm{x}}. \tag{41}$$

The resulting expressions are completely analogous to the corresponding nonrelativistic functionals, if the transverse interaction is neglected (which seems to be well-justified in the case of correlation). For instance, second-order perturbation theory with respect to the RKS Hamiltonian [22] yields the relativistic counterpart of the second-order Görling–Levy functional [23],

$$E_{\mathrm{c}}^{(2)} = E_{\mathrm{c}}^{\mathrm{MP2}} + E_{\mathrm{c}}^{\Delta\mathrm{HF}} \tag{42}$$

$$E_{\mathrm{c}}^{\mathrm{MP2}} = \frac{1}{2} \sum_{ijkl;\epsilon_F < \epsilon_k, \epsilon_l} \Theta_i \Theta_j \, \frac{(ij||kl)\,[(kl||ij) - (kl||ji)]}{\epsilon_i + \epsilon_j - \epsilon_k - \epsilon_l} \tag{43}$$

$$E_{\mathrm{c}}^{\Delta\mathrm{HF}} = \sum_{il;\epsilon_F < \epsilon_l} \frac{\Theta_i}{\epsilon_i - \epsilon_l} \left| \langle i | \alpha_\mu v_{\mathrm{x}}^\mu | l \rangle + e^2 \sum_j \Theta_j (ij||jl) \right|^2, \tag{44}$$

with the relativistic Slater integrals,

$$(ij||kl) = \int d^3r_1 \int d^3r_2 \, \frac{\phi_i^\dagger(\boldsymbol{r}_1) \phi_k(\boldsymbol{r}_1) \phi_j^\dagger(\boldsymbol{r}_2) \phi_l(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}, \tag{45}$$

and the matrix elements of the RKS exchange potential,

$$\langle i | \alpha_\mu v_{\mathrm{x}}^\mu | l \rangle = \int d^3r \, \phi_i^\dagger(\boldsymbol{r}) \alpha_\mu \phi_l(\boldsymbol{r}) v_{\mathrm{x}}^\mu(\boldsymbol{r}), \tag{46}$$

depending on the RKS four spinors. Similarly, resummation of all ring diagrams of RKS-based perturbation theory leads to the relativistic extension [24, 25] of the random phase approximation (RPA) of DFT [26–29].

Relativistic Optimized Potential Method

Orbital-dependent xc-functionals are often applied a posteriori, utilizing the KS orbitals resulting from LDA or GGA calculations. However, at least in the case of the exact exchange, a self-consistent application is of obvious interest, in order to obtain an RKS potential which is self-interaction-free.

The xc-potential $v_{\rm xc}^{\mu} = \delta E_{\rm xc}/\delta j_{\mu}$ for functionals of the type (38)–(44) has to be evaluated by the relativistic extension of the optimized potential method (ROPM) [30, 31]. The fundamental integral equation of the ROPM is most easily derived by direct functional differentiation of the orbital-dependent expression, if the chain rule is used to replace the derivative with respect to j^{μ} by ones with respect to the orbitals and eigenvalues [22],

$$\frac{\delta E_{\rm xc}}{\delta j^{\mu}(\boldsymbol{r})} = \int d^3r' \, \frac{\delta v_{\rm s}^{\rho}(\boldsymbol{r}')}{\delta j^{\mu}(\boldsymbol{r})} \sum_k \left\{ \int d^3r'' \left[\frac{\delta \phi_k^{\dagger}(\boldsymbol{r}'')}{\delta v_{\rm s}^{\rho}(\boldsymbol{r}')} \frac{\delta E_{\rm xc}}{\delta \phi_k^{\dagger}(\boldsymbol{r}'')} + c.c. \right] + \frac{\delta \epsilon_k}{\delta v_{\rm s}^{\rho}(\boldsymbol{r}')} \frac{\partial E_{\rm xc}}{\partial \epsilon_k} \right\}. \tag{47}$$

By virtue of the chain rule, the summation over k on the right-hand side of (47) includes all negative (and positive) energy states. As soon as the *no-pair* approximation is applied to $E_{\rm xc}$, all derivatives of $E_{\rm xc}$ with respect to negative energy states vanish, so that the summation over k reduces to states with $\epsilon_k > -2mc^2$. This constraint will, however, not be explicitly noted in the following. The derivatives on the right-hand side of (47) can be evaluated from the RKS equations. Introducing a small perturbation $\delta v_{\rm s}^{\mu}$ into (20) leads to a modification of the spinors and eigenvalues by $\delta\phi_k$ and $\delta\epsilon_k$, respectively,

$$\begin{aligned} &\left\{-i\hbar c\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + (\beta-1)mc^2 + \alpha_{\mu}\left[v_{\rm s}^{\mu}(\boldsymbol{r}) + \delta v_{\rm s}^{\mu}(\boldsymbol{r})\right]\right\}\left[\phi_k(\boldsymbol{r}) + \delta\phi_k(\boldsymbol{r})\right] \\ &\quad = \left[\epsilon_k + \delta\epsilon_k\right]\left[\phi_k(\boldsymbol{r}) + \delta\phi_k(\boldsymbol{r})\right]. \end{aligned}$$

For the derivatives in (47), only the terms linear in $\delta v_{\rm s}^{\mu}$ are relevant,

$$\begin{aligned} &\left\{-i\hbar c\boldsymbol{\alpha}\cdot\boldsymbol{\nabla} + (\beta-1)mc^2 + \alpha_{\mu}v_{\rm s}^{\mu}(\boldsymbol{r}) - \epsilon_k\right\}\delta\phi_k(\boldsymbol{r}) \\ &\quad = \delta\epsilon_k\phi_k(\boldsymbol{r}) - \alpha_{\mu}\delta v_{\rm s}^{\mu}(\boldsymbol{r})\phi_k(\boldsymbol{r}). \end{aligned} \tag{48}$$

Multiplication of this equation by $\phi_k^{\dagger}(\boldsymbol{r})$ and subsequent integration over $\boldsymbol{r}$ yields the shift of the eigenvalue,

$$\delta\epsilon_k = \int d^3r\, \phi_k^\dagger(\boldsymbol{r})\alpha_\mu \delta v_{\rm s}^\mu(\boldsymbol{r})\phi_k(\boldsymbol{r}) \qquad \Longrightarrow \qquad \frac{\delta\epsilon_k}{\delta v_{\rm s}^\mu(\boldsymbol{r})} = \phi_k^\dagger(\boldsymbol{r})\alpha_\mu\phi_k(\boldsymbol{r})\ .$$

Using this result, the differential equation for $\delta\phi_k$, Eq. (48), can be formally solved by means of the Greens function

$$G_k(\boldsymbol{r},\boldsymbol{r}') = \sum_{l\neq k} \frac{\phi_l(\boldsymbol{r})\,\phi_l^\dagger(\boldsymbol{r}')}{\epsilon_l - \epsilon_k} \tag{49}$$

$$\Longrightarrow \qquad \delta\phi_k(\boldsymbol{r}) = -\int d^3r'\, G_k(\boldsymbol{r},\boldsymbol{r}')\alpha_\mu \delta v_{\rm s}^\mu(\boldsymbol{r}')\phi_k(\boldsymbol{r}')$$

$$\Longrightarrow \qquad \frac{\delta\phi_k(\boldsymbol{r})}{\delta v_{\rm s}^\mu(\boldsymbol{r}')} = -G_k(\boldsymbol{r},\boldsymbol{r}')\alpha_\mu\phi_k(\boldsymbol{r}').$$

Moreover, $\delta v_{\rm s}^\rho(\boldsymbol{r}')/\delta j^\mu(\boldsymbol{r})$ is the inverse of the static current–current response function, which can be evaluated by differentiation of the four-current density (11),

$$\chi_{\rm s}^{\mu\nu}(\boldsymbol{r},\boldsymbol{r}') = \frac{\delta j^\mu(\boldsymbol{r})}{\delta v_{{\rm s},\nu}(\boldsymbol{r}')} = -\sum_k \Theta_k\, \phi_k^\dagger(\boldsymbol{r})\,\alpha^\mu G_k(\boldsymbol{r},\boldsymbol{r}')\,\alpha^\nu\phi_k(\boldsymbol{r}') + c.c.. \tag{50}$$

Multiplication of Eq. (47) by $\chi_{\rm s}^{\mu\nu}$ and subsequent integration leads to a set of four coupled integral equations,

$$\begin{aligned}
&\int d^3r'\, \chi_{\rm s}^{\mu\nu}(\boldsymbol{r},\boldsymbol{r}')\, v_{{\rm xc},\nu}(\boldsymbol{r}') \\
&= -\sum_k \int d^3r' \left[\phi_k^\dagger(\boldsymbol{r})\alpha^\mu G_k(\boldsymbol{r},\boldsymbol{r}')\frac{\delta E_{\rm xc}}{\delta\phi_k^\dagger(\boldsymbol{r}')} + c.c.\right] \\
&\quad + \sum_k \phi_k^\dagger(\boldsymbol{r})\alpha^\mu\phi_k(\boldsymbol{r})\frac{\partial E_{\rm xc}}{\partial\epsilon_k}.
\end{aligned} \tag{51}$$

It is worthwhile to emphasize that (by construction) orbital-dependent expressions are functionals of the complete four-current density (and not just functionals of n). This is reflected by the fact that the spatial components $v_{\rm xc}^i$ of the solution of (51) do not vanish in general.

Equation (51) has to be solved in each cycle of the self-consistent RKS procedure. In this process, one has to fix the gauge of $v_{\rm xc}^\mu$. In the case of $v_{\rm xc}^0$, this amounts to a normalization, since Eq. (51) determines $v_{\rm xc}^0$ only up to a global constant: the response function satisfies the identity

$$\int d^3r'\, \chi_{\rm s}^{\mu 0}(\boldsymbol{r},\boldsymbol{r}') = \int d^3r\, \chi_{\rm s}^{0\mu}(\boldsymbol{r},\boldsymbol{r}') = 0, \tag{52}$$

as can be verified by integration over (50) and use of the orthonormality of the ϕ_k. The normalization of $v_{\rm xc}^0$ is usually defined by the requirement $v_{\rm xc}^0(|\boldsymbol{r}| \to \infty) = 0$ for finite systems, while the average of $v_{\rm xc}^0$ in the unit cell is set to zero in the case of periodic systems.

The gauge freedom of the spatial components of $v_{\rm xc}^\mu$ is intrinsically related to the transversality of $\chi_{\rm s}^{\mu\nu}$. This property is easily verified as long as G_k, Eq. (49), is accounted for exactly. Differentiation first yields

$$\begin{aligned}\frac{\partial}{\partial r^j}\chi_{\rm s}^{j\nu}(\boldsymbol{r},\boldsymbol{r}') &= -\sum_k \Theta_k \sum_{l\neq k} \phi_k^\dagger(\boldsymbol{r})\boldsymbol{\alpha}\cdot\Big(\overleftarrow{\boldsymbol{\nabla}}+\overrightarrow{\boldsymbol{\nabla}}\Big)\frac{\phi_l(\boldsymbol{r})\,\phi_l^\dagger(\boldsymbol{r}')}{\epsilon_l-\epsilon_k}\alpha^\nu\phi_k(\boldsymbol{r}') + c.c.\\ &= \frac{-i}{\hbar c}\sum_k \Theta_k \sum_{l\neq k} \phi_k^\dagger(\boldsymbol{r})\phi_l(\boldsymbol{r})\,\phi_l^\dagger(\boldsymbol{r}')\alpha^\nu\phi_k(\boldsymbol{r}') + c.c.,\end{aligned} \tag{53}$$

where the RKS equations have been used to arrive at the second line,

$$\phi_k^\dagger(\boldsymbol{r})(\epsilon_k-\epsilon_l)\phi_l(\boldsymbol{r}) = i\hbar c\,\phi_k^\dagger(\boldsymbol{r})\boldsymbol{\alpha}\cdot\Big(\overleftarrow{\boldsymbol{\nabla}}+\overrightarrow{\boldsymbol{\nabla}}\Big)\phi_l(\boldsymbol{r}) \tag{54}$$

($\overleftarrow{\boldsymbol{\nabla}}$ denotes differentiation of the functions to the left of the $\boldsymbol{\nabla}$-operator). At this point, the completeness of the ϕ_l, $\sum_l \phi_l(\boldsymbol{r})\phi_l^\dagger(\boldsymbol{r}') = \delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}')$, can be utilized, provided the summation over l in (53) includes all negative energy states,

$$\begin{aligned}\frac{\partial}{\partial r^j}\chi_{\rm s}^{j\nu}(\boldsymbol{r},\boldsymbol{r}') &= \frac{-i}{\hbar c}\sum_k \Theta_k\,\phi_k^\dagger(\boldsymbol{r})\delta^{(3)}(\boldsymbol{r}-\boldsymbol{r}')\alpha^\nu\phi_k(\boldsymbol{r}') + c.c.\\ &\quad + \frac{i}{\hbar c}\sum_k \Theta_k\,\phi_k^\dagger(\boldsymbol{r})\,\phi_k(\boldsymbol{r})\,\phi_k^\dagger(\boldsymbol{r}')\alpha^\nu\phi_k(\boldsymbol{r}') + c.c.\\ &= 0.\end{aligned} \tag{55}$$

Note that this result holds for all possible choices for the occupation Θ_k, including the *no-pair* form (12). As a consequence, Eq. (51) determines the spatial components of $v_{\rm xc}^\mu$ only up to a gauge function $\boldsymbol{\nabla}\lambda$,

$$\int d^3r'\,\chi_{\rm s}^{\mu j}(\boldsymbol{r},\boldsymbol{r}')\,\frac{\partial}{\partial r'^j}\lambda(\boldsymbol{r}') = -\int d^3r'\left[\frac{\partial}{\partial r'^j}\chi_{\rm s}^{\mu j}(\boldsymbol{r},\boldsymbol{r}')\right]\lambda(\boldsymbol{r}') = 0\,.$$

If, however, the *no-pair* approximation is also applied to G_k, the l-summation in (49) is restricted to states with $\epsilon_k > -2mc^2$. As a result, the completeness relation is no longer available to ensure the transversality of $\chi_{\rm s}^{\mu\nu}$,

$$\frac{\partial}{\partial r^j}\chi_{\rm s}^{j\nu}(\boldsymbol{r},\boldsymbol{r}') \neq 0, \tag{56}$$

and the gauge freedom of $\boldsymbol{v}_{\rm xc}$ is lost.

The exact cancellation of the self-interaction by $E_{\rm x}$ is also visible in $v_{\rm xc}^{\mu}$. A detailed analysis of the integral equation (51) for finite systems allows one to derive the asymptotic behavior of $v_{\rm x}^0$ [22],

$$v_{\rm x}^0(\boldsymbol{r}) \stackrel{|\boldsymbol{r}|\to\infty}{\longrightarrow} -\frac{e^2}{|\boldsymbol{r}|}. \tag{57}$$

In order to see the physics behind this result, it has to be combined with the corresponding asymptotic forms of the external potential, $V^0 \sim -Ze^2/|\boldsymbol{r}|$ (Z = total charge of all nuclei), and of the Hartree potential, $v_{\rm H}^0 \sim Ne^2/|\boldsymbol{r}|$. The sum of the three potentials decays as $-(Z-N+1)e^2/|\boldsymbol{r}|$, as required by electrostatics, if one electron far outside the molecule experiences the net attraction of the nuclei and the other $N-1$ electrons attached to them.

The derivation of Eq. (51) is based on the four-current RDFT formalism. An analogous set of integral equations can be derived for RSDFT [25, 32],

$$\begin{aligned}
&\int d^3r' \left\{\chi_{nn}(\boldsymbol{r},\boldsymbol{r}')v_{\rm xc}(\boldsymbol{r}') + \boldsymbol{\chi}_{nm}(\boldsymbol{r},\boldsymbol{r}')\cdot\boldsymbol{B}_{\rm xc}(\boldsymbol{r}')\right\} \\
&= -\sum_k \int d^3r' \left[\phi_k^\dagger(\boldsymbol{r})G_k(\boldsymbol{r},\boldsymbol{r}')\frac{\delta E_{\rm xc}}{\delta\phi_k^\dagger(\boldsymbol{r}')} + c.c.\right] + \sum_k |\phi_k(\boldsymbol{r})|^2\frac{\partial E_{\rm xc}}{\partial\epsilon_k} \qquad (58)\\
&\int d^3r' \left\{\boldsymbol{\chi}_{mn}(\boldsymbol{r},\boldsymbol{r}')v_{\rm xc}(\boldsymbol{r}') + \boldsymbol{\chi}_{mm}(\boldsymbol{r},\boldsymbol{r}')\cdot\boldsymbol{B}_{\rm xc}(\boldsymbol{r}')\right\} \\
&= -\mu_B\sum_k \int d^3r' \left[\phi_k^\dagger(\boldsymbol{r})\beta\boldsymbol{\Sigma}G_k(\boldsymbol{r},\boldsymbol{r}')\frac{\delta E_{\rm xc}}{\delta\phi_k^\dagger(\boldsymbol{r}')} + c.c.\right] \\
&\quad +\mu_B\sum_k \phi_k^\dagger(\boldsymbol{r})\beta\boldsymbol{\Sigma}\phi_k(\boldsymbol{r})\frac{\partial E_{\rm xc}}{\partial\epsilon_k} \quad . \qquad (59)
\end{aligned}$$

The kernels are once more the response functions of the RKS system,

$$\chi_{nn}(\boldsymbol{r},\boldsymbol{r}') = \frac{\delta n(\boldsymbol{r})}{\delta v_{\rm s}(\boldsymbol{r}')} = -\sum_k \Theta_k\phi_k^\dagger(\boldsymbol{r})G_k(\boldsymbol{r},\boldsymbol{r}')\phi_k(\boldsymbol{r}') + c.c. \tag{60}$$

$$\boldsymbol{\chi}_{mn}(\boldsymbol{r},\boldsymbol{r}') = \frac{\delta \boldsymbol{m}(\boldsymbol{r})}{\delta v_{\rm s}(\boldsymbol{r}')} = -\mu_B\sum_k \Theta_k\phi_k^\dagger(\boldsymbol{r})\beta\boldsymbol{\Sigma}G_k(\boldsymbol{r},\boldsymbol{r}')\phi_k(\boldsymbol{r}') + c.c. \tag{61}$$

$$\boldsymbol{\chi}_{nm}(\boldsymbol{r},\boldsymbol{r}') = \frac{\delta n(\boldsymbol{r})}{\delta \boldsymbol{B}_{\rm s}(\boldsymbol{r}')} = \boldsymbol{\chi}_{mn}^\dagger(\boldsymbol{r}',\boldsymbol{r}) \tag{62}$$

$$\boldsymbol{\chi}_{mm}(\boldsymbol{r},\boldsymbol{r}') = \frac{\delta \boldsymbol{m}(\boldsymbol{r})}{\delta \boldsymbol{B}_{\rm s}(\boldsymbol{r}')} = -\mu_B^2\sum_k \Theta_k\phi_k^\dagger(\boldsymbol{r})\beta\boldsymbol{\Sigma}G_k(\boldsymbol{r},\boldsymbol{r}')\beta\boldsymbol{\Sigma}\phi_k(\boldsymbol{r}') + c.c., \tag{63}$$

and G_k is again given by (49).

In the case of collinear magnetization density, $\boldsymbol{m} = (0,0,m_z)$, these equations can be rewritten in terms of the generalized spin densities (34),

$$\sum_{\sigma'} \int d^3r' \, \chi_{\sigma\sigma'}(\boldsymbol{r},\boldsymbol{r}') v_{\mathrm{xc},\sigma'}(\boldsymbol{r}')$$

$$= -\sum_k \int d^3r' \left[\phi_k^\dagger(\boldsymbol{r}) P_\sigma G_k(\boldsymbol{r},\boldsymbol{r}') \frac{\delta E_{\mathrm{xc}}}{\delta \phi_k^\dagger(\boldsymbol{r}')} + c.c. \right]$$

$$+ \sum_k \phi_k^\dagger(\boldsymbol{r}) P_\sigma \phi_k(\boldsymbol{r}) \frac{\partial E_{\mathrm{xc}}}{\partial \epsilon_k} \tag{64}$$

$$\chi_{\sigma\sigma'}(\boldsymbol{r},\boldsymbol{r}') = \frac{\delta n_\sigma(\boldsymbol{r})}{\delta v_{\mathrm{s},\sigma'}(\boldsymbol{r}')} = -\sum_k \Theta_k \phi_k^\dagger(\boldsymbol{r}) P_\sigma G_k(\boldsymbol{r},\boldsymbol{r}') P_{\sigma'} \phi_k(\boldsymbol{r}') + c.c., \tag{65}$$

with P_σ given by (37). Each of the spin-channels $v_{\mathrm{x},\sigma}$ of (64) satisfies (57) in the case of finite systems [25].

It remains to remark that the ROPM equation of the purely n-dependent RDFT formalism [33, 34] is given by the time-like component of Eq. (51).

Relativistic Corrections in E_{xc}: I. Transverse Exchange

The exact exchange allows an unambiguous assessment of the importance of relativistic corrections in the xc-functional and in particular of the transverse interaction: for this functional, one can directly compare the fully relativistic expression (38) with the Coulomb exchange (39) and the Coulomb–Breit approximation, which includes the transverse interaction only to leading order in $1/c^2$,

$$E_{\mathrm{x}}^{\mathrm{B}} = \frac{e^2}{4} \sum_{k,l} \Theta_k \Theta_l \int d^3r \int d^3r' \sum_{i,j=1}^{3} \frac{\phi_k^\dagger(\boldsymbol{r}) \alpha^i \phi_l(\boldsymbol{r}) \, \phi_l^\dagger(\boldsymbol{r}') \alpha^j \phi_k(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}$$

$$\times \left(\delta_{ij} + \frac{(r^i - r'^i)(r^j - r'^j)}{|\boldsymbol{r}-\boldsymbol{r}'|^2} \right) \tag{66}$$

(The Breit exchange can be derived from (40) by expansion in powers of $1/c$ and subsequent use of (54)).

Results obtained by solution of the ROPM equation (51) for these three exchange-only functionals (REXX) are given in Tables 1–4 as well as Fig. 1. Table 1 lists REXX ground state energies of closed-subshell atoms. On the one hand, a comparison of the Coulomb with the Coulomb–Breit energies shows the well-known magnitude of the Breit correction, ranging from marginal for light atoms to keV-size for the heaviest ones. On the other hand, the corrections resulting from the retardation effects by which the complete transverse interaction differs from the Breit interaction are almost two orders of magnitude smaller. As to be expected, the retarded exchange of photons effectively leads to a reduction of the interaction strength compared to the instantaneous Coulomb interaction: both

Table 1 Exchange-only ground state energies from finite differences REXX calculations for closed subshell atoms: Coulomb (C)- and Coulomb–Breit (C+B)-limit in comparison with complete transverse exchange (C+T) [22]. In all calculations, the nuclei were represented by uniformly charged spheres with nuclear radii given by $R_{\mathrm{nuc}} = (1.0793A^{1/3} + 0.73587)\,\mathrm{fm}$, A being the atomic mass (weighted by isotopic abundances) taken from Table III.7 of [35]. The speed of light was set to $c = 137.0359895$ a.u. (all energies in mHa)

	$-E^{\mathrm{C+T}}$	$E - E^{\mathrm{C+T}}$	
		C+B	C
He	2,861.8	0.0	−0.1
Ne	128,673.6	0.0	−16.7
Ar	528,546.1	0.0	−132.2
Zn	1,793,840.0	2.6	−758.4
Kr	2,787,429.4	7.2	−1,418.5
Cd	5,589,495.8	34.1	−3,803.1
Xe	7,441,173.0	63.9	−5,702.6
Yb	14,053,749.7	247.3	13,871.3
Hg	19,626,704.9	490.2	22,121.0
Rn	23,573,354.2	707.1	−28,615.2
No	36,687,172.7	1,633.8	−53,452.8

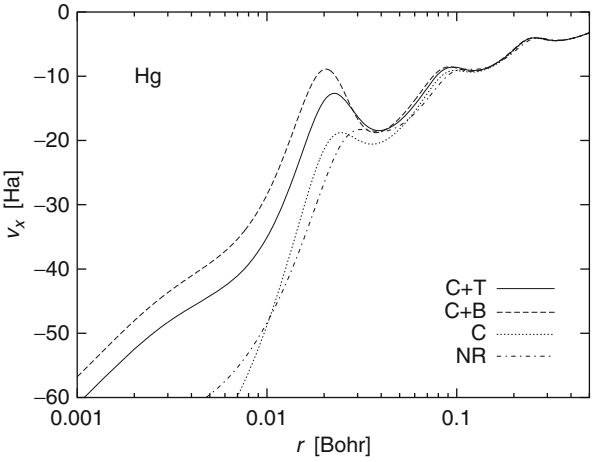


Fig. 1 REXX exchange potential for neutral Hg: self-consistent Coulomb (C), Coulomb–Breit (C+B), and fully transverse (C+T) results in comparison with nonrelativistic limit (NR)

the retardation of the Coulomb exchange term in (38) by the modulation factor $\cos(\omega_{kl}|\boldsymbol{r} - \boldsymbol{r}'|)$ and the current–current-coupling act repulsive, so that the Breit interaction is repulsive. However, the lowest order representation of both effects by expansion of $\cos(\omega_{kl}|\boldsymbol{r} - \boldsymbol{r}'|)$ in $1/c$ necessarily overestimates the true reduction, so that the transverse corrections beyond the Breit-limit are attractive.

As an example, the corresponding exchange potentials of the Hg atom are shown in Fig. 1, together with the nonrelativistic v_{x}. As is clear from the preceding discussion, the transverse exchange potential $v_{\mathrm{x}}^{\mathrm{T}}$ is predominantly repulsive. Even

Table 2 Exchange-only single particle energies ($-\epsilon_{nlj}$) for neutral Hg from self-consistent REXX calculations, using (i) the complete relativistic EXX potential (C+T), (ii) its Coulomb–Breit approximation (C+B), and (iii) its Coulomb (C) limit. Also given are RGGA results which have been obtained with the relativistic extension [36] of the Becke functional [37], either restricting to exchange-only (RB88) or including the Lee–Yang–Parr correlation GGA (RBLYP) [38] (all energies in mHa; nuclear charge distribution and c as in Table 1)

	REXX			RB88	RBLYP
Level	C+T	C+B	C	C+T	C+T
$1s1/2$	3,036,871	3,032,278	3,047,431	3,036,453	3,036,485
$2s1/2$	538,444	537,853	540,057	538,051	538,085
$2p1/2$	516,198	515,546	518,062	516,097	516,132
$2p3/2$	445,422	445,013	446,683	445,276	445,311
$3s1/2$	127,956	127,858	128,273	127,703	127,738
$3p1/2$	117,994	117,885	118,351	117,857	117,893
$3p3/2$	102,302	102,236	102,537	102,152	102,187
$3d3/2$	86,069	86,036	86,202	85,959	85,994
$3d5/2$	82,692	82,665	82,808	82,582	82,617
$4s1/2$	28,361	28,351	28,428	28,037	28,072
$4p1/2$	24,090	24,075	24,162	23,819	23,854
$4p3/2$	20,321	20,315	20,364	20,024	20,059
$4d3/2$	13,397	13,397	13,412	13,151	13,186
$4d5/2$	12,689	12,690	12,701	12,441	12,476
$4f5/2$	3,766	3,770	3,757	3,571	3,607
$4f7/2$	3,613	3,616	3,603	3,417	3,453
$5s1/2$	4,394	4,394	4,404	4,278	4,313
$5p1/2$	3,004	3,002	3,013	2,886	2,920
$5p3/2$	2,360	2,360	2,364	2,219	2,253
$5d3/2$	507	507	506	367	399
$5d5/2$	440	441	440	300	332
$6s1/2$	330	330	330	222	249

for the L-shell, this correction is of the order of 1 Ha ($\langle r(2s_{1/2})\rangle = 0.069$ Bohr). Unlike for the integrated energies, the local error of the Breit exchange potential is quite substantial.

The RKS eigenvalues of Hg obtained with these potentials are listed in Table 2. Both the K- and L-shell energies clearly reflect the shift of the total RKS potential by inclusion of $v_{\mathrm{x}}^{\mathrm{T}}$. Somewhat surprisingly, the $2p_{1/2}$-level percentage-wise experiences a slightly larger shift than the $1s_{1/2}$-state (this phenomenon is observed quite often for very heavy atoms). For all low-lying states (including the M-shell), $v_{\mathrm{x}}^{\mathrm{T}}$ dominates over the correlation potential. In spite of the orthogonality constraint, however, the $6s_{1/2}$-level remains almost unaffected by $v_{\mathrm{x}}^{\mathrm{T}}$: $v_{\mathrm{x}}^{\mathrm{T}}$ amounts to less than 0.3 % of the total RKS potential for the $1s$-shell, so that only a minor deformation of the $1s$-orbital is observed. As a result, the eigenvalue of the $6s_{1/2}$-level changes by only 0.15 %, which is irrelevant on the absolute scale.

Table 3 REXX ionization potentials of neutral atoms calculated from total energy differences between neutral and multiply ionized states: self-consistent inclusion of the transverse exchange (C+T) versus neglect of E_x^T (C). In the case of group IIA and IIB atoms, both s-electrons are removed, for the IVB atoms both $p_{1/2}$-electrons, for the noble gas atoms all $p_{3/2}$-electrons (all energies in mHa; point nuclei have been used)

Atom	C+T	C
Sr	558	558
Ba	501	501
Cd	838	838
Hg	942	941
Sn	744	743
Pb	769	767
Xe	3,785	3,784
Rn	3,406	3,407

Table 4 REXX electron removal energies of highly ionized Hg calculated from total energy differences: self-consistent inclusion of the transverse exchange (C+T) versus perturbative evaluation of E_x^T on basis of self-consistent calculations with E_x^C (C+pT) and complete neglect of E_x^T (C) (all energies in Ha; nuclear charge distribution and c as in Table 1)

ΔE	C+T	C+pT	C
$Hg^{70+} \longrightarrow Hg^{74+}$	2,810.227	2,810.219	2,813.784
$Hg^{74+} \longrightarrow Hg^{76+}$	1,642.950	1,642.937	1,646.533
$Hg^{76+} \longrightarrow Hg^{78+}$	1,718.586	1,718.582	1,720.295
$Hg^{78+} \longrightarrow Hg^{80+}$	6,994.905	6,994.891	7,002.469

It is worthwhile to point out that all the potentials in Fig. 1 asymptotically obey Eq. (57) (which is not visible in the figure due to the restricted r-range). As a result, the REXX eigenvalues for the most weakly bound $6s_{1/2}$-state are quite close to the REXX ionization energy for a $6s_{1/2}$-electron obtained by subtraction of the ground state energy of Hg^+ from that of neutral Hg (312 mHa, see below). This behavior represents a clear improvement over the potentials of the GGA functionals, which do not satisfy (57) and thus yield eigenvalues quite different from the ionization energy (353 mHa for RB88).

The limited relevance of the transverse exchange for standard electronic structure properties is confirmed by the atomic ionization potentials (IPs) listed in Table 3. In this table, fully self-consistent results including the transverse exchange are compared with energies obtained by neglecting E_x^T. In all cases, complete subshells are ionized, in order to avoid all ambiguities related to the handling of current contributions for open subshells. Table 3 demonstrates that the impact of E_x^T on these IPs is marginal even for the heaviest atoms. One has to consider the removal of the innermost electrons, in order to see a sizable effect, such as the removal of the L-shell electrons of Hg shown in Table 4.

The results collected so far raise the question whether a perturbative treatment of E_x^T is sufficient? Corresponding data are also included in Table 4. It is obvious that

the a posteriori evaluation of the transverse exchange with the RKS spinors obtained by a self-consistent calculation with the Coulomb exchange gives highly accurate results for both the L- and the K-shell ionization energies.

Relativistic Corrections in E_{xc}: II. Coulomb Exchange

Once the transverse exchange is set aside, it remains to analyze the role of relativity in E_x^C. In the case of 1-electron systems, E_x^C trivially reduces to a simple density functional

$$E_x^{C,1-\text{electron}} = -\frac{e^2}{2}\int d^3r \int d^3r' \frac{n(\boldsymbol{r})\, n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}.$$

Similarly, the Fock expression (39) can be rewritten as

$$E_x^{C,2-\text{electron}} = -\frac{e^2}{4}\int d^3r \int d^3r' \frac{n(\boldsymbol{r})\, n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}$$

for spherically symmetric 2-electron systems with $j = 1/2$ (e.g., an atomic $1s$-shell). In these limits, the relativistic corrections to the functional dependence of E_x^C on the density vanish, since the corresponding nonrelativistic functionals have exactly the same form. In the general situation, however, the functional $E_x^C[j]$ differs from the exact nonrelativistic exchange functional. Unfortunately, the orbital-dependent expression (39) combines the relativistic corrections to $E_x^C[j]$ itself with the relativistic corrections in j^μ in an inextricable way. The discussion of the former corrections is therefore continued later in the context of the relativistic LDA.

The present section focuses on the importance of an exact treatment of the relativistic Coulomb exchange as compared to the use of nonrelativistic LDA or GGA functionals in R(S)DFT, without the attempt to resolve the role of the relativistic corrections in the functional $E_x^C[j]$. The starting point is a brief comparison of the purely n-dependent RKS formalism, Eqs. (24), (25), with collinear RSDFT, Eqs. (33)–(37). Often Eq. (24) is used to discuss open-shell systems, ignoring, however, all current contributions to E_{xc}. Prototype results of this approach are provided in Table 5, which lists the IPs of a number of atoms. While the ionization of s-shells is reproduced rather well by the purely n-dependent RKS formalism, the deviations become sizable as soon as p- or d-electrons are involved. Hence, reliable results should only be expected from the RSDFT approach, which is therefore consistently used in the following.

For heavy open-shell atoms, the alignment of spins favored by the exchange interaction is competing with spin–orbit coupling, favoring good j. The ordering of the RKS eigenvalues (and, eventually, the occupation of the RKS states) depends on the strength of the exchange coupling and therefore on the exchange functional. A comparison of results obtained with the exact Coulomb exchange (39), using Eqs. (64) and (65), with the nonrelativistic LDA is shown in Fig. 2 [25]. Figure 2

Table 5 REXX IPs of selected atoms calculated from total energy differences: purely n-dependent RKS equations (24) versus spin-density-dependent RKS equations (33) (all energies in mHa; point nuclei have been used)

	Eq. (24)	Eq. (33)
Zr	204	217
Hf	203	224
Ag	229	231
Au	280	282
Cd	269	267
Hg	313	312
Sn	229	246
Pb	239	240
Sb	275	309
Bi	238	253
I	365	345
At	318	304

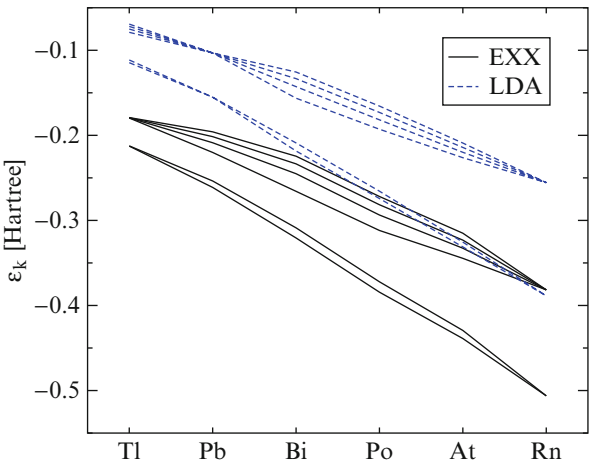


Fig. 2 RKS eigenvalues of the $6p$-subshell for $6p$-elements with occupation $6p^n$: exact relativistic Coulomb exchange versus nonrelativistic LDA (including correlation [39]). Lines are drawn to guide the eye

demonstrates that, even for heavy elements, the splitting of states resulting from the exact exchange is significantly larger than that produced by the LDA. This is particularly noteworthy for Pb, for which the LDA yields the usual degeneracy of the closed $6p_{1/2}$-subshell.

The strength of the exchange coupling is particularly important for s-d transfer energies – reproducing these energies has been a long-standing challenge for DFT [40–42]. As an example, Fig. 3 shows the energy differences between the lowest-lying states of the Cr atom, i.e., the $3d^5(6S)4s\ ^7S_3$ ground state, the first excited state, $3d^5(6S)4s\ ^5S_2$, and the $3d^44s^2\ ^5D$ multiplet [25]. The excitation to the $3d^5(6S)4s\ ^5S$ state requires the inversion of the $4s$ spin, that to the $3d^44s^2\ ^5D$ multiplet the transfer of a d-electron to the $4s$-state. Figure 3 demonstrates that

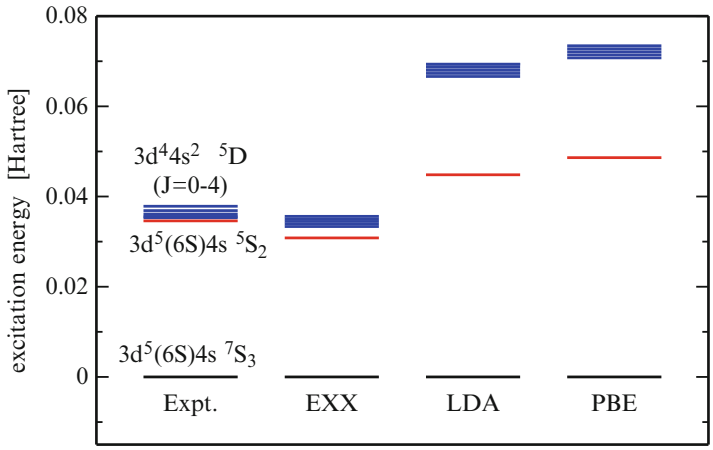


Fig. 3 Low-lying levels of Cr: EXX results versus nonrelativistic LDA [39], PBE-GGA [43], and experimental [44] data (all DFT results were obtained by solution of the RKS equations (33), using (71) in the case of the LDA and GGA). The experimental $3d^5(6S)4s\ ^5S_2$ state is only 0.7 mHa lower than the lowest state ($J = 0$) of the $3d^44s^2\ ^5D$ multiplet and can therefore not be resolved on the scale used. In the case of the DFT results, the $3d^44s^2$-energies correspond to different occupations of the $3d$-substates

the exact (relativistic) exchange provides a much better account of the exchange coupling than the LDA or GGA.

Relativistic Corrections in E_{xc}: III. Coulomb Correlation

Given the results for the RDFT exchange, neglect of the transverse interaction in the correlation functional seems legitimate. Prototype RDFT results for Coulomb correlation energies are given in Tables 6 and 7. Table 6 lists the nonrelativistic correlation energies of the helium isoelectronic series together with the corresponding relativistic corrections. In the nonrelativistic limit, exact reference data are available, so that an unambiguous examination of density functionals is possible. One observes that the orbital-dependent expression (42) leads to very accurate E_c for highly charged ions, consistent with its first-principles character. For this reason, one would expect the relativistic correlation energies obtained with $E_c^{(2)}$ to be quite reliable. Neither the nonrelativistic LDA nor the PW91-GGA reproduces the Z-dependence of the difference between the relativistic and the nonrelativistic $E_c^{(2)}$-data, indicating the need for relativistic corrections to these functionals.

Corresponding energies for neutral noble gas atoms are listed in Table 7. A comparison of $E_c^{(2)}$ with the reference energies available for light atoms indicates a substantial overestimation of atomic correlation energies by $E_c^{(2)}$. In fact, the deviation increases with increasing electron number. Although the relativistic

Table 6 Coulomb correlation energies of the He isoelectronic series: nonrelativistic values (E_c^{NR}) and relativistic corrections $E_c^R - E_c^{NR}$ from $E_c^{(2)}$ [24], Eq. (42), LDA [39] and PW91-GGA [45]. Also given are the exact nonrelativistic correlation energies [46,47] and the relativistic corrections from the RLDA [21,48]. All results have been calculated a posteriori with the KS states obtained by self-consistent EXX calculations (all energies in mHa)

	$-E_c^{NR}$				$E_c^{NR} - E_c^R$			
Ion	LDA	GGA	$E_c^{(2)}$	exact	LDA	GGA	RLDA	$E_c^{(2)}$
He	112.8	45.9	48.21	42.04	0.0	−0.0	0.0	0.00
Ne^{8+}	203.0	61.7	46.81	45.69	0.1	−0.1	0.3	−0.07
Zn^{28+}	267.2	71.3	46.67	46.34	0.9	−0.6	6.6	−0.19
Sn^{48+}	297.7	76.0	46.65	46.47	2.8	−1.7	24.4	0.72
Yb^{68+}	318.0	79.3	46.63	46.53	5.9	−3.3	58.3	3.71
Th^{88+}	333.2	81.7	46.62	46.56	11.0	−5.6	117.1	11.00

Table 7 As Table 6, but for neutral noble gas atoms

	$-E_c^{NR}$				$E_c^{NR} - E_c^R$			
Atom	LDA	GGA	$E_c^{(2)}$	exact	LDA	GGA	RLDA	$E_c^{(2)}$
Ne	746	382	477	390	0	−0	1	0
Ar	1,431	771	866	722	1	−0	3	1
Kr	3,284	1,914	2,151		8	−2	23	13
Xe	5,199	3,149	3,487		23	−4	76	36
Rn	9,027	5,706	6,260		82	−18	303	208

corrections calculated with $E_c^{(2)}$ should provide a reasonable account of the true corrections (since they originate primarily from the innermost states for which $E_c^{(2)}$ gives accurate results), their inclusion seems to be of secondary importance for standard electronic structure calculations: they are completely masked by the misrepresentation of correlation effects for the valence electrons even in the case of first-principles functionals.

Relativistic Local Density Approximation

The relativistic LDA (RLDA) approximates the xc-energy density of the actual system by that of a relativistic homogeneous electron gas (RHEG), e_{xc}^{RHEG}, with a density equal to the local density $n(\boldsymbol{r})$ of the actual system,

$$E_{xc}^{RLDA}[n] = \int d^3r\, e_{xc}^{RHEG}(n(\boldsymbol{r})). \tag{67}$$

Since $\boldsymbol{j}$ vanishes in the uniform electron gas, there is no current dependence in the RLDA. It should be noted, however, that the RLDA is an approximation to the functional $E_{xc}[n, \boldsymbol{j} = \boldsymbol{0}]$ of four-current RDFT rather than to the functional

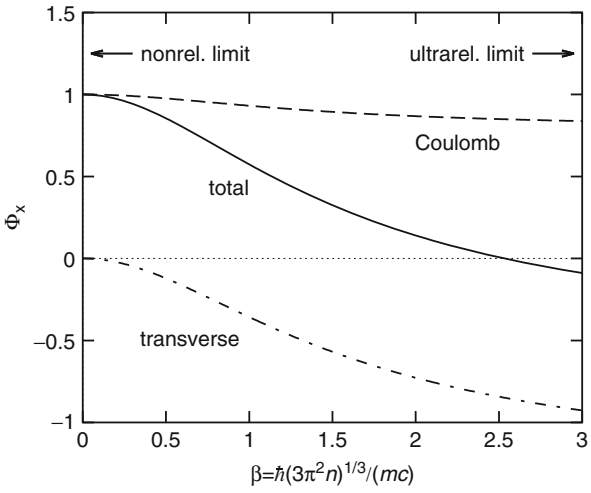


Fig. 4 Relativistic correction factor for the LDA exchange energy density: Coulomb contribution, transverse contribution and total correction Φ_x

$E_{xc}[n, \boldsymbol{j}[n]]$ of the purely n-dependent form of RDFT in which current contributions are represented as density functionals. The exchange contribution [3, 4, 49–52] can be written as product of the nonrelativistic LDA exchange energy density e_x^{HEG},

$$e_x^{HEG}(n) = -\frac{3(3\pi^2)^{1/3}}{4\pi} e^2 n^{4/3}, \tag{68}$$

and a relativistic correction factor Φ_x,

$$e_x^{RHEG}(n) = e_x^{HEG}(n)\ \Phi_x(\beta). \tag{69}$$

Φ_x is most suitably expressed in terms of the ratio of the local Fermi momentum and mc,

$$\beta = \frac{\hbar(3\pi^2 n)^{1/3}}{mc}. \tag{70}$$

The total e_x^{RHEG} can be split into a Coulomb and a transverse component according to Eqs. (39) and (40), in order to resolve the two sources of corrections. The variation of the corresponding factors $\Phi_x^{C/T}$ as well as of the total Φ_x with β is shown in Fig. 4. The β-dependence of Φ_x^C reflects the fact that the expression (39), when written as a functional of the density, in general differs from the exact nonrelativistic $E_x[n]$: while the exact nonrelativistic exchange also has the form of the Fock term, it is evaluated with nonrelativistic KS orbitals rather than the RKS spinors of (39). However, Φ_x^C varies only weakly with β, in contrast to Φ_x^T. In fact, Φ_x^T dominates for high densities and ultimately leads to a sign change of e_x^{RHEG}.

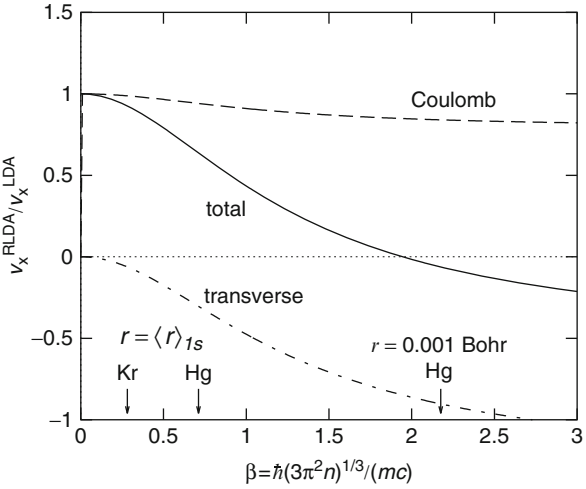


Fig. 5 Relativistic correction factor of LDA exchange potential. The β-values corresponding to the densities of Kr and Hg at the r-expectation values of the $1s$-orbitals ($r =< r >_{1s}$) and the density of Hg at $r = 0.001$ Bohr are also indicated (nuclear charge distribution and c as in Table 1)

Table 8 Transverse exchange energies (E_x^T) for closed subshell atoms: self-consistent REXX, RLDA, and B88-RGGA results (all energies in Ha; nuclear charge distribution and c as in Table 1)

Atom	REXX	RLDA	RGGA
He	0.000064	0.000147	0.000059
Ne	0.0167	0.0350	0.0167
Ar	0.132	0.249	0.132
Zn	0.758	1.318	0.757
Kr	1.417	2.391	1.415
Cd	3.797	6.131	3.796
Xe	5.693	9.039	5.691
Yb	13.842	21.418	13.837
Hg	22.071	33.957	22.054
Rn	28.547	43.979	28.519
No	53.313	84.222	53.101

Figure 5 shows the corresponding β-dependence of the RLDA exchange potential. The relevant range of densities is indicated by the β-values found at the r-expectation values of the $1s_{1/2}$-orbitals of Kr and Hg. In addition, the density of Hg at $r = 0.001$ Bohr is marked. At this density, the total RLDA correction factor for v_x has already changed its sign, so that the RLDA exchange potential becomes repulsive. This result is, however, in contradiction to the exact v_x, which remains negative at $r = 0.001$ Bohr (see Fig. 1).

The drastic overestimation of the transverse exchange in atoms by the RLDA is also obvious from the corresponding exchange energies listed in Table 8. The results in Fig. 5 and Table 8 point at the fundamental differences between the highly

Table 9 Self-consistent exchange-only ground state energies of closed subshell atoms: (R)LDA and (R)GGA results versus REXX reference data. E_x^T has been included in the RLDA, RGGA, and REXX calculations, but neglected to generate the RCLDA and RCEXX data. The PW91 form [53] has been applied for the GGA (all energies in mHa; nuclear charge distribution and c as in Table 1)

Atom	$-E$	$E - E^{\mathrm{REXX}}$				$-E$	$E - E^{\mathrm{RCEXX}}$
	REXX	LDA	RLDA	GGA	RGGA	RCEXX	RCLDA
He	2,862	138	138	6	6	2,862	138
Ne	128,674	1,038	1,080	−43	−24	128,690	1,062
Ar	528,546	2,159	2,458	−111	41	528,678	2,341
Zn	1,793,840	3,119	4,702	−1,146	−263	1,794,598	4,140
Kr	2,787,429	3,671	6,543	−1,683	−22	2,788,848	5,565
Cd	5,589,496	3,197	10,556	−4,537	−35	5,593,299	8,213
Xe	7,441,173	2,315	13,161	−6,705	83	7,446,876	9,800
Yb	14,053,750	−4,778	20,888	−17,660	−894	14,067,621	13,272
Hg	19,626,705	−11,491	29,161	−27,253	−257	19,648,826	17,204
Rn	23,573,354	−17,409	35,207	−35,145	−9	23,601,969	19,677
No	36,687,173	−43,631	56,937	−68,097	−1,344	36,740,625	25,787

localized core states in atoms and the infinitely extended states of a high-density electron gas, for which the finite speed of the photons exchanged between states plays a completely different role.

Table 9 provides some total RLDA energies for atoms. The table shows that the combination of the underestimation of the true energies by the LDA on the nonrelativistic level (obvious from the LDA data for light atoms) with the overestimation of E_x^T by the RLDA leads to particularly large errors for heavy atoms. As soon as E_x^T is dropped, the RLDA performs as well as the LDA does in nonrelativistic DFT, as can be seen from the Coulomb exchange data in Table 9.

The consistent application of the RKS equations (29) or (33) requires the use of $\boldsymbol{m}$-dependent xc-functionals. The corresponding RLDA is based on a spin-polarized RHEG [15, 54–56]. Restricting the discussion to first-order corrections in $1/c^2$, the xc-energy as well as the charge and magnetization density of this system can be expressed in terms of spin-up and spin-down Fermi momenta, $k_{\mathrm{F},\sigma}$. Inversion of the functions $n(k_{\mathrm{F},\uparrow}, k_{\mathrm{F},\downarrow})$ and $|\boldsymbol{m}(k_{\mathrm{F},\uparrow}, k_{\mathrm{F},\downarrow})|$ then yields the desired functional $e_x^{\mathrm{RHEG}}(n, |\boldsymbol{m}|)$ [56]. Prototype ground state energies for open-shell atoms obtained with this functional (XRR) are given in Table 10. It turns out, however, that the XRR functional yields essentially the same results as the combination of the $\boldsymbol{m}$-independent expression (67) with the spin-dependence of the nonrelativistic exchange functional [57], applied to $n_\pm$,

$$E_x^{\mathrm{RLSDA}}[n_+, n_-] = \frac{1}{2}\left\{E_x^{\mathrm{RLDA}}[2n_+] + E_x^{\mathrm{RLDA}}[2n_-]\right\}. \tag{71}$$

Table 10 Dependence of atomic exchange-only ground state energies ($-E$) on the treatment of spin: magnetization-dependent form of the weakly relativistic LDA (XRR [56]) versus combination of the purely n-dependent RLDA (69) (with the relativistic corrections restricted to first order in $1/c^2$) with Eq. (71) (all energies in mHa; nuclear charge distribution and c as in Table 1)

Atom	XRR	Eq. (71)
Cr	1,045,939.4	1,045,939.3
Fe	1,267,112.3	1,267,112.0
Eu	10,813,724.3	10,813,723.0
W	16,099,343.5	16,099,343.5
Au	18,962,228.4	18,962,228.4
U	27,909,051.4	27,909,051.3
Am	30,313,657.0	30,313,656.6

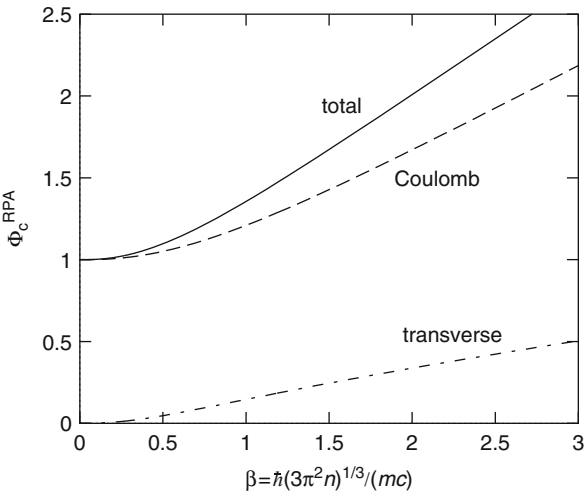


Fig. 6 Relativistic correction factor for the RPA contribution to the correlation energy density of the RHEG: Coulomb contribution, transverse contribution and total correction Φ_c^{RPA}

If the functional (71) is consistently restricted to first order in $1/c^2$, the resulting atomic ground state energies are extremely close to those obtained with the XRR approach (see Table 10).

The correlation energy density e_c^{RHEG} of the RHEG is only available in the RPA [21,48]. The corresponding relativistic correction factor is shown in Fig. 6, together with its Coulomb and transverse components (the latter is attractive in the case of RPA correlation). The RPA includes the leading relativistic correction, but misses the second-order exchange which contributes substantially for realistic densities. Accurate results from the RPA can therefore only be expected in the ultrarelativistic regime. In view of this restriction and of the well-known limitations of the LDA, it is no surprise that the RPA-based RLDA misrepresents the relativistic corrections in atomic correlation energies even if the transverse interaction is neglected (see Tables 6 and 7).

Relativistic Generalized Gradient Approximation

First-principles gradient corrections to the RLDA can be derived from the long-wavelength expansion of the current–current response function of the RHEG [7]. This expansion determines in particular the lowest order gradient term proportional to $(\nabla n)^2$. In view of the lacking information on the RHEG response function, however, a first-principles relativistic gradient correction is not available so far.

A semiempirical relativistic GGA (RGGA) has been constructed in [36, 58]. Restricting the discussion to exchange, this functional has the form

$$E_{\mathrm{x}}^{\mathrm{RGGA}} = E_{\mathrm{x}}^{\mathrm{RLDA}}[n] + \int d^3r\, e_{\mathrm{x}}^{\mathrm{GGA}}(n,\xi)\, \Phi_{\mathrm{x}}^{\mathrm{GGA}}(\beta), \tag{72}$$

with the dimensionless gradient $\xi = [\nabla n/(2(3\pi^2 n)^{1/3} n)]^2$ and $e_{\mathrm{x}}^{\mathrm{GGA}}(n,\xi)$ denoting the energy density of a nonrelativistic GGA – both the B88- and the PW91-form have been used for $e_{\mathrm{x}}^{\mathrm{GGA}}$ in order to check the consistency of the ansatz (72) (the functional (72) has to be combined with Eq. (71), when dealing with magnetic systems). The relativistic correction factor $\Phi_{\mathrm{x}}^{\mathrm{GGA}}$ has been optimized to reproduce the exact relativistic E_{x} of closed-subshell atoms, with the two GGAs leading to very similar shapes for $\Phi_{\mathrm{x}}^{\mathrm{GGA}}(\beta)$ [36, 58]. The resulting RGGA consistently improves the energetics of atoms over both the nonrelativistic GGA and the RLDA, as can for instance be seen from Tables 8 and 9.

The same basic approach has been used to set up a RGGA for correlation [58], with analogous results for atoms. However, relativistic and correlation effects are rarely important simultaneously (since they usually involve different shells), so that no details are given at this point.

Relativistic Corrections in E_{xc}: IV. Bonding Properties

The accuracy of the RGGA allows an analysis of the significance of relativistic corrections to the xc-functional for standard electronic structure properties [59–62]. Corresponding results are given in Tables 11 and 12. In both cases, Au is considered which exhibits the effects of relativity most clearly [66]. The particular importance of relativity for Au is illustrated by a comparison of relativistic with fully nonrelativistic LDA results (obtained by solution of the nonrelativistic KS equations). A substantial overestimation of bond lengths and underestimation of bond energies and elastic constants by the fully nonrelativistic approach is observed, reflecting the missing relativistic contraction of the $6s_{1/2}$-orbital (compare [61,65,67]). In contrast to relativistic kinematics, however, the relativistic corrections to the xc-functional have little impact on these data: neither for the spectroscopic constants of the Au dimer nor for the cohesive properties of bulk Au and Pt, a significant difference between GGA and RGGA results is found. As one might have expected, the sizable relativistic corrections to the ground state energies of the dimer and the bulk on the

Table 11 Spectroscopic constants of noble metal dimers: LDA and BP86-GGA [37, 63] versus relativistic BP86-GGA. For comparison also fully nonrelativistic LDA results are given

	Cu_2			Au_2		
Method	R_e (Bohr)	D_e (eV)	ω_e (cm^{-1})	R_e (Bohr)	D_e (eV)	ω_e (cm^{-1})
nonrel. calc., LDA[64, 65]	4.10	2.65	330	5.08	1.95	135
LDA[59, 61]	4.05	2.86	307	4.64	3.00	196
GGA[59]	4.16	2.28	287	4.75	2.30	179
RGGA[59]	4.17	2.27	285	4.76	2.27	177
Expt.	4.20	2.05	265	4.67	2.30	191

Table 12 Lattice constant a_0 and cohesive energy E_{coh} of Au and Pt obtained by LAPW calculations with the nonrelativistic as well as the relativistic forms of the LDA and the PW91-GGA [62] in comparison to experiment [68, 69]. The atomic ground state energies have here been evaluated with RSDFT, Eq. (33), unlike in Ref. [62], in which the purely n-dependent RKS approach was used. For comparison also fully nonrelativistic LDA results are given (these data have been obtained with the plane-wave pseudopotential approach, with the valence space including the $5s$, $5p$, $5d$, and $6s$-states)

	Au		Pt	
Method	a_0 (Bohr)	$-E_{coh}$ (eV)	a_0 (Bohr)	$-E_{coh}$ (eV)
nonrel. calc., LDA	8.06	3.30	7.68	5.91
LDA	7.68	3.97	7.36	6.35
RLDA	7.68	3.94	7.37	6.32
GGA	7.87	2.73	7.51	4.86
RGGA	7.88	2.72	7.52	4.82
Expt.	7.71	3.81	7.41	5.84

one hand and the atoms on the other hand cancel out in the energy surface. The accuracy of the data listed is completely dominated by the basic type of functional used, as can be seen from a comparison of (R)GGA and (R)LDA numbers.

Summary

At the present level of sophistication, the inclusion of relativistic corrections in LDA- or GGA-type approximations for the xc-energy functional does not seem to be necessary, at least in standard electronic structure calculations in which an accurate description of the inner shells is irrelevant. On the one hand, the RLDA overestimates the effect of the transverse interaction drastically: the high densities required for the transverse correction to become sizable are only found for the extremely localized innermost shells, which, however, have little in common with a uniform electron gas. On the other hand, the transverse correction has little effect on the valence states, so that its more realistic description by suitably modeled

GGA-type functionals is not really visible in the properties which one usually addresses with DFT calculations. It remains to be seen, however, whether truly current-dependent approximations for the xc-functional change this picture.

Acknowledgements I am very grateful to R. M. Dreizler and D. Ködderitzsch for many valuable discussions on the topics of this contribution.

References

1. Engel E, Dreizler RM (2011) Density functional theory: an advanced course. Springer, Berlin
2. Rajagopal AK, Callaway J (1973) Phys Rev B 7:1912
3. Rajagopal AK (1978) J Phys C 11:L943
4. MacDonald AH, Vosko SH (1979) J Phys C 12:2977
5. Eschrig H, Seifert G, Ziesche P (1985) Solid State Commun 56:777
6. Hohenberg P, Kohn W (1964) Phys Rev 136B:864
7. Engel E, Müller H, Speicher C, Dreizler RM (1995) In: Gross EKU, Dreizler RM (eds) Density functional theory. NATO ASI series B, vol 337. Plenum, New York, p 65
8. Engel E (2002) In: Schwerdtfeger P (ed) Relativistic electronic structure theory, part 1, chapter 10. Fundamentals. Elsevier, Amsterdam, p 524
9. Eschrig H, Servedio VDP (1999) J Comput Chem 20:23
10. Eschrig H (2003) The fundamentals of density functional theory. Edition am Gutenbergplatz, Leipzig
11. Lieb EH (1983) Int J Quantum Chem 24:243
12. Huang KN, Aoyagi M, Chen MH, Crasemann B, Mark H (1976) At Data Nucl Data Tables 18:243
13. Labzowsky L, Goidenko I, Tokman M, Pyykkö P (1999) Phys Rev A 59:2707
14. Liberman D, Waber JT, Cromer DT (1965) Phys Rev 137A:27
15. Ramana MV, Rajagopal AK (1981) J Phys C 14:4291
16. Diener G, Gräfenstein J (1989) J Phys Condens Matter 1:8445
17. Jansen HJF (1999) Phys Rev B 59:4699
18. Wang F, Liu WJ (2003) J Chin Chem Soc (Taipei) 50:597
19. Anton J, Fricke B, Engel E (2004) Phys Rev A 69:012505
20. Laskowski R, Madsen GKH, Blaha P, Schwarz K (2004) Phys Rev B 69:140408(R)
21. Ramana MV, Rajagopal AK (1981) Phys Rev A 24:1689
22. Engel E, Facco Bonetti A, Keller S, Andrejkovics I, Dreizler RM (1998) Phys Rev A 58:964
23. Görling A, Levy M (1994) Phys Rev A 50:196
24. Facco Bonetti A (2000) Implizite Dichtefunktionale: Untersuchung eines störungstheoretische Zugangs für das Korrelationsenergie funktional. Ph.D. thesis, Univ. Frankfurt
25. Engel E, Ködderitzsch D, Ebert H (2008) Phys Rev B 78:235123
26. Kotani T (1998) J Phys Condens Matter 10:9241
27. Kotani T (2000) J Phys Condens Matter 12:2413
28. Furche F (2001) Phys Rev B 64:195120
29. Fuchs M, Gonze X (2002) Phys Rev B 65:235109
30. Sharp RT, Horton GK (1953) Phys Rev 90:317
31. Talman JD, Shadwick WF (1976) Phys Rev A 14:36
32. Ködderitzsch D, Ebert H, Engel E (2008) Phys Rev B 77:045101
33. Shadwick BA, Talman JD, Norman MR (1989) Comput Phys Commun 54:95
34. Engel E, Keller S, Facco Bonetti A, Müller H, Dreizler RM (1995) Phys Rev A 52:2750
35. Hisaka K et al (Particle Data Group) (1992) Phys Rev D 45(11):Part II
36. Engel E, Keller S, Dreizler RM (1996) Phys Rev A 53:1367
37. Becke AD (1988) Phys Rev A 38:3098

38. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785
39. Vosko SH, Wilk L, Nusair M (1980) Can J Phys 58:1200
40. Harris J, Jones RO (1978) J Chem Phys 68:3316
41. Gunnarsson O, Jones RO (1985) Phys Rev B 31:7588
42. Lagowski JB, Vosko SH (1989) Phys Rev A 39:4972
43. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77:3865
44. Ralchenko Y, Kramida AE, Reader J, Team NA (eds) (2008) NIST atomic spectra database (version 3.1.5), [Online]. Available: http://physics.nist.gov/asd3 2 Sept 2008. National Institute of Standards and Technology, Gaithersburg
45. Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Singh DJ, Fiolhais C (1992) Phys Rev B 46:6671
46. Davidson ER, Hagstrom SA, Chakravorty SJ, Umar VM, Froese Fischer C (1991) Phys Rev A 44:7071
47. Chakravorty SJ, Gwaltney SR, Davidson ER, Parpia FA, Froese Fischer C (1993) Phys Rev A 47:3649
48. Facco Bonetti A, Engel E, Dreizler RM, Andrejkovics I, Müller H (1998) Phys Rev A 58:993
49. Akhiezer IA, Peletminskii SV (1960) Zh. Eksp. Teor. Fiz. 38:1829 [Sov. Phys. JETP 11, 1316 (1960)]
50. Zapolsky HS (1960) Lns report, Cornell University
51. Jancovici B (1962) Nuovo Cim XXV:428
52. Freedman BA, McLerran LD (1977) Phys Rev D 16:1130, 1147, 1169
53. Perdew JP (1991) In: Ziesche P, Eschrig H (eds) Electronic structure of solids 1991. Akademie Verlag, Berlin, p 11
54. Ramana MV, Rajagopal AK (1979) J Phys C 12:L845
55. MacDonald AH (1983) J Phys C 16:3869
56. Xu BX, Rajagopal AK, Ramana MV (1984) J Phys C 17:1339
57. Oliver GL, Perdew JP (1979) Phys Rev A 20:397
58. Engel E, Keller S, Dreizler RM (1998) In: Dobson JF, Vignale G, Das MP (eds) Electronic density functional theory: recent progress and new directions. Plenum, New York, p 149
59. Mayer M, Häberlen OD, Rösch N (1996) Phys Rev A 54:4775
60. Varga S, Engel E, Sepp WD, Fricke B (1999) Phys Rev A 59:4288
61. Liu W, van Wüllen C (2000) J Chem Phys 113:2506
62. Schmid RN, Engel E, Dreizler RM, Blaha P, Schwarz K (1998) Adv Quantum Chem 33:209
63. Perdew JP (1986) Phys Rev B 33:8822
64. Painter GS, Averill FW (1983) Phys Rev B 28:5536
65. van Wüllen C (1995) J Chem Phys 103:3589
66. Pyykkö P (1988) Chem Rev 88:563
67. Takeuchi N, Chan CT, Ho KM (1989) Phys Rev B 40:1565
68. Dewaele A, Loubeyre P, Mezouar M (2004) Phys Rev B 70:094112
69. Brewer L (1977) Tech. Rep. Report LBL-3720 Rev., Lawrence Berkeley Laboratory, Univ. of California, Berkeley

Part V

Relativistic Quantum Chemical Methods and Applications

Jochen Autschbach and Jun Li

Relativistic Many-Body Aspects of the Electron Electric Dipole Moment Searches Using Molecules

19

Bhanu P. Das, Malaya Kumar Nayak, Minori Abe, and V. S. Prasannaa

Contents

Introduction 582
Relativistic Configuration Interaction (RCI) Method 590
Calculations of E_{eff} in YbF 593
Calculations of E_{eff} in ThO 596
Relativistic Coupled-Cluster Method for E_{eff} 600
General Aspects 600
Theory of the Coupled-Cluster Method 600
Calculation of E_{eff} in YbF 602
Calculation of E_{eff} in Mercury Monohalides 605
Summary 607
References 607

Abstract

The existence of an intrinsic electric dipole moment for the electron (eEDM), which is a consequence of simultaneous violations of parity and time reversal symmetries, is a very important problem in physics. This is because it can

B.P. Das (✉)
Theoretical Physics and Astrophysics Group, Indian Institute of Astrophysics, Bangalore, India
e-mail: bpdas.iia@gmail.com, das@iiap.res.in

M.K. Nayak
Bhabha Atomic Research Centre, Trombay, Mumbai, India
e-mail: mk.nayak72@gmail.com

M. Abe
Department of Chemistry, Tokyo Metropolitan University, Hachioji-city, Tokyo, Japan
e-mail: abeminoriabe@gmail.com

V.S. Prasannaa
JST, CREST, Kawaguchi, Saitama, Japan
e-mail: srinivasaprasannaa@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_31

test theories beyond the Standard Model and provide insights into the matter-antimatter asymmetry in the universe. The eEDM is determined by combining the measured shift in the energy of an atom or a molecule and the effective electric field, which can be calculated. This article deals with the calculations of effective electric fields in the leading molecular candidates in the search for eEDM, using ab initio relativistic quantum chemical methods.

Keywords

Electron electric dipole moment • Simultaneous violation of parity and time reversal symmetries • Physics beyond the standard model • Matter-antimatter asymmetry • Atoms and molecules • Energy shift • Effective electric field • Configuration interaction • YbF • ThO • Relativistic coupled-cluster method • Mercury monohalides

Introduction

The electric dipole moment of the electron (eEDM) is an excellent probe of physics beyond the Standard Model [1, 2]. The Standard Model (SM), which encompasses the electroweak theory and quantum chromodynamics (QCD), is by far the most successful model to date, of particle physics, having withstood many experimental tests. However, phenomena like neutrino oscillations and questions on the nature of dark matter make it necessary to extend physics beyond the domain of SM [3]. Popular theories that go beyond SM include different variants of supersymmetric (SUSY) models.

The eEDM is an intrinsic property of an electron, which arises due to simultaneous violations of two discrete symmetries, namely, parity reversal (P) and time reversal (T) symmetries [4]. P would mean that $\vec{r} \rightarrow -\vec{r}$, where $\vec{r}$ refers to spatial coordinates. By definition, the parity operator $\boldsymbol{P}$ reverses the sign of the position operator $\boldsymbol{r}$. That is, $\boldsymbol{PrP^{-1}} = -\boldsymbol{r}$. Also, it is a unitary operator, which means that $\boldsymbol{P}=\boldsymbol{P^{-1}}=\boldsymbol{P^{\dagger}}$.

Consider a system with its electric dipole moment (EDM) operator given by $\boldsymbol{d}$. Since it contains in it the displacement operator, and since $\boldsymbol{PrP^{-1}} = -\boldsymbol{r}$, $\boldsymbol{PdP^{-1}} = -\boldsymbol{d}$. In the absence of an external electric field, a stationary state, $|\varphi\rangle$, has a nonzero EDM, given by

$$\langle \boldsymbol{d} \rangle = \langle \varphi | \boldsymbol{d} | \varphi \rangle \tag{1}$$

If the action of $\boldsymbol{P}$ on $|\varphi\rangle$ gives $|\varphi'\rangle$, then

$$\begin{aligned} \langle \boldsymbol{d} \rangle &= \langle \varphi | \boldsymbol{P^{\dagger} P d P^{-1} P} | \varphi \rangle \\ &= -\langle \varphi' | \boldsymbol{d} | \varphi' \rangle \end{aligned} \tag{2}$$

The stationary state satisfies the equation

$$\boldsymbol{H}|\varphi\rangle = E|\varphi\rangle \tag{3}$$

That is,

$$\begin{aligned} \boldsymbol{PHP^{-1}P}|\varphi\rangle &= E\boldsymbol{P}|\varphi\rangle \\ \boldsymbol{H}|\varphi'\rangle &= E|\varphi'\rangle \end{aligned} \tag{4}$$

Here, the assumption is that the Hamiltonian is invariant under space inversion, that is, $\boldsymbol{PHP^{-1}} = \boldsymbol{H}$

Therefore, $|\varphi\rangle$ and $|\varphi'\rangle$ describe stationary states with the same energy, E. If the energy level is nondegenerate, then $|\varphi'\rangle = c|\varphi\rangle$, with c being an eigenvalue of the parity operator, that is, it is either $+1$ or -1. Therefore,

$$\begin{aligned} \langle\varphi|\boldsymbol{d}|\varphi\rangle &= -\langle\varphi'|\boldsymbol{d}|\varphi'\rangle \\ &= -c^2\langle\varphi|\boldsymbol{d}|\varphi\rangle \\ &= -\langle\varphi|\boldsymbol{d}|\varphi\rangle \end{aligned} \tag{5}$$

This is true only if $\langle\boldsymbol{d}\rangle$ is zero. That is, if the Hamiltonian is invariant under parity and if the stationary state is nondegenerate, then the EDM of the system is zero.

$\boldsymbol{T}$ would mean that $t \rightarrow -t$, where t refers to time. Physically, time reversal symmetry can be understood as motion reversal.

If the Hamiltonian is rotationally invariant, then the eigenvectors of $\boldsymbol{H}$, $\boldsymbol{J^2}$, and $\boldsymbol{J_z}$ form a complete set $|E, j, m\rangle$. It is assumed that the degeneracy of these eigenvectors is associated only with the $(2j + 1)$ values of m. The EDM in one of these states is given by the Wigner-Eckart theorem:

$$\langle E, j, m|\boldsymbol{d}|E, j, m\rangle = C_{E,j}\langle E, j, m|\boldsymbol{J}|E, j, m\rangle \tag{6}$$

Here, $\boldsymbol{J}$ refers to the angular momentum operator, and $C_{E,j}$ is not a function of m. It can be shown that if $|u'\rangle = \boldsymbol{T}|u\rangle$, and if $|v'\rangle = \boldsymbol{T}|v\rangle$, then $\langle u'|v'\rangle = \langle u|v\rangle^*$ [5]. If $|u\rangle$ is $|\varphi\rangle$, and $|v\rangle$ is $\boldsymbol{d_\alpha}|\varphi\rangle$, where $\boldsymbol{d_\alpha}$ is a component of the EDM operator, then

$$\begin{aligned} |u'\rangle &= |\varphi'\rangle = \boldsymbol{T}|\varphi\rangle \\ |v'\rangle &= \boldsymbol{T}\boldsymbol{d}_\alpha|\varphi\rangle = \boldsymbol{d_\alpha}|\varphi'\rangle \end{aligned}$$

The last equation makes use of the fact that $\boldsymbol{Td_\alpha} = \boldsymbol{d_\alpha T}$. Therefore,

$$\langle\varphi'|\boldsymbol{d_\alpha}|\varphi'\rangle = \langle\varphi|\boldsymbol{d_\alpha}|\varphi\rangle^* \tag{7}$$

Since $\boldsymbol{d_\alpha}$ is Hermitian,

$$\langle\varphi'|\boldsymbol{d_\alpha}|\varphi'\rangle = \langle\varphi|\boldsymbol{d_\alpha}|\varphi\rangle \tag{8}$$

Similarly,

$$\langle\varphi'|\boldsymbol{J}|\varphi'\rangle = -\langle\varphi|\boldsymbol{J}|\varphi\rangle \tag{9}$$

This makes use of the fact that $\boldsymbol{d_\alpha J} = -\boldsymbol{J d_\alpha}$. Also,

$$\boldsymbol{J}_z|E, j, m\rangle = \hbar m|E, j, m\rangle \tag{10}$$

$$\boldsymbol{T J}_z \boldsymbol{T}^{-1}\boldsymbol{T}|E, j, m\rangle = \hbar m \boldsymbol{T}|E, j, m\rangle \tag{11}$$

$$\boldsymbol{J}_z(\boldsymbol{T}|E, j, m\rangle) = -\hbar m(\boldsymbol{T}|E, j, m\rangle) \tag{12}$$

Since $\boldsymbol{J}_z|E, j, -m\rangle = -\hbar m|E, j, -m\rangle$, $\boldsymbol{T}|E, j, m\rangle$ and $|E, j, -m\rangle$ vary only by a phase factor (recalling that degeneracy is only due to m). Hence, if $|\varphi\rangle = |E, j, m\rangle$, and $|\varphi'\rangle = \boldsymbol{T}|E, j, m\rangle$, then

$$\langle\varphi|\boldsymbol{d_\alpha}|\varphi\rangle = \langle E, j, m|\boldsymbol{d_\alpha}|E, j, m\rangle \tag{13}$$

$$\langle\varphi'|\boldsymbol{d_\alpha}|\varphi'\rangle = \langle E, j, -m|\boldsymbol{d_\alpha}|E, j, -m\rangle \tag{14}$$

Substituting the above two expressions into Eq. (8), one obtains

$$\langle E, j, m|\boldsymbol{d_\alpha}|E, j, m\rangle = \langle E, j, -m|\boldsymbol{d_\alpha}|E, j, -m\rangle \tag{15}$$

Similarly,

$$\langle\varphi'|\boldsymbol{J}|\varphi'\rangle = \langle E, j, -m|\boldsymbol{J}|E, j, -m\rangle \tag{16}$$

$$\langle\varphi|\boldsymbol{J}|\varphi\rangle = \langle E, j, m|\boldsymbol{J}|E, j, m\rangle \tag{17}$$

$$\because \langle\varphi'|\boldsymbol{J}|\varphi'\rangle = -\langle\varphi|\boldsymbol{J}|\varphi\rangle,$$

$$\langle E, j, -m|\boldsymbol{J}|E, j, -m\rangle = -\langle E, j, m|\boldsymbol{J}|E, j, m\rangle \tag{18}$$

Using Eqs. (15) and (18) in Eq. (6) gives EDM = 0. That is, under the combined assumptions of rotational invariance and time reversal invariance, that degeneracy is only due to m, the EDM vanishes. The proof discussed above follows Ballentine's approach [5] and is not specific only to eEDM, but of a more general nature.

Therefore, there must be simultaneous violation of both these symmetries in order for the electron or any other system to have an intrinsic EDM. T violation implies CP violation, due to the well- known CPT theorem [6].

The electron would not have an EDM, if it were a point particle. However, the electron is actually surrounded by a cloud of virtual particles, and it is an asymmetry in this electron cloud that gives it its EDM.

The SM eEDM arises only at the three-loop level, while in SUSY, for example, eEDM arises even at the one-loop level. This can be understood from the fact that in SUSY, the complex phases from emission and reabsorption need not be the same and hence do not cancel out, as in the Standard Model [1]. The eEDM predicted by SM is, hence, much smaller than that predicted by SUSY. The former predicts an eEDM less than 10^{-38} e cm, while SUSY and other extensions of SM predict values of eEDM that are a few orders of magnitude larger [1].

A combination of theoretical calculations and experiment provides upper limits on the eEDM. This imposes stringent constraints on the extensions of SM. The current best limits for eEDM using atoms come from thallium (Tl), which is 20×10^{-28} e cm, with 90 % confidence level. The best limit using molecules is from ThO [7], which is 0.87×10^{-28} e cm, followed by YbF [8], which is 10.5×10^{-28} e cm, both with 90 % confidence level. Mercury halides are promising candidates for eEDM searches, with their effective electric fields being about five times that of YbF and one and a half times that of ThO.

Knowledge of eEDM may also shed light on the matter-antimatter asymmetry of the universe [9]. It is believed that the early universe had equal amounts of matter and antimatter. However, observations indicate that the observable universe is dominated by matter. The baryon asymmetry in the universe (BAU) can be expressed by a quantity, η, which is given by

$$\eta = \frac{n_B}{n_\gamma} \approx 6.1^{+0.3}_{-0.2} \times 10^{-10} \tag{19}$$

Here, the quantity n_B is the number density of net baryons, that is, the difference between the number density of baryons and that of antibaryons. n_γ refers to the number density of the cosmic microwave background radiation (CMBR) photons.

Sakharov put forth the three necessary conditions for baryogenesis, that is, having BAU from an initial baryon symmetric state, namely, (1) baryon number violation, (2) C and CP violation, and (3) out-of-thermal equilibrium conditions. There are many models of baryogenesis, which satisfy these three conditions, with one of the well-known models being electroweak baryogenesis.

The common theme that underlies both eEDM and the matter-antimatter asymmetry is CP violation. CP violation in the SM is predominantly from the CKM (Cabibbo-Kobayashi-Maskawa) matrix. The SM predicts a smaller value of eEDM than its various extensions. Also, CP violation in SM predicts η to be of the order of 10^{-18}, whereas the value of η from observations gives $\approx 6.1^{+0.3}_{-0.2} \times 10^{-10}$. The search for a model that predicts the right amount of CP violation to describe both these phenomena is a very important problem.

The SM has one Higgs doublet, which cannot produce CP violation, because only one real field remains after the Higgs mechanism. A complex phase is required for CP violation. Including a second Higgs doublet, that is, a two Higgs doublet model (2HDM), results in CP violation from the Higgs sector too. Kazarian and Kuzmin estimated the scalar CP violating parameter from cosmology and obtained an estimate for the eEDM, $|d_e| > 10^{-27}$ e cm [9]. Nataraj and Das modified the idea and extended it further, and after imposing a cutoff for d_e from the best limit then available (Tl), preliminary calculations indicated that η was 6–7 orders lower than the expected value [10]. This may be an indication that more sources of CP violation are required. As the eEDM limit improves, we can get a better handle on η. If one finds the right model that explains both d_e and η, then there would be a unified framework that predicts the right amount of CP for both the phenomena.

Applying an electric field, $\vec{E}$, to a quantum mechanical system such as an electron would shift its energy by ΔE, which is given by

$$\Delta E = -\vec{d}_e \cdot \vec{E} \tag{20}$$

Applying an electric field on an electron would accelerate it, so better systems to observe eEDM are atoms and molecules, commonly with one unpaired electron.

Consider an atomic system. The Hamiltonian, $\boldsymbol{H_a}$, is given by

$$\boldsymbol{H_a} = \sum_i \frac{\boldsymbol{p_i^2}}{2m} + \sum_{i>j} \frac{e^2}{\boldsymbol{r_{ij}}} - d_e \sum_i \boldsymbol{\sigma_i} \cdot \boldsymbol{E_i^{intl}} \tag{21}$$

The summation is over the number of electrons in the atom, p_i is the momentum of the ith electron, m is the mass of an electron, Z is the atomic number of the atom, r_i refers to the distance of the ith electron from the nucleus, and $r_{ij} = |\vec{r}_i - \vec{r}_j|$. d_e is the eEDM, σ_i are the (4×4) Pauli matrices, and E_i^{intl} is the electric field experienced by the electron due to the nucleus and the other electrons in the atom. The last term describes the interaction of the spin of an electron that has an eEDM with the internal electric field that it experiences. Since d_e is small, this term can be treated as a perturbation, $\boldsymbol{H_{eEDM}}$, to the atomic Hamiltonian, and hence,

$$\boldsymbol{H_a} = \boldsymbol{H_0} + \boldsymbol{H_{eEDM}} \tag{22}$$

Also,

$$\boldsymbol{H_0}|\psi_\alpha^0\rangle = E_0|\psi_\alpha^0\rangle \tag{23}$$

Here, $|\psi_\alpha^0\rangle$ refers to the unperturbed state of $\boldsymbol{H_0}$.

The application of an external electric field, E^{extnl}, induces an EDM given by $\sum_i e\vec{r}_i$, with the summation i going over the number of electrons. Assume that the applied electric field is along the z direction. Hence, the atomic EDM operator, $\boldsymbol{D_a}$, becomes

$$\boldsymbol{D_a} = \sum_i (\boldsymbol{d_e \sigma_{z,i}} + e\boldsymbol{z_i}) \tag{24}$$

Then, the atomic EDM is given by

$$\langle \boldsymbol{D_\alpha} \rangle = \langle \psi_\alpha | \boldsymbol{D_a} | \psi_\alpha \rangle; \tag{25}$$

$$|\psi_\alpha\rangle = |\psi_\alpha^0\rangle + d_e|\psi_\alpha^1\rangle + d_e^2|\psi_\alpha^2\rangle + \ldots \tag{26}$$

Here, $|\psi_\alpha\rangle$ is the state of $\boldsymbol{H_a}$. Terms that are second order and beyond in d_e are neglected, since d_e is very small. Note that $\boldsymbol{d_e\sigma}_z$ is even under parity, while ez is odd under parity. Also, $|\psi_\alpha^0\rangle$ and $|\psi_\alpha^1\rangle$ are of opposite parities. Therefore,

$$\begin{aligned}\langle \boldsymbol{D_\alpha}\rangle &= (d_e\langle\psi_\alpha^0|\sum_i(\boldsymbol{d_e\sigma}_{z,i}|\psi_\alpha^0\rangle)\\ &+ (d_e e\langle\psi_\alpha^0|\sum_i \boldsymbol{z_i}|\psi_\alpha^1\rangle\\ &+ d_e e\langle\psi_\alpha^1|\sum_i \boldsymbol{z_i}|\psi_\alpha^0\rangle) \qquad (27)\\ &= \langle\boldsymbol{D^0}\rangle + \langle\boldsymbol{D^1}\rangle; \qquad (28)\\ \langle\boldsymbol{D^0}\rangle &= d_e\langle\psi_\alpha^0|\sum_i(\boldsymbol{d_e\sigma}_{z,i}|\psi_\alpha^0\rangle \qquad (29)\\ \langle\boldsymbol{D^1}\rangle &= d_e e\langle\psi_\alpha^0|\sum_i z_z|\psi_\alpha^1\rangle\\ &+ d_e e\langle\psi_\alpha^1|\sum_i z_z|\psi_\alpha^0\rangle \qquad (30)\end{aligned}$$

And

$$|\psi_\alpha^1\rangle = \sum_{I\neq\alpha}\frac{|\psi_I^0\rangle\langle\psi_I^0|\boldsymbol{H}_{\mathbf{eEDM}}|\psi_\alpha^0\rangle}{E_\alpha^0 - E_I^0} \qquad (31)$$

It can be shown that

$$\langle\boldsymbol{D^1}\rangle = -d_e\langle\psi_\alpha^0|\sum_i(\boldsymbol{d_e\sigma}_{z,i}|\psi_\alpha^0\rangle \qquad (32)$$

$$\because \langle\boldsymbol{D_\alpha}\rangle = 0 \qquad (33)$$

That is, in a nonrelativistic scenario, although the electron has an EDM, due to the various interactions in the atom, the atomic EDM is zero [11].

Consider the relativistic case, where the usual Hamiltonian is replaced by the Dirac-Coulomb Hamiltonian, and the wave function is built from four-component spinors. The atomic Hamiltonian is [12]

$$\begin{aligned}\boldsymbol{H_a} &= \boldsymbol{H_0} + \boldsymbol{H}_{\mathbf{eEDM}}\\ &= \sum_i c\boldsymbol{\alpha}\cdot\boldsymbol{p_i} + \beta mc^2 + \sum_{i>j}\frac{e^2}{\boldsymbol{r_{ij}}} - d_e\sum_i \beta\boldsymbol{\sigma_i}\cdot\boldsymbol{E_i^{intl}} \qquad (34)\end{aligned}$$

In the above equation, c is the speed of light, α and β refer to the Dirac matrices, $\boldsymbol{p_i}$ is the momentum of the ith electron, and m is the mass of the electron.

And the atomic EDM becomes

$$\begin{aligned}\langle \boldsymbol{D_\alpha} \rangle &= (d_e \langle \psi_\alpha^0 | \sum_i (\boldsymbol{d_e \beta \sigma}_{z,i} | \psi_\alpha^0 \rangle) \\ &+ (d_e e \langle \psi_\alpha^0 | \sum_i z_z | \psi_\alpha^1 \rangle \\ &+ d_e e \langle \psi_\alpha^1 | \sum_i z_i | \psi_\alpha^0 \rangle) \\ &= \frac{2cd_e}{\hbar} \sum_{I \neq \alpha} \frac{\langle \psi_\alpha^0 | z | \psi_I^0 \langle \psi_I^0 | i \beta \gamma_5 \boldsymbol{p^2} | \psi_\alpha^0}{E_\alpha^0 - E_I^0} + h.c. \\ &\neq 0\end{aligned} \tag{35}$$

In the above equation, h.c. stands for Hermitian conjugate. And, the relation $\boldsymbol{H_{eEDM}^{eff}} = \frac{2icd_e \beta \gamma_5 \boldsymbol{p^2}}{\hbar}$ has been used, where $\boldsymbol{H_{eEDM}^{eff}}$ is an effective eEDM Hamiltonian and γ_5 is the product of the four Dirac matrices. This demonstrates that the eEDM calculations can only be performed in a relativistic framework.

Therefore, the shift in energy of the system is given by

$$\begin{aligned}\Delta E &= -\langle \boldsymbol{D_a} \cdot \boldsymbol{E^{extnl}} \rangle \\ &= -RE^{\text{extnl}} d_e \\ &= -E_{\text{eff}} d_e\end{aligned} \tag{36}$$

Here, $R \equiv \frac{\langle \boldsymbol{D_\alpha} \rangle}{d_e}$, and is called the enhancement factor, and $E_{\text{eff}} = RE_{\text{external}}$. E_{eff} is called the effective electric field, and it is the electric field that the electron experiences in the atom. Note that the effective electric field for an atom depends on an external electric field. Typically, these fields are of the order of kV/cm. Thallium atom, which is currently the candidate among atoms that has given the best upper limit for eEDM, has $R = -466.31$

Consider a diatomic molecule. The interaction Hamiltonian due to the electron having eEDM is given by [12]

$$\boldsymbol{H_{eEDM}} - d_e \sum_{i=1}^{N_e} \boldsymbol{\beta \sigma}_i \cdot \boldsymbol{E_i^{intl}} \tag{37}$$

The internal electric field here is from the two nuclei as well as from all the other electrons in the molecule. N_e refers to the number of electrons in the molecule. The quantity that is measured in the experiment is the energy shift due to the effective electric field, which can be written as

$$\begin{aligned}\Delta E &= \langle\psi|\boldsymbol{H}_{\mathbf{eEDM}}|\psi\rangle \\ &= -d_e \sum_{i=1}^{N_e} \langle\psi|\boldsymbol{\beta\sigma}_{\boldsymbol{i}} \cdot \boldsymbol{E}_{\boldsymbol{i}}^{\mathbf{intl}}|\psi\rangle \\ &= -d_e E_{\text{eff}} \end{aligned} \tag{38}$$

$$\therefore E_{\text{eff}} = \sum_{i=1}^{N_e} \langle\psi|\boldsymbol{\beta\sigma}_{\boldsymbol{i}} \cdot \boldsymbol{E}_{\boldsymbol{i}}^{\mathbf{intl}}|\psi\rangle \tag{39}$$

E_{eff} is evaluated using a relativistic many-body approach. Therefore, the energy shift is measured from the experiment, and the effective electric field is obtained from theory. The ratio of the two gives the eEDM. In the above equation, $|\psi\rangle$ refers to the wave function of a molecular state. The calculations must be performed in the relativistic framework, since in a nonrelativistic treatment, the effective electric field is zero. In such a case, the single particle wave function is built from the four-component spinors.

Since the evaluation of the two-body term in the expression for the internal electric field is complicated, the following relationship can be used [13]:

$$\begin{aligned}-d_e \sum_{i=1}^{N_e} \beta\boldsymbol{\sigma}_{\boldsymbol{i}} \cdot \boldsymbol{E}^{\mathbf{intl}} &= [-\frac{d_e}{e}\beta\boldsymbol{\sigma} \cdot \nabla, \boldsymbol{H_0}] \\ &\quad + 2ic d_e \sum_{i=1}^{N_e} \beta\gamma_5 \boldsymbol{p}^2 \end{aligned} \tag{40}$$

where the summation over the electronic coordinates is given by i, and H_0 refers to the Dirac-Coulomb Hamiltonian, which is given by

$$\begin{aligned}\boldsymbol{H_0} = &\sum_i \left[c\boldsymbol{\alpha} \cdot \boldsymbol{p}_{\boldsymbol{i}} + \beta mc^2 - \sum_A \frac{Z_A}{|\boldsymbol{r}_{\boldsymbol{i}} - \boldsymbol{r}_{\boldsymbol{A}}|} \right] \\ &+ \sum_{i \neq j} \frac{1}{|\boldsymbol{r}_{\boldsymbol{i}} - \boldsymbol{r}_{\boldsymbol{j}}|} \end{aligned} \tag{41}$$

Here too, the summation over the electronic coordinates is indicated by i. The summation over the nuclear coordinates is indicated by A. $\boldsymbol{r_i}$ is the position vector from the origin to the site of an electron, and $\boldsymbol{r_A}$ is the position vector from the origin to the coordinate of a nucleus. Z_A is the atomic number of the Ath nucleus.

Using the above relation, one can show that

$$E_{\text{eff}} = \frac{2ic}{e} \sum_{i=1}^{N_e} \langle\psi|\beta\gamma_5 \boldsymbol{p}_{\boldsymbol{i}}^2|\psi\rangle \tag{42}$$

This expression of the effective electric field is easier to evaluate than that given by Eq. (39), because $\beta\gamma_5 \boldsymbol{p}_i^2$ is a one-body operator.

It is important to note that the eEDM is associated with CP violation, which is a particle physics phenomenon. The signature of this effect is observed in an experiment using an atom/molecule. This would fall under the domain of atomic, molecular, and optical physics. The calculation of the effective field requires relativistic quantum chemistry. Hence, the search for eEDM involves the synergy of three distinct fields, in which theory and experiment are indispensable.

In the last few years, molecules have been generally preferred over atoms. This is due to the fact that in molecules, there is hybridization of the atomic orbitals and the matrix elements of the effective eEDM operator between some orbitals may be large. The effective electric field is a function of the internal electric field (of the order of MV/cm), unlike in atoms, where their effective electric fields are a function of an applied field, which is typically of the order of kV/cm.

Experimentally, the shift in energy of a molecule in a particular state due to eEDM in the presence of an internal electric field is given by

$$\Delta E = -d_e E_{\text{eff}} \eta(E_{\text{external}}); \tag{43}$$

$$\eta(E_{\text{external}}) = \langle z.n \rangle \tag{44}$$

where η is called the polarization factor, defined as the dot product between the z axis, defined by the external electric field, and the internuclear axis, n. If η is one, which is an ideal case, then the molecule is completely polarized, that is, the external field and the internuclear axis are along the same axis. The polarization factor can be obtained from the experiment. As seen earlier, the shift in energy is from experiment too, and E_{eff} is calculated from relativistic many-body theory, and a combination of these is used to set an upper limit on eEDM.

In the following sections, we discuss the relativistic many-body methods that have been used to determine the effective electric fields of the heavy polar molecules that are currently the leading candidates in the eEDM searches.

Relativistic Configuration Interaction (RCI) Method

In this section, the method of relativistic configuration interaction (RCI) is described for obtaining the exact energy of a many-electron state and the corresponding wave function of that state. The discussions that follow, on this method, are based on the book by Szabo and Ostlund [14], where its nonrelativistic version is expounded. This also involves the method of obtaining the correlation energy of the many-electron state. Among all the approaches developed so far, the CI method is conceptually the simplest but computationally challenging. The basic idea is to diagonalize the N-electron Hamiltonian in a basis of N-electron functions (Slater determinants). In other words, we represent the exact wave function of any state as a linear combination of N-electron trial functions and use the variational method to

optimize the energy of the state. If the basis were complete, we would obtain the exact energies of ground as well as of all excited states of the system. In principle, CI provides an exact solution of the many-electron system. In practice, however, we can handle only a finite set of N-electron trial function; consequently, CI provides only upper bounds to the exact energies.

The starting point of RCI method is the Dirac Fock (DF) wave function, $|\Phi_0\rangle$. After solving the self-consistent field (SCF) equations in a finite basis set, we obtain a set of relativistic molecular orbitals, $\{\chi_i\}$. The determinant formed from the N lowest energy orbitals is $|\Phi_0\rangle$. Other than $|\Phi_0\rangle$, we can form a large number of N-electron determinants from the orbitals. It might be convenient to describe these other determinants by stating how they differ from $|\Phi_0\rangle$. Thus, the set of all possible determinants include $|\Phi_0\rangle$, the singly excited determinants $|\Phi_a^r\rangle$ (which differ from $|\Phi_0\rangle$, in having the spin orbital χ_a replaced by χ_r), the doubly excited determinant $|\Phi_{ab}^{rs}\rangle$ (differing from $|\Phi_0\rangle$ by replacing χ_a with χ_r and χ_b with χ_s), and N-tuple excited determinants. We can use these many-electron wave functions as a basis to expand the exact many-electron wave function $|\Psi\rangle$. If $|\Phi_0\rangle$ is a reasonable approximation to $|\Psi\rangle$, then a better approximation according to variation principle (which becomes exact as the basis becomes complete) is

$$|\Psi\rangle = c_0|\Phi_0\rangle + \sum_{a,r} c_a^r|\Phi_a^r\rangle + \sum_{\substack{a<b\\r<s}} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c\\r<s<t}} c_{abc}^{rst}|\Phi_{abc}^{rst}\rangle + \cdots \tag{45}$$

This is the form of full CI wave function. The restriction on the summation indices (i.e., $a < b$, $r < s$, etc.) is to avoid multiple counting of a given excited determinant. The above expression can be simplified as

$$|\Psi\rangle = c_0|\Phi_0\rangle + \sum_{a,r} c_a^r|\Phi_a^r\rangle + \left(\frac{1}{2!}\right)^2 \sum_{a<b,r<s} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \left(\frac{1}{3!}\right)^2 \sum_{\substack{a<b<c\\r<s<t}} c_{abc}^{rst}|\Phi_{abc}^{rst}\rangle + \cdots \tag{46}$$

where a factor of $(1/n!)^2$ is included in front of the summation to avoid multiple counting in the case of unrestricted summation indices.

The trial wave function given above can be used to determine the wave functions and the energies, by using the linear variational method. This consists of forming the matrix representation of the Hamiltonian in the basis of N-electron functions which we are used to expand the above equation. Then, the eigenvalues of this matrix have to be found by diagonalizing the Hamiltonian matrix. This is called the full CI matrix, and the method is referred to as full CI. The lowest eigenvalue will be an upper bound to the ground state energy of the system, and the higher eigenvalues will be upper bounds to exited states. The difference between the lowest eigenvalue ($\mathcal{E}_0$) and the DF energy (E_0) obtained within the same one-electron basis is called the basis set correlation energy. As the one-electron basis set approaches completeness, this basis set correlation energy approaches the exact correlation energy. However, the basis set correlation energy obtained by performing a full CI is exact within the subspace spanned by the one-electron basis.

To construct the full CI matrix and examine its properties, it is convenient to rewrite the previous full CI wave function in a symbolic form

$$|\Psi\rangle = c_0|\Phi_0\rangle + c_S|\Phi_S\rangle + c_D|\Phi_D\rangle + c_T|\Phi_T\rangle + c_Q|\Phi_Q\rangle + \cdots \tag{47}$$

where $|\Phi_S\rangle$ represents the terms involving single excitations, $|\Phi_D\rangle$ represents terms involving double excitations, and so on. Similarly, c_S and c_D are coefficients corresponding to terms involving single and double excitations, respectively. Using this notation, the full CI matrix takes the following form:

$$\begin{bmatrix} \langle\Phi_0|H|\Phi_0\rangle & \langle\Phi_0|H|S\rangle & \langle\Phi_0|H|D\rangle & \langle\Phi_0|H|T\rangle & \cdots \\ & \langle S|H|S\rangle & \langle S|H|D\rangle & \langle S|H|T\rangle & \cdots \\ & & \langle D|H|D\rangle & \langle D|H|T\rangle & \cdots \\ & & & & \cdots \\ & & & & \end{bmatrix}$$

H is the Hamiltonian for which Ψ is the eigenfunction. Since the matrix is Hermitian, the lower triangle will be the same as the upper triangle. As mentioned earlier, we have to diagonalize this full CI matrix to get the eigenvalues and the corresponding eigenvectors, but the following observations of the full CI matrix are important:

1. There is no coupling between the DF state and single excitations (i.e., $\langle\Phi_0|H|S\rangle = 0$). This is a consequence of Brillouin's theorem which states that all matrix elements of the form $\langle\Phi_0|H|\Phi_a^r\rangle$ are zero.
2. There is no coupling between $|\Phi_0\rangle$ and triples or quadruples. Similarly, there is no mixing between singles and quadruples. This is a consequence of the fact that all matrix elements of the Hamiltonian between Slater determinants which differ by more than two spin orbitals are zero (Slater-Condon rules). This also indicates that the blocks which are not zero are sparse. For instance, the matrix element $\langle D|H|T\rangle$ represents

$$\langle D|H|T\rangle \leftrightarrow \langle\Phi_{ab}^{rs}|H|\Phi_{cde}^{tuv}\rangle$$

   For a matrix element of this type to be nonzero, the indices a and b must be included in the set $\{c, d, e\}$ and also the indices r and s must be included in the set $\{t, u, v\}$:
3. Since there is no mixing of single excitations with $|\Phi_0\rangle$ directly, it can be expected that they have a very small influence on the ground state energy. There effect is not zero because they do mix indirectly, i.e., they interact with doubles which in turn interact with $|\Phi_0\rangle$. Although they have almost negligible effect on the ground state energy, they can influence one-electron properties like dipole moment significantly.

4. Because the double excitations mix directly with $|\Phi_0\rangle$, it may be expected that their contribution is important. In fact, for small systems, they have a dominant contribution in the correlation energy. Moreover, it turns out that quadruple excitations are more important than triple or single excitations, if one is concerned solely with the ground state energy.

As mentioned earlier, the number of determinants is extremely large even for a small system, and it becomes practically impossible to do a full CI calculation. One of the common approximations is to truncate the full CI matrix or equivalently the CI expansion for the exact many-electron wave function at certain excitation level. If one includes only single excitations in the trial function for the CI expansion, the scheme is called singly excited CI (SCI). Similarly, including single and double excitations in the CI expansion is called singly and doubly excited CI (SDCI) and so on.

For heavier system, even performing an SDCI calculation is impossible. For such systems, a better approximation is the restricted active space configuration interaction (RASCI) method. In this method, one considers a limited number of occupied orbitals in $|\Phi_0\rangle$ and a limited number of virtual (un-occupied) orbitals with respect to $|\Phi_0\rangle$. These orbitals are called active orbitals. Within these active orbitals, one could do a truncated CI calculation at certain excitation level.

Furthermore, in the RASCI method, the total active orbitals are divided into three active subspaces: (a) RAS1 with a restricted number of holes allowed, (b) RAS2 where all possible configurations are permitted, and (c) RAS3 with an upper limit on the number of electrons allowed.

In the CI method, Eq. (46) is used, to calculate the expectation value of an operator, O.

Calculations of E_{eff} in YbF

Before we describe a relativistic CI calculation of E_{eff} in YbF, we briefly discuss some of the early calculations of this quantity.

The effective electric field E_{eff} in YbF was first calculated by Kozlov using a semiempirical method [15, 16] based on experimental hyperfine structure data. Ab initio calculation of E_{eff} in the molecule was first calculated by Titov et al. [17] using a generalized relativistic effective core potential (GRECP) at the level of self-consistent field (SCF) and restricted active space (RAS). It is expected that the GRECP procedure provides reasonable accuracy of the computed quantity with small computational cost. However, it should be noted that this is not an all-electron treatment.

When the core electrons of a heavy-atom molecule do not play an active role, the effective Hamiltonian in GRECP approach, H^{Ef}, can be written as

$$\mathrm{H}^{Ef} = \sum i_v \left[h^{\mathrm{Schr}}(i_v) + \mathrm{U}^{Ef}(i_v)\right] + \sum_{i_v \geq j_v} \frac{1}{r_{i_v}, j_v} \tag{48}$$

written only for the valence and some outer core electrons, which are denoted by the indices i_v and j_v, and U^{Ef} is a relativistic effective core potential (RECP) operator simulating interactions of the explicitly treated ("pseudo-valence") electrons with those which are excluded from the RECP calculations [17]. h^{Schr} refers to the one-body part and the last term is the two-body part of the Hamiltonian.

Titov and coworkers have further improved their results by considering an effective operator (EO) technique to their previous calculations using GRECP procedure at the level of SCF and RASSCF [18]. Their EO technique is based on the framework of the second-order many-body perturbation theory (MBPT(2)), which can include a significant part of the core-valence electron correlation [19]. Assuming that the valence-valence electron correlation effect is negligible, Parpia [20] estimated E_{eff} from all-electron unrestricted Dirac-Fock (DF) (UDF) method. Quiney et al. [21] reported E_{eff} computed at the core-polarization level, with all-electron DF orbitals, neglecting the effect of pair correlation (PC) and higher order corrections to E_{eff}. The calculations cited above predict the values of E_{eff} in the interval of -18.81 to -26.05 GV/cm, which is quite large. Therefore, a more precise estimation of E_{eff} is necessary in order to set a reliable limit on d_e

In early 2006, E_{eff} for the ground $(^2\Sigma_{1/2})$ states of the YbF [8] molecule was computed, using all-electron DF orbitals at the RASCI level. Later, MBPT (2) was employed, to calculate E_{eff} for the same system [22]. The active space considered in the earlier calculation [8] using the RASCI method was not so large, but some triple excitations were considered, which was generating a large number of configuration state functions (CSFs). However, it was found that there was no significant contribution by the inclusion of the triple excitations. Later, the same RASCI approach was used, but with a sufficiently large CI space. Also, all possible single and double excitations within the active space were considered, to study E_{eff} [23]. However, the deep-lying occupied and high-lying virtual orbitals are excluded in these calculations.

The effective electric field E_{eff} for the ground state of YbF estimated from our recent RASCI calculation is compared with other calculations [17, 18, 20, 21] as well as with our previous calculation [8, 22] and the semiempirical result [16] in Table 1. As Table 1 shows, the recently estimated value of E_{eff} using the same RASCI method, but with a slightly different technique, improved over our earlier calculation [8], and the magnitude of the computed quantity has increased by around 6 % over our earlier result [8]. These calculations [23] have considered (a) all the 31 active occupied spinors containing 31 electrons in RAS1 with a maximum of two holes allowed, (b) 7 active virtual spinors in RAS2, and (c) the remaining active virtual spinors in RAS3 with a maximum of two electrons are permitted. The basic idea behind this choice is to consider all possible single and double excitations, within the active space, with respect to the DF reference configuration. Although from a methodology point of view, both the recent calculations [23] and our earlier calculations [8] are the same, there are some technical differences, such as employing a sufficiently large CI space and considering all possible single and double excitations, within the active space, with respect to the DF reference function, while in the earlier calculations, a relatively small CI space was used and

Table 1 Effective electric field E_{eff} for the ground $^2\sum_{1/2}$ state of YbF molecule

Mothods	E_{eff} (GV/CM)
Semi-empirical [16]	−26.05
GRECP-RASSCF [17]	−18.81
DF+CP [21]	−24.80
UDF (unpaired electron) [20]	−19.89
UDF (all electrons) [20]	−24.87
GRECP-RASSCF-EO [18]	−24.93
DF (Nayak et al.) [8]	19.91
RASCI (Nayak et al.) [8]	−22.49
MBPT(2) (Nayak et al.) [22]	−21.56
RASCI [23]	−24.06

considered all possible single, double, and some triple excitations, within the active space, with respect to the DF reference function. For example, the active spaces for YbF molecule considered in the recent calculations are composed of 31 electrons and a maximum of 76 active orbitals, whereas the active spaces considered in the earlier calculations [8] are composed of 31 electrons and a maximum of 56 active orbitals. The recently estimated result of E_{eff} is off by ∼3 % from that of Mosyagin et al. [18] at the level of RASSCF with the EO technique within the GRECP approach (GRECP-RASSCF-EO) and differs by ∼7 % from the semiempirical result of Kozlov [16], with the accuracy of the latter method expected to be 80 %. Similarly, our recently estimated result of E_{eff} is off by ∼3 % from the calculation of Quiney et al. [21] at the level of first-order core polarization with all-electron DF orbitals (DF + core polarization) and Parpia's all-electron UDF [20] calculation. Although the core polarization contribution is the most important, the effects of PC and higher- order terms are non-negligible. The inclusion of electron correlation through UDF method is generally not recommended, as the UDF theory suffers from spin contamination.

The inclusion of electron correlation to E_{eff} via the CI method is straightforward but computationally challenging as a large number of electrons and orbitals need to be included in the CI space. The effect of electron correlation is analyzed using the RASCI method. There are 39 doubly occupied orbitals and one singly occupied orbital in YbF, of which the 25th occupied orbital corresponds to $5s$ occupied spin orbital of Yb. As the contribution of the $5s$ orbital of Yb to E_{eff} is quite significant [15–17], this orbital is included in the CI space. The occupied orbitals above the 25th are also included in the RASCI space from energy considerations [the $4f$ orbitals of Yb and the $2p$ orbitals of F in YbF are energetically quite close (see Table 12 of Ref. [20])]. Hence, 31 active electrons (16α and 15β) are included in the CI space. In the recent calculations for E_{eff}, six sets of RASCI spaces are constructed from 31 active electrons and 36, 46, 56, 66, 76, and 86 active orbitals to analyze the result of E_{eff}. From this analysis, it was found that when more active orbitals are included in the CI space, the magnitude of E_{eff} for YbF increases gradually and reaches a value of −24.06 GV/cm for the active space containing 76 active orbitals. Increasing in the size of the active space further has almost negligible effect on the result of

E_{eff}. Therefore, using a large CI space is quite important to study E_{eff}. Also, the contribution of triple excitations, omitted in the recent calculation, but considered in our earlier calculation [8], is negligible.

Calculations of E_{eff} in ThO

The first calculation of E_{eff} for the $^3\Delta_1$ state of the ThO molecule was performed by Meyer et al. [24]. An approximate method, based on the nonrelativistic molecular structure calculations, perturbed by the Hamiltonian arising from the electron electric dipole moment (eEDM) interaction, was used in the work. The relativistic effect is accounted for in a semiempirical way, as described in [24]. Augmented correlation-consistent polarized valence quadruple-zeta (aug-cc-pvqz) basis of Dunning was considered to describe the O atom's $s - f$ orbitals. For the Th atom, a relativistic 78 electron effective core potential (ECP78MWB) was used, to describe the core of Th, and utilized the ECP78MBW aug-cc-pvqz basis set to describe the $s - f$ orbitals of the outer core and valence region. In addition, the geometry was optimized by varying the Th and O bond distance to get an equilibrium bond length, R_e, of $3.47a_0$ in the ground state, where a_0 is the Bohr radius. The optimized bond length thus obtained is used to calculate E_{eff} in the first excited state ($^3\Delta_1$) of molecule. A final result of 104 GV/cm was obtained. Due to approximate relativistic matching of nonrelativistic orbitals and semiempirical methods to incorporate the relativistic effects, the uncertainty was estimated to be $\sim 3\,\%$.

The next detailed study of the E_{eff} on the excited $^3\Delta_1$ state of ThO was by Skripnikov et al. [25]. In their calculations, they used relativistic two-component coupled-cluster formalism to include electron correlation effects. Based on this formalism, the transition energy between $^1\Sigma^+$ and excited $^3\Delta_1$ states was calculated and so were the molecular dipole moment, effective electric field, E_{eff}, electron-nucleus scalar-pseudoscalar constant $W_{P,T}$, and parallel hyperfine structure constant $A_\parallel$. A single-reference approach was used for the calculations of the above-mentioned spectroscopic constants and symmetry-violating properties. In their coupled-cluster approach, they considered single and double excitations (CCSD) and the perturbative treatment of triple cluster amplitudes (CCSD(T)). The basis set for Th is based on GRECP formalism, where the inner-core electrons ($1s$-$4f$) of Th were excluded from the molecular correlation calculations using the GRECP operator. Thus, the outermost 38 electrons in ThO molecule were treated explicitly, and the basis set for Th was constructed based on a generalized correlated scheme [18]. The contracted basis set chosen for Th was (30s, 20p, 17d, 11f, 4g, 1h)/[30s, 8p, 6d, 4f, 4g, 1h]. Only the p-, d-, and f-type Gaussians are contracted. For O, the aug-ccpvqz basis set, with two g-type basis functions removed, was employed, which is (13s, 7p, 4d, 3f)/[6s, 5p, 4d, 3f]. In addition to that, basis set enlargement corrections to the considered parameters were also analyzed, by using extended basis set for Th (with added f, g, h, and i Gaussian functions). Also, the ground- and excited-state geometries were optimized at the CCSD(T) level, with the

extended basis set used for Th. A value of 3.482 a.u for $^1\Sigma+$ and 3.515 a.u. for $^3\Delta_1$ was obtained, which is in very good agreement with the experimental results (3.478 a.u for $^1\Sigma^+$ and 3.511 a.u. for $^3\Delta_1$) [26]. Hence, the experimental bond length of ThO was used for the study of E_{eff} on the excited $^3\Delta_1$ state. Based on the method, they obtained a final result of $E_{\text{eff}} = 84.0\,\text{GV/cm}$ for the $^3\Delta_1$ state at the CCSD(T) level using 38 correlating electrons. The uncertainty in their calculation is given to be within 15 % of the computed quantity.

A more detailed and elaborate study has been performed recently by Fleig et al. [27]. Their work is based on the generalized active space (GAS) concept, introduced into the configuration interaction (CI) method of DIRAC 11 program. In the GAS-CI formalism, the total active orbital space is subdivided into an arbitrary number of subspaces with arbitrary occupation constraints. In other words, different excitation levels can be assigned to different subspaces of the total active orbital space. This is a more general and extended formalism than the RAS concept discussed earlier. For the study of the ThO molecule, Fleig et al. [27] used the Dirac-Coulomb Hamiltonian and an all-electron DF calculation to optimize the molecular spinors. In their study, they used fully uncontracted atomic Gaussian basis sets of double-zeta, triple-zeta, and quadruple-zeta quality, for the description of the electronic shells of ThO. For Th atom, Dyall's basis sets were used, whereas for O atom, Dunning's cc-pvnz-DK basis set (where $n \in \{2, 3, 4\}$) and the aug-cc-pvtz-DK set were used. More details of this work can be found in Ref. [27]. The basic model used in their Hartree-Fock open-shell calculations is based on an average-of-configuration Fock operator for two electrons in Th ($7s$, $6\delta d$) Kramers pairs with all other (96) electrons of ThO restricted to the closed shell. This model is appropriate for the region close to the equilibrium bond length of the molecule, where the dominant configurations correspond to the system $Th^{2+}\ O^{2-}$. The open-shell averaging ensures a balanced description of the low-lying electronic states of interest.

As mentioned earlier, Fleig et al. [27] have used the concept of GAS-CI in their electronic structure calculation and for the study of E_{eff} in the $^3\Delta_1$ state of the ThO molecule. For this purpose, they constructed different GAS-CI spaces for defining the CI wave functions of varying quality. As shown in Fig. 1, they partitioned the active space into seven subspaces, which contain different number of Kramers-paired spinors, five of which are active for excitations. Based on this partitioning, they assign four parameters (m,n,p,q) that define the accumulated electron occupation constraints of the subspaces. They choose five different CI wave function models for their calculations. The values of the parameters (m,n,p,q) and the corresponding GAS-CI wave function models are presented in Table 2.

The parameter, K, has been introduced to define the variable active valence spinor space. For example, $K = 3$ includes only the ($7s$, $6\delta d$) spinors of Th, resulting in 3 Kramers pairs in the fourth active space as shown in Fig. 1, thus providing a model for a balanced description of the ground state ($\Omega = 0$) and the excited states $\Omega \in \{1, 2, 3\}$ arising from $^3\Delta$ in the $\Lambda - S$ coupling picture. Similarly, $K = 5$ includes two more Kramers pairs containing π-type spinors to the fourth active space, and $K = 7$ adds two more Kramers pairs of σ-type spinors. Again,

	# of Kramers pairs	Accumulated # of electrons	
		min	max
Deleted	(176)		
Virtual	183-X	36	36
Th: 6d$\sigma\pi$, 7p, 8s, 8p Th: 7s, 6dδ	K	36-m	36
Th: 6s, 6p O: 2s, 2p	8	34-n	34
Th: 5d	5	18-p	18
Th: 5s, 5p	4	8-q	8
Frozen core	(31)		

Fig. 1 Generalized active space models for the CI wave function used in Tho molecule. The space with 183-K virtual Kramers pairs (for vTZ basis set) is considered as active virtual, whereas 31 Kramers pairs are kept a frozen and 176 virtual are deleted

Table 2 Parameters defining the occupations constraint and the label of the model space for the CI wave function of ThO molecule

Parameter values	Label of correlation model
m = 2, n = 2, p = 2, q = 2	MR_K-CISD(18)
m = 2, n = 3, p = 0, q = 0	MR_K^{+T}-CISD(18)
m = 3, n = 2, p = 0, q = 0	MR_K-CISDT(18)
m = 2, n = 2, p = 2, q = 0	MR_K-CISD(28)
m = 2, n = 2, p = 2, q = 0	MR_K-CISD(36)

$K = 9$ results from adding two more Kramers pairs of π-type, $K = 10$ includes one more Kramers pair of σ-type spinors, and finally $K = 12$ results from adding two more Kramers pairs containing π-type spinors to the fourth active space. In addition to this, the CI wave function models contain the number of total correlated electrons (in parenthesis) and include the excitation ranks, where "SDT" stands for single, double, and triple excitations, for example. The value of the parameter, for example, $n = 2$, denotes the maximum hole rank of the respective active space. In this particular example of $n = 2$, Slater determinants with zero up to two holes in the Th $(6s, 6p)$ and O $(2s, 2p)$ (which are included in third active space) would be included in the expansion of the CI wave function. Based on the GAS-CI concept discussed above, Fleig et al. [27] defined various models of CI wave functions in their study, which are labeled as MR_3-CISD(18), $MR_5 - CISD(18)$, MR_7-CISD(18), MR_9-CISD(18), MR_{10}-CISD(18), and MR_{12}-CISD(18). These are all basic model spaces which are at the level of single and double excitations and contain 18 electrons in total. Furthermore, they defined a few additional models such

as MR_3-CISDT(18) and MR_3^{+T}-CISD(18). The former contains Slater determinants with triple excitations into the virtual spinor space, whereas the latter contains three holes in the Th $(6s, 6p)$ and O $(2s, 2p)$ sub-valence spinors.

Fleig et al. [27], in their calculations, first establish the molecular wave functions that describe accurately the excitation energy of the $\Omega = 1$ electronic state. The experimental internuclear distance of $R = 3.477\, a_0$ was used for their calculations. Basis sets of different qualities were considered, and it was found that the vtz basis set leads to a large correction of $-15\,\%$, whereas the vqz set only yields another $-4\,\%$ i.e., less than $200\,\mathrm{cm}^{-1}$ in absolute value. Hence, their investigation proceeded with the vtz basis set. The next criterion considered is the number of electronic shells included in the explicit treatment of dynamic electron correlation. For that, cases starting from 2 correlated electrons (Th $(7s, 6d)$ shells) were considered and went up to 38 electrons (Th $(5s, 5p, 5d, 6s, 6p, 6d, 7s)$ and O $(2s, 2p)$). It was observed that the correlation among the valence and sub-valence electrons of both atoms are quite important as compared to the core-valence and core-core correlations, where the latter changes the excitation energy by only $-3\,\%$. The third step was to consider the size and structure of the active space. It was observed that increasing the size of the active space leads to non-negligible corrections. In particular, the active spaces arising from $X = 7$ to $X = 10$ provide important correlation effects. A very accurate vertical excitation energy of $T_e = 5{,}410\,\mathrm{cm}^{-}1$ was obtained, as compared to the experimental result $5{,}317\,\mathrm{cm}^{-}1$, for the $\Omega = 1$ state, from their CI wave function model MR_{12}-CISD(18). Having studied the quality of the wave function in describing the excited state ($\Omega = 1$), which is relevant for the measurement of electron EDM, the next goal was to determine the effective electric E_{eff} and the parallel component of the magnetic hyperfine constant $A_\parallel$ for this state.

This effective one-body form of the electron EDM operator has been used by Fleig et al. [27] in their calculation of E_{eff} for the ThO molecule.

Similar analysis was performed for E_{eff} and $A_\parallel$ as was performed for the excitation energy T_e. It was found that E_{eff} is virtually insensitive to the size of the basis set employed. The hyperfine constant $A_\parallel$ changes by hardly 1 % in magnitude, as the basis set cardinal number is increased from 2 to 4. Furthermore, it was found that both E_{eff} and $A_\parallel$ are largely unaffected when more than 18 valence/sub-valence electrons are included in the explicit treatment of electron correlation. The final results were $E_{\mathrm{eff}} = 75.2\,\mathrm{GV/cm}$ and the $A_\parallel = -1{,}339\,\mathrm{MHz}$ in the $(\Omega = 1)$ excited state of ThO, using the CI wave function MR_{12}-CISD(18). The value of E_{eff} thus obtained is 10% smaller than the result previously reported by Skripnikov et al. [25]. From their detailed analysis, the uncertainty (error bar) in their calculation for the computed quantity E_{eff} is estimated to be around 3 %, which is quite small as compared to the error bars in the previous studies [24, 25]. In addition to that, the newly computed E_{eff} (which is 10 % smaller in magnitude) will have consequences for the recently determined upper bound on the eEDM [7], which had been obtained based on the earlier reported result $E_{\mathrm{eff}} = 84.0\,\mathrm{GV/cm}$ [25]. It is because of this 10 % smaller value of the E_{eff} that the upper bound of the eEDM d_e requires adjustment, with the revised bound being $|d_e| \leq 9.7 \times 10^{-29}\,\mathrm{e\,cm}$.

Relativistic Coupled-Cluster Method for E_{eff}

General Aspects

In this section, we discuss a relativistic coupled-cluster (RCC) method [28, 29], following the relativistic CI method. The single-reference RCC method is considered to be the current gold standard of many-body theory of heavy atoms and diatomic molecules, where the electronic state is mainly governed by a single determinant. For example, the RCC method is very effective and accurate in systems consisting of an alkali-earth-like atom and a halide atom, such as YbF and HgF. These molecules are also considered to be promising candidates for eEDM measurement, since the effective electric field, E_{eff}, is expected to be large, due to the high electronegativity of the halides. In this section, we briefly explain the theory of RCC and how to calculate E_{eff} using this method. We then discuss the results of E_{eff} in the YbF molecule with detailed calibrations of the dependence of basis sets and core correlation effects. To assess the accuracy of our RCC wave function, we also calculate the parallel component of the hyperfine coupling constant ($A_{\parallel}$) and the dipole moment (DM) along the molecular axis and compare them with the corresponding experimental values. From this assessment, we see that our RCC method is accurate and is within 10 % of error. Finally, we report on E_{eff} for mercury monohalide systems (HgF, HgCl, HgBr, and HgI). These systems are promising candidates for a new generation of EDM experiments, since their E_{eff} is more than 100 GV/cm, the largest values ever reported, and the sensitivity of measurement can be improved by generating cold molecules of mercury monohalides by photo association.

Theory of the Coupled-Cluster Method

The coupled-cluster wave function can be written as

$$|\psi\rangle = e^{\boldsymbol{T}}|\Phi_0\rangle \tag{49}$$

where $|\Phi_0\rangle$ refers to the DF wave function of the ground state of the molecule, which is built from single-particle four-component spinors. This is a single-reference CCM, that is, $|\Phi_0\rangle$ is taken to be a single determinant. $\boldsymbol{T}$ is the cluster operator, which is given by $\boldsymbol{T_1} + \boldsymbol{T_2}$, where $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ indicate the single (S) and double (D) excitation operators, respectively. This is called the CCSD (coupled-cluster singles and doubles) approximation. The excitation operators in CCSD are given by

$$\boldsymbol{T_1} = \sum_{ap} t_a^p a_p^\dagger a_a \tag{50}$$

$$\boldsymbol{T_2} = \sum_{p>q,a>b} t_{ab}^{pq} a_p^\dagger a_q^\dagger a_b a_a \tag{51}$$

t_a^p and t_{ab}^{pq} are the cluster amplitudes, with a and b referring to holes and with p and q referring to particles. If $a_p^\dagger a_a$ acts on a state, then a hole "a" is destroyed from that state, and a particle "p" is created. When $a_p^\dagger a_a$ acts on a model state, which is the Slater determinant, then the resulting state is denoted by $|\Phi_a^p\rangle$.

The CCSD amplitude equations are

$$\langle \Phi_a^p | e^{-\boldsymbol{T}} \boldsymbol{H_N} e^{\boldsymbol{T}} | \Phi_0 \rangle = 0 \tag{52}$$

$$\langle \Phi_{ab}^{pq} | e^{-\boldsymbol{T}} \boldsymbol{H_N} e^{\boldsymbol{T}} | \Phi_0 \rangle = 0 \tag{53}$$

The term $e^{-\boldsymbol{T}} \boldsymbol{H_N} e^{\boldsymbol{T}}$ can be written as $\{\boldsymbol{H_N} e^{\boldsymbol{T}}\}_C$, using the linked-cluster theorem [30], where the subscript "C" means that each term is connected and H_N is the normal-ordered Hamiltonian. The important features of our relativistic CCSD method are that it uses the Dirac-Coulomb approximation shown in Eq. (40), it is size extensive, and the correlation effects are treated to all orders in the residual Coulomb interaction for all possible single and double excitations [13]. In the coupled-cluster method, the expectation value of any operator, $\boldsymbol{O}$, is given by

$$\begin{aligned} \langle \boldsymbol{O} \rangle &= \frac{\langle \psi | \boldsymbol{O} | \psi \rangle}{\langle \psi | \psi \rangle} \\ &= \frac{\langle \Phi_0 | e^{\boldsymbol{T}} \dagger \boldsymbol{O} e^{\boldsymbol{T}} | \Phi_0 \rangle}{\langle \Phi_0 | e^{\boldsymbol{T}} \dagger e^{\boldsymbol{T}} | \Phi_0 \rangle} \\ &= \frac{\langle \Phi_0 | e^{\boldsymbol{T}} \dagger \boldsymbol{O_N} e^{\boldsymbol{T}} | \Phi_0 \rangle}{\langle \Phi_0 | e^{\boldsymbol{T}} \dagger e^{\boldsymbol{T}} | \Phi_0 \rangle} + \frac{\langle \Phi_0 | \boldsymbol{O} | \Phi_0 \rangle \langle \Phi_0 | e^{\boldsymbol{T}} \dagger e^{\boldsymbol{T}} | \Phi_0 \rangle}{\langle \Phi_0 | e^{\boldsymbol{T}} \dagger e^{\boldsymbol{T}} | \Phi_0 \rangle} \end{aligned} \tag{54}$$

The previous line uses the fact that $\boldsymbol{O} = \boldsymbol{O_N} + \langle \Phi_0 | \boldsymbol{O} | \Phi_0 \rangle$. Using the expression [31],

$$\frac{\langle \Phi_0 | e^{\boldsymbol{T}} \dagger \boldsymbol{O}_{\boldsymbol{N}} e^{\boldsymbol{T}} | \Phi_0 \rangle}{\langle \Phi_0 | e^{\boldsymbol{T}} \dagger e^{\boldsymbol{T}} | \Phi_0 \rangle} = \langle \Phi_0 | e^{\boldsymbol{T}^\dagger} \boldsymbol{O_N} e^{\boldsymbol{T}} | \Phi_0 \rangle_C \tag{55}$$

results in the expectation value of the operator $\boldsymbol{O}$ becoming

$$\langle \boldsymbol{O} \rangle = \langle \Phi_0 | e^{\boldsymbol{T}} \dagger \boldsymbol{O_N} e^{\boldsymbol{T}} | \Phi_0 \rangle_C + \langle \Phi_0 | \boldsymbol{O} | \Phi_0 \rangle \tag{56}$$

In the present work, the expectation value of a normal-ordered operator was calculated by considering only the linear terms in the coupled-cluster wave function as they make the most important contributions. The expectation value can, therefore, be expressed as

$$\langle \Phi_0 | (\boldsymbol{1} + \boldsymbol{T_1} + \boldsymbol{T_2})^\dagger (\boldsymbol{O_N}) (\boldsymbol{1} + \boldsymbol{T_1} + \boldsymbol{T_2}) | \Phi_0 \rangle_C + \langle \Phi_0 | \boldsymbol{O} | \Phi_0 | \tag{57}$$

An expectation value problem, such as the expression for E_{eff}, can be calculated using Eq. 57, where T_1 and T_2 amplitudes are determined from the CCSD amplitude equations. In the framework of coupled-cluster theory, the normal CC method [32] is the most widely used approach to calculate expectation values. However, this method requires substantially more computational time than the CC method described above. In addition, our approach captures the dominant correlation effects.

To perform RCC calculation for E_{eff}, two of the most widely used relativistic codes, REL4D in UTChem [33] and DIRAC08 [34], were combined and modified. UTChem was used for the generation of the DF orbitals and the molecular orbital (MO) integral transformations [35]. A suitable computational algorithm was developed in UTChem to evaluate one-electron integrals of the effective EDM Hamiltonian used in Eq. (41). We used the C_8 double group symmetry, since the adaptation of this symmetry drastically reduces the computational costs for both the MO transformation and the RCCSD calculations. Using the MO integrals, the DIRAC08 code was used to obtain the RCCSD amplitudes [13].

Calculation of E_{eff} in YbF

Dyall's four-component valence double-zeta (dz), triple-zeta (tz), and quadruple-zeta (qz) basis sets were used for ytterbium to check the basis set dependence in E_{eff} [36]. For fluorine, Watanabe's four-component basis sets [37] were employed. Some diffuse and polarization functions from the Sapporo basis sets [38] were used as well. All of the basis sets were used in their uncontracted form. This is because of the fact that although the DF results obtained from contracted as well as primitive basis sets match, while at the CCSD level, the former gives rise to large T_1 diagonostics, which indicate that the CC calculations are not stable. The basis sets used in the calculations are as follows: dz for Yb (24s, 19p, 13d, 8f, 1g), tz for Yb (30s, 24p, 18d, 14f, 3g, 2h), and qz for Yb (35s, 30p, 19d, 13f, 5g, 3h, 2i) and for F (13s, 10p, 4d, 3f). The total numbers of basis set orbitals for dz, tz, and qz are hence 270, 423, and 499, respectively. A cutoff value of 80 a.u. (atomic units) was imposed for the energies of the virtual molecular spinors in our CCSD calculations. The qz basis is the most accurate among the ones considered, with its accuracy confirmed by Gomes et al. [39]. The bond length and harmonic frequency they obtained with the qz basis were 2.0196 angstrom and 503.2 cm^{-1}, respectively. These results agree well with the experimental values, 2.0161 angstrom and 506.6674 cm^{-1} [40], and are also close to the extrapolated values, 2.0174 angstrom and 507.6 cm^{-1}, obtained from the results of dz, tz, and qz basis sets [39].

Another important aspect for accurate calculations is correlation space in electron correlation calculations. Here, three different relativistic CCSD calculations were performed, in order to understand the importance of core correlation effects, which are referred to as 49e-CCSD, 69e-CCSD, and 79e-CCSD. In the first case, 49 electrons are excited, that is, the $1s$, $2s$, $2p$, $3s$, $3p$, and $3d$ orbitals of Yb and the 1s orbital of F are frozen. In the second case, 69 electrons of YbF were excited, that is, the $1s$, $2s$, and $2p$ orbitals of Yb are frozen. In the last case, all 79 electrons of

Table 3 Summary of the calculated results of the present work

Basis set	Method	Total energy (a.u.)	$T_{1,\mathrm{dia}}$	E_{eff} (GV/cm)	DM (D)	$A_{\parallel}$ (MHz)
DZ	DF	−14,167.289602	–	17.9	3.21	–
TZ	DF	−14,167.321791	–	18.2	3.21	–
QZ	DF	−14,167.323266	–	18.2	3.21	6,239
DZ	49e-CCSD(197)	−14,169.344299	0.0432	21.4	3.37	–
TZ	49e-CCSD(255)	−14,169.899971	0.0588	21.1	3.46	–
QZ	49e-CCSD(293)	−14,170.080575	0.0397	22.7	3.59	–
QZ	49e-CCSD(303)	−14,170.026999	0.0339	22.8	3.59	–
QZ	69e-CCSD(293)	−14,170.501826	0.0334	23.1	3.60	–
QZ	79e-CCSD(293)	−14,170.522807	0.0311	23.1	3.60	7,913
Exp.	–	–	–	–	3.91(4)	7,424(81)

YbF are excited. The experimental value of bond length, which is 2.0161 angstrom, is chosen for the calculations.

Table 3 summarizes the calculated results. At the DF level, the values of E_{eff} and DM for the three basis sets are very close. However, at the CCSD level, there is a basis set dependence for the 49-e calculations. For the tz basis, a relatively large value for T_1 diagnostic (0.0558) was obtained, which indicates instability of the single-reference calculations with the basis. Gomes et al. also encountered the same issue at the tz basis and discussed its influence on the spectroscopic constants [39]. In our qz basis calculation, the T_1 diagnostic is lower (0.0397) than that for the tz basis. Therefore, it is reasonable to assume that our qz basis provides reliable results for E_{eff} and DM. The difference in the values of DM for the three CCSD-qz (49e-CCSD, 69e-CCSD, and 79e-CCSD) calculations was negligible. However, the corresponding difference in the case of E_{eff} was significantly larger, especially in between the 49e-CCSD and 69e-CCSD levels. The values of E_{eff} and DM of 69e-CCSD are the same as those of 79e-CCSD to three significant figures, indicating that the results of 69e-CCSD are saturated. We also compared two types of space of the virtual orbitals (293 and 303) in the QZ-49e-CCSD calculations. The extension of the virtual orbital space produces a very small change in the value of E_{eff}.

In addition, the DF and the individual correlation contributions to the E_{eff} for 79e-CCSD has been calculated. In the calculation of the expectation values shown in Eq. 57, the bra state can be expanded as DF, the singly excited (S), or doubly excited (D) determinants, and the ket state can be similarly expanded as DF, S, or D. The value of E_{eff} can be broken down into nine contributions at the level of 79e-CCSD (23.1 GV/cm), which are listed in Table 4. The dominant contribution comes from the DF-DF term (i.e., the value of E_{eff} at the DF level). The S-DF and DF-S terms contribute the most to E_{eff} among other terms.

In summary, the best result obtained was 23.1 GV/cm for E_{eff} and 3.60 Debye for DM with the qz basis set. The value of DM was within 8 % percent of the experimental value (3.91(4) Debye). Both E_{eff} and DM depend on the mixing of orbitals of opposite parities. Another important property that has similarities with

Table 4 Contributions of the nine combination terms for E_{eff} at the level of 79e-CCSD

	DF	S	D
DF	18.16	3.14	0.00
S	3.14	−1.31	0.09
D	0.00	0.09	−0.17

Table 5 Calculated values of E_{eff} HFCC ($A_{\parallel}$) and DM in the previous and present works

Method	Ref	E_{eff} GV/cm	$A_{\parallel}$ MHz	DM Debye
Semi-empirical	Kozlov et al. [15]	26.1	–	–
GRECP-RASSCF	Titov et al. [17]	18.8	4,975	–
GRECP-RASSCF-effective operator	Mosyagin et al. [18]	24.9	8,000	–
UDF-unpaired	Parpia [20]	19.9	–	–
UDF-all	Parpia [20]	24.9	–	4.00
MBPT	Quiney et al. [21]	24.8	7,985	–
(DF + Core polarization)				
DF	Nayak et al. [43]	19.9	–	3.98
RASCI	Nayak et al. [43]	22.5	–	3.90
RASCI	Nayak et al. [43]	24.1	–	–
DF/qz	Present work	18.2	6,239	3.21
79e-CCSD/qz	Present work	23.1	7,913	3.60
Experiment	[40, 41]	–	7,424(81)	3.91(4)

E_{eff} is the hyperfine coupling constant (HFCC). This is due to the fact that both these quantities are sensitive to the behavior of the wave function in the region of nucleus. The parallel component of HFCC ($A_{\parallel}$) was calculated, based on Quiney's formulation [21]. $A_{\parallel}$ was obtained as 6,239 and 7,913 MHz at the DF-qz and 79e-CCSD-qz levels, respectively, with the corresponding experimental values being 7,822(5) MHz by Van Zee et al., in 1978 [41], and 7,424(81) MHz by Steimle et al., in 2007 [42]. Our best calculation of $A_{\parallel}$ is within 7 % of the latest experimental value (Table 5).

E_{eff} in YbF has been previously calculated by different methods. The various methods by which E_{eff} was calculated are summarized in Table 6. The earliest work by Titov et al. in 1996 [17] was based on the restricted-active-space self-consistent-field (RASSCF) method with the generalized relativistic effective core potential (GRECP), with which they obtained an E_{eff} of 18.8 GV/cm. This result was later improved to 24.9 GV/cm with an effective operator for core polarization by Mosyagin et al. in 1998 [18]. Another study by Kozlov in 1997 [15], which was based on a semiempirical method, calculated E_{eff} to be 26.1 GV/cm. A comparison with the four-component relativistic methods is necessary to assess the accuracy of the abovementioned approximations.

Within the framework of the four-component relativistic method, Parpia [20] calculated $E_{\text{eff}} = 19.9$ GV/cm from the unpaired spinor and obtained 24.9 GV/cm from all the occupied spinors at the unrestricted DF level in 1998. In the same year,

Table 6 DF and CCSD results of E_{eff} in mercury monohalide systems

Molecule	Method	$T_{1,\mathbf{dia}}$	E_{eff} (GV/cm)
HgF	DF	–	104.25
HgCl	DF	–	103.57
HgBr	DF	–	97.89
HgI	DF	–	96.85
HgF	CCSD	0.0268	115.42
HgCl	CCSD	0.0239	113.56
HgBr	CCSD	0.0255	109.29
HgI	CCSD	0.0206	109.30

Quiney et al. [21] calculated $E_{\text{eff}} = 24.8\,\text{GV/cm}$ at the restricted DF level with the first-order core polarization taken into account using MBPT. In 2006 and 2009, Nayak et al. [43] accounted for electron correlation in E_{eff} using the restricted-active-space configuration interaction (RASCI) method at the four-component level and obtained $E_{\text{eff}} = 24.1\,\text{GV/cm}$ from a 31-electron correlated calculation in a space of 76 active orbitals. However, this active space is not sufficiently large. Our relativistic CCSD calculation has advantages over the relativistic CI approach used by Nayak et al. [43], because CCSD is size extensive unlike the approximate CI used in the latter work, the size of our qz basis is substantially larger than that used in Ref. [43], and all the core-correlation effects were included, while only 31 electrons were excited in the CI calculation. Also, the number of spinors in our virtual space (293 orbitals) is much larger than in Ref. [43] (60 orbitals).

Electron correlation increased the DF value of E_{eff} by about 20 % in our work. This trend is similar to the previous correlation calculations, especially where core-polarization effects were included.

Calculation of E_{eff} in Mercury Monohalides

Based on the experience of the detailed calculation of E_{eff} in YbF, the quantity was calculated for mercury monohalides. The mercury monohalides are expected to have larger values of E_{eff} than YbF molecule, because mercury is heavier than ytterbium. Since the present systems contain more number of electrons than in YbF, Dyall's c2v basis sets were chosen for Hg [44], and correlation-consistent polarized valence double-zeta (ccpvdz) basis sets were chosen for F, Cl, and Br [45], all in uncontracted form. For iodine, Dyall's basis (uncontracted) was chosen. Their details are as follows: Hg $(22s, 19p, 12d, 9f, 1g)$, F $(9s, 4p, 1d)$, Cl $(12s, 8p, 1d)$, Br $(14s, 11p, 6d)$, and I $(8s, 6p, 6d)$. No core orbitals were frozen, and no virtuals truncated in these calculations. The bond lengths (in angstrom) chosen for these calculations are HgF (2.00686) [46], HgCl (2.42), HgBr (2.62), and HgI (2.81) [47].

A summary of the calculations, both at the DF and the CCSD levels, is given in Table 7. The values for E_{eff} are very large for all of the chosen mercury halides: about one and a half times and about five times that of ThO [48] and YbF [13],

Table 7 Effective electric field, E_{eff}, in the HgF molecule

Work	E_{eff} (GV/cm)
Dmitriev et al. [49]	99.26
Meyer et al. [24]	68
This work	115.42

Table 8 Contributions of the nine combination terms for E_{eff} in HgF

	DF	S	D
DF	104.25	10.08	0.00
S	10.08	−3.91	0.22
D	0.00	0.22	−5.52

respectively. This is because there is strong mixing between the valence $6s$ and the virtual $6p_{1/2}$ orbital of the Hg atom and also the large value of the matrix element of the effective eEDM Hamiltonian between these two orbitals.

In the table below, our result for E_{eff} in HgF is compared with those available in the literature.

Dmitriev et al. computed E_{eff} for HgF using relativistic effective core potentials. They used five relativistic valence orbitals $5d_{3/2}$, $5d_{1/2}$, $6s_{1/2}$, $6p_{1/2}$, and $6p_{3/2}$ for Hg, and the minimal atomic basis set for F. Meyer et al. computed E_{eff} for HgF using a quasi-relativistic method and has compared their results with those obtained by other methods. The present work is the first application of an all-electron and fully relativistic method to mercury halides.

The contribution of the correlation effects is approximately 10 % of the effective electric field. The eight terms containing operators that are linear in T that contribute to the correlation show that there are cancellations between them, due to some terms being positive and others negative. This point is illustrated in Table 8, where the contributions of the individual terms to the expectation value in equation 57 are shown, for HgF. Among the correlation terms, the DF-S and the S-DF terms together contribute 20.16 GV/cm, while the S-S and the D-D terms together contribute −9.43 GV/cm. (These tendencies are similar to YbF.) The 9 correlation terms hence add up to 11.17 GV/cm. The DF-D and the D-DF terms are zero, due to the Slater-Condon rules and the fact that the H_{EDM} operator is a one-body operator. This is true, for example, also for the $H^{\text{eff}}_{\text{EDM}} T_1^2$ term.

The eEDM sensitivity of experiments is proportional to $E_{\text{eff}}\sqrt{N}$, where N is the number of molecules whose spin precession is detected. In order to achieve high sensitivity in experiment, a large value of both E_{eff} and N would be very desirable. Our calculations show a large value of E_{eff} for the mercury monohalides. In addition, they can be produced in large quantities at ultracold temperatures, e.g., by photo-association of laser-cooled Hg with magnetically trapped halogen atoms [50]. Hence, mercury monohalides can have both large values of E_{eff} and N and hence be attractive candidates for the next generation of eEDM experiments.

Summary

The first section explains why simultaneous violation of parity and time reversal symmetries are needed for the electron to have an electric dipole moment and the importance of eEDMs in the matter- antimatter asymmetry in the universe. It then explains in detail how atoms and molecules are used to get an upper limit on eEDM and why molecules are nowadays preferred over atoms. The second section describes the configuration interaction method and explains how one applies this method in the calculation of the effective electric fields of YbF and ThO molecules, respectively, along with other details such as the choice of basis and frozen cores. The third section introduces the reader to the current gold standard of electronic structure calculations, namely, the relativistic coupled cluster method, and how it is applied to calculate effective electric fields in YbF and the mercury monohalides, with the latter set of molecules showing great promise in the search for the eEDM.

References

1. Fukuyama T (2012) Searching for new physics beyond the standard model in electric dipole moment. Int J Mod Phys A 27(16):1230015
2. Ibrahim T, Itani A, Nath P (2014) Electron electric dipole moment as a sensitive probe of pev scale physics. Phys. Rev. D 90:055006
3. Nagashima Y (2014) Beyond the standard model of elementary particle physics. Wiley-VCH, Weinheim
4. Landau L (1957) On the conservation laws for weak interactions. Nucl Phys 3(1):127–131
5. Ballentine LE (1998) Quantum mechanics: a modern development. World Scientific Publishing, Singapore
6. Lüuders G (2000) Proof of the {TCP} theorem. Ann Phys 281(1–2):1004–1018
7. The ACME Collaboration, Baron J, Campbell WC, DeMille D, Doyle JM, Gabrielse G, Gurevich YV, Hess PW, Hutzler NR, Kirilov E, Kozyryev I, O'Leary BR, Panda CD, Parsons MF, Petrik ES, Spaun B, Vutha AC, West AD (2014) Order of magnitude smaller limit on the electric dipole moment of the electron. Science 343(6168):269–272
8. Hudson JJ, Kara DM, Smallman IJ, Sauer BE, Tarbutt MR, Hinds EA (2011) Improved measurement of the shape of the electron. Nature 473(7348):493–496
9. Kuzmin SV, Kazarian AM, Shaposhnikov ME (1992) Cosmological lower bound on the {EDM} of the electron. Phys Lett B 276(1–2):131–134
10. Nataraj HS (2009) Electric dipole moment of the electron and its implications on matter-antimatter asymmetry in the universe. PhD thesis
11. Sandars PGH (1967) Measurability of the proton electric dipole moment. Phys Rev Lett 19:1396–1398
12. Salpeter EE (1958) Some atomic effects of an electronic electric dipole moment. Phys Rev 112:1642–1648
13. Das BP (1989) Aspects of many-body effects in molecules and extended systems. p 411
14. Ostlund NS, Szabo A (1996) Modern quantum chemistry: introduction to advanced electronic structure theory, New edition edn. Dover Publications Inc.
15. Kozlov MG, Ezhov VF (1994) Enhancement of the electric dipole moment of the electron in the YbF molecule. Phys Rev A 49:4502–4507

16. Kozlov MG (1997) Enhancement of the electric dipole moment of the electron in the YbF molecule. J Phys B 30:L607
17. Titov AV, Mosyagin NS, Ezhov VF (1996) T-odd spin-rotational hamiltonian for YbF molecule. Phys Rev Lett 77:5346–5349
18. Mosyagin NS, Kozlov MG, Titov AV (1998) Electric dipole moment of the electron in the YbF molecule. J Phys B: At Mol Opt Phys 31:L763
19. Titov AV, Mosyagin NS, Petrov AN, Isaev TA, DeMille DP (2006) T-parity violation effects in polar heavy-atom molecules. Prog Theor Chem Phys 15:253
20. Parpia FA (1998) Ab initio calculation of the enhancement of the electric dipole moment of an electron in the YbF molecule. J Phys B: At Mol Opt Phys 31(7):1409
21. Skaane H, Quiney HM, Grant IP (1998) Hyperfine and PT-odd effects in YbF 2 sigma. J Phys B: At Mol Opt Phys 31:L85
22. Nayak MK, Chaudhuri RK (2007) Perturbative calculations of P, T-odd effects in heavy polar diatomic molecules. J Phys Conf Ser 80:012051
23. Nayak MK, Chaudhuri RK (2009) Re-appraisal of the P; T-odd interaction constant Wd in YbF: relativistic configuration interaction approach. Pramana J Phys 73:581
24. Meyer ER, Bohn JL (2008) Prospects for an electron electric dipole moment search in metastable ThO and ThF+. Phys Rev A 78:010502
25. Skripnikov LV, Petrov AN, Titov AV (2013) Theoretical study of ThO for the electron electric dipole moment search. J Chem Phys 139:221103
26. Huber KP, Herzberg G (1979) Molecular spectra and molecular structure: constants of diatomic molecules, vol 4. Van Nostrand, New York
27. Fleig T, Nayak MK (2014) Electron electric dipole moment and hyperfine interaction constants for ThO. J Mol Spectrosc 300:16
28. Kaldor U (1997) Atomic and molecular applications of the coupled cluster method, microscopic quantum many-body theories and their applications. In: Proceedings of a European summer school, Valencia, 8–19 Sept 1997, vol 510, pp 71–92
29. Nataraj HS, Sahoo BK, Das BP, Mukherjee D (2008) Intrinsic electric dipole moments of paramagnetic atoms: rubidium and cesium. Phys Rev Lett 101:033002
30. Bishop RF (1997) The coupled cluster method, microscopic quantum many-body theories and their applications. In: Proceedings of a European summer school, Valencia, 8–19 Sept 1997, vol 510, pp 1–70
31. Cizek J (1967) On the use of the cluster expansion and the technique of diagrams in calculations of correlation effects in atoms and molecules. In: Lefebvre WC (ed) Advances in chemical physics: correlation effects in atoms and molecules, Sep 08, 2009. Wiley, Hoboken
32. Arponen J (1983) Variational principles and linked-cluster exp S expansions for static and dynamic many-body problems. Ann Phys 151(2):311–382
33. Yanai T et al (2003) UTChem – a program for ab initio quantum chemistry. In: Sloot PMA et al (eds) ICCS 2003. LNCS 2660. Springer, Berlin/Heidelnerg, pp 84–95; Yanai T, Nakajima T, Ishikawa Y, Yanai T et al (2001) Computational science ICCS 2003:6526; Yanai T, Nakajima T, Ishikawa Y, Hirao K (2002) J Chem Phys 114:10122; Hirao K. J Chem Phys 116
34. Visscher L, Lee TJ, Dyall KG (1996) Formulation and implementation of a relativistic unrestricted coupled-cluster method including noniterative connected triples. J Chem Phys 105(19):8769–8776
35. Nakajima T, Abe M, Yanai T, Hirao K (2004) A 4-index transformation of Dirac's 4-component relativistic theory. Chem Phys Lett 388:68–73
36. Dyall KG (2009) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the 4s, 5s, 6s, and 7s elements. J Phys Chem A 113:12638
37. Watanabe Y, Tatewaki H, Koga T, Matsuoka O (2006) Relativistic gaussian basis sets for molecular calculations: fully optimized single-family exponent basis sets for H-Hg. J Comput Chem 27(1):48–52
38. Noro T, Sekiya M, Koga T (2012) Segmented contracted basis sets for atoms H through Xe: Sapporo-(dk)-nzp sets ($n = d, t, q$). Theor Chem Acc 131:1124

39. Gomes ASP, Dyall KG, Visscher L (2010) Relativistic double-zeta, triple-zeta, and quadruple-zeta basis sets for the lanthanides la–lu. Theor Chem Acc 127(4):369–381
40. Sauer BE, Cahn SB, Kozlov MG, Redgrave GD, Hinds EA (1999) Perturbed hyperfine doubling in the $A^2\Pi_{1/2}$ and [18.6]0.5 states of YbF. J Chem Phys 110(17):8424–8428
41. Van Zee RJ, Seely ML, DeVore TC, Weltner W (1978) Electron spin resonance of the ytterbium uoride molecule at 4 k. J Phys Chem 82(10):1192–1194
42. Steimle TC, Ma T, Linton C (2007) The hyperfine interaction in the A 2Pi1/2 and X 2Sigma+ states of ytterbium monofluoride. J Chem Phys 127(23):234316
43. Nayak MK, Chaudhuri RK (2006) Ab initio calculation of p,t-odd effects in ybf molecule. Chem Phys Lett 419(1–3):191–194
44. Dyall KG (1998) Relativistic and nonrelativistic finite nucleus optimized double zeta basis sets for the 4p, 5p and 6p elements. Theor Chem Acc 99:366–371; addendum (2002) 108:365; Relativistic quadruple-zeta and revised triple-zeta and double-zeta basis sets for the 4p, 5p, and 6p elements (2006) 115:441
45. Elsetha-gen T, Sun L, Gurumoorthi V, Chase J, Li J, Schuchardt KL, Didier BT, Windus TL (2007) Basis set exchange: a community database for computational sciences. J Chem Inf Model 47(3):1045–1052
46. van Meer R, Visscher L, Reiher M, Knecht S, Fux S, Saue T (2011) Mossbauer spectroscopy for heavy elements: a relativistic benchmark study of mercury. Theor Chem Acc 129(3–5):631650
47. Cheung Nh, Cool TA (1979) Franck-condon factors and r-centroids for the b 2 sigma - x 2 sigma systems of hgcl, hgbr and hgi. J Quant Spectrosc Radiat Transf 21(10):397–432
48. Fleig T, Nayak MK (2014) Electron electric dipole moment and hyperfine interaction constants for tho. J Mol Spectrosc 300:16–21
49. Dmitriev YuYu, Khait YuG, Kozlov MG, Labzovsky LN, Mitrushenkov AO, Shtoff AV, Titov AV (1992) Calculation of the spinrotational Hamiltonian including p- and p, t-odd weak interaction terms for hgf and pbf molecules. Phys Lett A 167(3):280–286
50. Doherty WG, Rennick CJ, Lam J, Softley TP (2014) Magnetic trapping of cold bromine atoms. Phys Rev Lett 112(2):023002

Relativistic Calculations of Atomic Clock

20

Bijaya Kumar Sahoo

Contents

Introduction 612
Classes of Clocks and Their Status 615
Operational Principles of Atomic Clocks 615
Other Possible Candidates 619
Typical Systematics 620
Instrumental Effects 621
External Effects 622
Relativistic Many-Body Methods 630
General Approach 631
Random-Phase Approximation (RPA) 638
Demonstration of Few Theoretical Results 639
A_{hf}-Constants and g_J-Factors 639
Static Polarizabilities (In Au) 641
Quadrupole Moments 643
Clocks for Fundamental Physics 647
Summary 651
References 652

Abstract

A pedagogic introduction to the atomic clock physics, mainly from a theoretical viewpoint, is presented, and the need for sophisticated relativistic many-body methods for their studies is emphasized. Few general aspects on the working principles of atomic clocks, their necessities in the daily life as well as for the fundamental sciences, and their present status are outlined. Special attention has been paid to keep the discussion at the graduate course level. Basic physics related to major systematics in an atomic system exposed to external

B.K. Sahoo (✉)
Theoretical Physics Division, Physical Research Laboratory (PRL), Ahmedabad, Gujarat, India
e-mail: bijaya@prl.res.in

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_30

electromagnetic fields and theoretical approaches for their accurate estimations for the atomic clock frequency measurements are highlighted. In this context, we discuss the roles of many-body methods to evaluate Stark, Zeeman, quadrupole, and multipolar blackbody radiation shifts and to search for magic wavelengths in the atomic systems. A few examples are given to show usefulness of theoretical perception for scrutinizing suitability of a candidate as an atomic clock a priori to performing measurement. Since the modern atomic clocks offer most precise measurements of the atomic transition frequencies, they are also used as tools for probing both temporal and spatial variation of many fundamental physical constants at the low-energy scale. The requirement of relativistic many-body methods to yield information on the temporal variation of fine-structure constant from the atomic clock studies is demonstrated. Several theoretical methods, capable of calculating atomic properties very accurately, are prescribed to facilitate better comprehension to the subject.

Keywords

Atomic clock • BBR shift (black-body radiation shift) • Configuration interaction (CI) • Coupled-cluster (CC) • Doppler effect • Fine structure constant • Frequency standard • Instability • Linewidth • Magic wavelengths • Micromotion • Microwave clock • Optical clock • Quadrupole shift • Quality factor • Random phase approximation (RPA) • Reproducibility • Secular motion • Signal-to-noise ratio • Stark shift • Zeeman shift

Introduction

An **atomic clock** is a device that uses resonance frequency of an electromagnetic transition between two energy levels in an atomic system as frequency standard to count ticks between two consecutive seconds. More than 50 years ago, Essen and Parry had proposed for the first time to use transition frequency between the hyperfine sublevels of the ground state in Cs atom to define unit of time. Today, this frequency is measured up to 9,192,631,770 oscillations per unit of time (known as second) and is conceded as primary **frequency standard**. This sort of accuracy is several orders higher than the accuracies of a typically used quartz clock in the daily life activities. The guiding cause for accrediting atomic transition frequencies as the time or frequency standards lies in their origin as they are the consequences of the fundamental interactions between the elementary particles; their values must remain the same irrespective of their locations or time of measurements. A selection criterion for a particular transition to be considered for this purpose depends mainly on three critical factors: a transition frequency that can be probed by a short-term stable local oscillator, the availability of a suitable frequency-counting mechanism, and a weak forbidden transition having intrinsic narrow natural **linewidth**. Broadly, atomic clocks are classified into either active or passive depending on whether the frequency standard is derived directly from the electromagnetic radiation or it is probed by an electromagnetic radiation of an external oscillator.

One of the quantities that signifies a better frequency standard is the **quality factor** (Q), defined as $Q = \frac{v}{\delta v}$ where ν and $\delta\nu$ are the transition frequency and linewidth of the transition, respectively. The ultimate objective would be to search for an atomic clock with Q-factor as large as possible. The accuracy of an atomic clock cannot be, in fact, defined uniquely. In practice, it quantifies either the offset of a frequency standard from its normal operating frequency or how well the frequency of the device can be related to the standard international (SI) unit of frequency. We shall return back to the estimation of uncertainties in the determination of the offset from its original value of the frequency in an atomic clock later. Another important aspect that also needs to be looked into before accepting candidates for atomic clocks is their **instability** and **reproducibility** limitations. Stability (as opposed to instability) is a measure of variability of the frequency standard over a specified period of time. Gauging this quantity requires a suitable reference standard having either lower or equally compatible stability with the considered atomic clock. All the three factors (reproducibility, accuracy, and instability) of an atomic frequency standard depend on the sensitivity of the reference frequency to the environmental perturbations and to the extent within which these perturbations can be controlled. The stability of the frequency standard is generally characterized by either two-sample or Allan variance given by [2]

$$\sigma_y^2(\tau) = \left\langle \frac{1}{2}\Sigma_k(\bar{y}_k - \bar{y}_{k+1})^2 \right\rangle, \tag{1}$$

where $\bar{y}_k$ is the mean fractional deviation frequency measured at time t_k with an interval dt_k from its nominal operating frequency ν_0 over the averaging period τ and reads

$$\bar{y}_k = \frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t_k)dt_k = \frac{1}{\tau}\int_{t_k}^{t_k+\tau} \frac{\nu(t_k)-\nu_0}{\nu_0}dt_k. \tag{2}$$

For better statistical accuracy, the average value needs to be evaluated by integrating over a large period of time τ. The final deviation is determined as the square root of the variance.

The present primary Cs clock is a **microwave clock**; however, most of the modern clocks are preferred to operate in the optical frequencies. The **optical clocks** operate at frequencies about five orders of magnitude higher than the Cs clock. The local reference in an optical clock is generally a narrow linewidth laser that is stabilized to a narrow optical atomic reference transition. Candidates chosen for the optical clocks can be grouped into two major classes: neutral atoms trapped and cooled using magnetic optical traps (MOT) and a single singly and multiply charged ion trapped using the Paul-type trap. The real state of the art for considering optical transitions as the frequency standards lies in their stabilities and to the precisions at which the uncertainties associated with their measurements can be elucidated. Instability in the fractional frequency shift due to the quantum fluctuations in an atomic absorption signal can be estimated by

$$\sigma_y(\tau) = \frac{n}{Q\ (S/N)}\tau^{1/2}, \tag{3}$$

where S/N is the **signal-to-noise ratio** for a 1 Hz detection bandwidth and ν value depends on the shape of the atomic resonance line and the method used to determine its central frequency. From the above expression, it can be understood that stability is a measure of the precision within which a given quantity can be measured and is usually stated as a function of averaging time. Owing to this fact, very significant reduction in the instability is gained in the optical clocks. The reduction factors depend exactly on the signal-to-noise ratio at which the atomic absorption signal is observed and proportional to $N^{1/2}$, where N is the number of atoms detected in the measurement process. Thus, the atomic clocks based on the optical lattices have relatively large stability compared with the microwave clocks as well as from a single ionized optical clock. The other concern in setting up an atomic clock is the uncertainties in the systematics, arising due to the sensitivity of an experiment to the environmental perturbations, which can shift the measured frequency from its unperturbed natural atomic frequency. It is, indeed, indispensable to expect suppressed systematics for an ideal frequency standard. Thus, it is imperative to understand all the physical processes happening surrounding the experiments and enable to estimate their systematics rigorously. It is also important that clock frequency can be reproducible irrespective of the location and time of an experiment. Contemplating both reproducibility and sensitivity to the environmental perturbations, there seems to be strong competition between the optical clocks based on the cold neutral atoms trapped in the optical lattices and a singly charged ion trapped using the Paul trap to replace the Cs microwave clock to become the next-generation primary frequency standard. In this scenario it is very much useful to have comprehensive understanding about the present status of different atomic clocks that are under consideration in different laboratories around the world and the roles of various environmental perturbations that are accountable for limiting the accuracies of the atomic clocks.

Among other prominent applications, atomic clocks are the essential components of the global positioning system (GPS). Each GPS satellite contains multiple atomic clocks. GPS receivers decode detected signals by synchronizing each receiver to the atomic clocks. In this process, it provides very precise time to an observed GPS signal and helps in finding out position and time of an object. It provides us the time to within 100 billionths of a second. Atomic clocks are also employed for synchronizing various signals from different instruments to carry out high-precision measurements. This is very important to differentiate tiny signatures that are observed in the high-resolution instruments like the measurements that are carried out using high-energy-based accelerators in the particle physics to probe any possible subtle effects governed by the fundamental interactions. In addition, these clocks are of immense use in the radio astronomy, telecommunications, meteorology, military services, etc. Besides, there are of intense investigations carried out for studying fundamental sciences directly using atomic clock signals. At the end of this chapter, a special section is devoted describing principles of probing variation in fundamental constants, especially the **fine-structure constant** α_e, using the atomic clock studies.

The main objective in this chapter would be to demonstrate the requirement of relativistic many-body methods in the context of studying physics related to atomic clocks. One may wonder how theoretical calculations in the considered atomic systems for the frequency standards could be useful at the standpoint where everything seems to be very precise. Note that most of the clock candidates are heavy atomic systems and performing ultrahigh precision calculations in these systems is insurmountable. Although the systematics of atomic clocks themselves are small but the physical quantities of atomic systems associated with these systematics are essentially large enough to be calculated accurately. All the physical quantities measured in the atomic clock experiments are undoubtedly very precise; hence, these observed results can also act as benchmarks for testing the capability of a (relativistic) many-body method to reproduce the experimentally measured quantities. Indeed, it requires powerful many-body methods for theoretical studies that are capable of taking care both the relativistic and electron correlation effects adequately. It is also worth mentioning that appropriateness of a candidate for an atomic clock can be prejudged from the knowledge of the calculated quantities a priori to actual measurement. Thus, it would be vital to understand differences and similarities among the theoretical methods used for the calculations to validate their reported results. In this regard, an attempt is made to formulate various many-body methods with the common starting point (reference state).

Classes of Clocks and Their Status

Operational Principles of Atomic Clocks

The advent of frequency comb is one of the major discoveries in the last century. It helps in stabilizing mode-locked femtosecond lasers to a very high degree and acts as a frequency divider connecting with frequencies of the entire range of the optical spectra. In fact, this can compare two atomic clocks hand in hand even falling under far apart range like optical to microwave frequencies. Frequency combs are used as common references to provide better stabilities to the optical clocks, which also helps to define accuracies of atomic clocks in an elegant manner. Confining atoms or ions for optical clocks in small regions under ultrahigh vacuum makes them well isolated from the environmental perturbations, and they produce narrow linewidths for the transitions as per the obligation. Howbeit, thermal motions of atoms and ions trapped using the electromagnetic fields in the optical clocks put tremendous challenges to reduce systematics during the clock frequency measurements. Also, operating the clocks at room temperature can give rise to a velocity distribution for the atoms causing Doppler shifts. In the Paul traps, the residual thermal motion and the **micromotion** of the ions can produce second-order Doppler shifts. Various sub-Doppler techniques have been invented recently to resolve atomic resonances much narrower than those of the Doppler distributions to eliminate **Doppler effects**. The advanced cooling and trapping techniques that make use of an array of lasers and external electric and/or magnetic fields are able to reduce these effects acutely. Techniques like the use of ultralow expansion glasses to measure quantum fluctuations on a fast time scale with a high S/N ratio have also been developed

to minimize many systematics. Depending on the choices of the candidates, special skills are being exploited to overcome these challenges. For example, it is argued that by confining microwave hyperfine transitions in the ground states of Cs and Rb atoms in an optical lattice generated by a circularly polarized laser field and by applying an external magnetic field with appropriately chosen direction may cancel out the dynamic Stark frequency shifts [3]. This may improve the accuracy in the frequency of Cs clock as in this case the clock transition is insensitive to the strengths of both the laser and external magnetic fields.

On the basis of ionic charges, general characteristics of optical clocks distinguish from each other and few of their working principles are discussed below case-wise by highlighting their advantages and drawbacks in constraining accuracies of the clocks.

Optical Lattice Clocks

Optical lattice clocks employ neutral atoms trapped in a specially engineered standing-wave laser fields, termed as magic trapping that uses magnetic optical trap (MOT) to form a lattice kind of potential structure with regular spacing. Its main advantage is that since the interactions between the neutral atoms are fairly short ranged, millions of atoms can be trapped and interrogated simultaneously. It extraordinarily improvises the stability of the clock owing to $\sqrt{N}$-factor offered by N-number of ensemble atoms trapped in the system. In this case, individual energy levels of a clock transition may be perturbed very strongly by the trapping fields; however, at the magic conditions both the clock levels are shifted almost identically. As a result, magic optical trapping potentials for clock transitions are defined for specific tailored trapping fields in which differential shifts of the clock transitions consequently vanish. The effects of optical laser trapping fields on energy levels are quantified using the ac **Stark shifts** that can be expressed in terms of the dynamic polarizabilities of the states and strengths of the applied electric fields. Even when the electric field oscillates, as in lasers, the differential Stark shift of a transition between two energy levels remains time independent. In this approach very narrow hyperfine-induced transitions, such as the $ns^2\ {}^1S_0 \rightarrow nsnp\ {}^3P_0$ transitions (n being the principal quantum number of the ground state) in the fermionic stable isotopes of alkaline earth elements (Be, Mg, Ca, Sr, Ba, and Ra atoms), can be exercised to make felicitous frequency standards. The analogous transitions from the rare-gas Zn, Cd, Hg, and Yb atoms can also fulfill the same errand. In fact, the atomic $ns^2\ {}^1S_0 \rightarrow nsnp\ {}^3P_0$ transitions of the corresponding bosonic stable isotopes are strictly forbidden by the angular momentum and parity selection rules. Applying small external magnetic fields, the excited $nsnp\ {}^3P_0$ states of these transitions can mix with their almost-degenerate fine-structure $nsnp\ {}^3P_1$ states. Under this condition $nsnp\ {}^3P_0$ the states can gain finite lifetimes and can be used as the interrogation time when the $ns^2\ {}^1S_0 \rightarrow nsnp\ {}^3P_0$ transitions are considered for frequency standards. These transitions thrive as perfect atomic clocks because of their contrived linewidths which can be manipulated to yield optimum stability by selecting the strengths of the applied magnetic fields. Thus far the most stable clock frequency of the above hyperfine-induced transition in ^{87}Sr has been measured

Table 1 Comparative analysis of important clock properties among few leading optical clocks

System	Clock transition	λ (in nm)	Natural linewidth (in Hz)	Observed linewidth (in Hz)	Fractional uncertainty (in $\times 10^{-15}$)	Instability (in $10^{-15}\tau^{-1/2}$)
Atoms						
^{40}Ca	$^1S_0 \leftrightarrow ^3P_1$	657	375.0	250.0	4.2	1.8
^{87}Sr	$^1S_0 \leftrightarrow ^3P_0$	698	0.001	0.5	0.006	0.4
^{171}Yb	$^1S_0 \leftrightarrow ^3P_0$	578	0.01	1.0	0.36	1.5
^{174}Yb	$^1S_0 \leftrightarrow ^3P_0$	578	0.0	4.0	1.5	5.5
^{199}Hg	$^1S_0 \leftrightarrow ^3P_0$	266	0.1	11.0	5,000	5.4
^{201}Hg	$^1S_0 \leftrightarrow ^3P_0$	266	0.1	11.0	5,000	5.4
Ions						
^{27}Al$^+$	$^1S_0 \leftrightarrow ^3P_0$	267	0.008	2.7	0.0086	0.6
^{40}Ca$^+$	$S_{1/2} \leftrightarrow D_{5/2}$	729	0.2	30.0	2.4	
^{88}Sr$^+$	$S_{1/2} \leftrightarrow D_{5/2}$	674	0.4	5.0	0.021	1.1
^{171}Yb$^+$	$S_{1/2} \leftrightarrow D_{3/2}$	436	3.1	10.0	0.45	
	$S_{1/2} \leftrightarrow F_{7/2}$	467	10^{-9}	7.0	0.071	
^{199}Hg$^+$	$S_{1/2} \leftrightarrow D_{5/2}$	282	1.7	6.7	0.019	

within 3×10^{-17} fractional uncertainty and a stability of better than $0.4 \times 10^{-15}\tau^{-1/2}$ [2,4,5].

A synchronous frequency comparison has been demonstrated of two optical lattice clocks using ^{87}Sr and ^{88}Sr atoms with the Allan standard deviation 1×10^{-17} in an averaging time of 1,600 s [2,4,5]. Similarly in ^{174}Yb, magnetically induced transition has been observed for a width of 5 Hz with the reported fractional uncertainty 1.7×10^{-15} and a stability better than $5.5 \times 10^{-16}\tau^{-1/2}$ [2, 4, 5]. Particularly, clock transitions from ^{199}Hg and ^{201}Hg exhibit small blackbody radiation shifts (**BBR shifts**), and their clock frequencies are observed within the fractional uncertainties to 5×10^{-12} [2,4,5]. In Table 1, we compare many pivotal properties of few leading atomic clocks accumulating from the review articles [2,4,5] to enliven about their adequacies and to explore about other better possible ways to boost their further advancements.

Singly Charged Ions

Although it appears as if use of a single singly charged trapped ion is a drawback in view of instability for an optical clock, but owing to developed competent principles to keep the ions isolated from the environmental perturbations makes them expedient contenders for the atomic clocks. Narrow transitions in several ions have been identified which are potentially pertinent for the optical frequency standards. Each of these ions has its own advantage and disadvantage in terms of clock transition parameters and intrinsic sensitivity to the environmental perturbations as well as in their technical complexities for running the corresponding experiments. The typical working scheme of a singly charged ion, presumed to be suitable for optical frequency standard, comprises a strong allowed transition for the laser cooling and a weak forbidden transition as the frequency reference.

The frequency is readily observed by detecting the resonance fluorescence from the strong cooling transition; however, detecting absorption of individual photon at the frequency of narrow reference transition is, in practice, not an easy affair. This is achieved by an electron shelving technique, whereby quantum jump in the cooling laser fluorescence signal is observed by driving the ion to the upper state of the reference transition by the clock laser. This trick empowers the narrow reference transition to be detected within incredible efficiency. The line profile of a clock transition can be built up by measuring quantum jump probability as a function of the probe laser frequency. The probe laser can be stabilized to the reference transition by repeatedly stepping its frequency back and forth between two estimated half-maximum intensity points of the resonance curve and monitoring the quantum jump rate imbalance between these two points. This quantum jump imbalance can provide a frequency discriminant from which a correcting steer to the frequency of the probe laser can be derived. Some of the prominent and routinely used ions are $^{27}Al^+$, $^{40}Ca^+$, $^{43}Ca^+$, $^{87}Sr^+$, $^{88}Sr^+$, $^{115}In^+$, $^{171}Yb^+$, $^{173}Yb^+$, $^{199}Hg^+$, and $^{201}Hg^+$ in which a lot of developments are carried out toward frequency standards. Few important key parameters of these standards are already compared in Table 1. Clock transitions from the alkaline earth ions, Yb^+ and Hg^+, involve ground S-states and one of the metastable $^2D_{3/2,5/2}$ states that lie below the $^2P_{1/2,3/2}$ states. Therefore, these D-states decay to their ground S-states via the electric-quadrupole transitions with natural linewidths in the range of 0.2–3 Hz, while the $P_{1/2}$ to $S_{1/2}$ transitions are used for cooling the ions. The highest observed Q-factor 1.6×10^{14} is achieved for $^{199}Hg^+$ with linewidth of 6.7 Hz [2, 4, 5]. The other important set of ions are $^{27}Al^+$ and $^{115}In^+$ having electronic configurations similar to the alkaline earth elements; their $ns^2\ ^1S_0 \rightarrow nsnp\ ^3P_0$ transitions are analogous to the previously discussed lattice clock transitions, which are also considered for frequency standards. Mixing of the $nsn\ ^3P_0$ states with the $nsnp\ ^3P_1$ and $nsnp\ ^1P_1$ states due to their hyperfine interactions allows to measure the transition frequencies of the above transitions in sufficiently finite time. These transitions are ideal for the frequency standards because of their longer stability and for exhibiting low systematic frequency shifts. A direct comparison of measurement between the $^{199}Hg^+$ and $^{27}Al^+$ frequency standards has been carried out with a relative uncertainty in their ratio as 5.2×10^{-17} for the total systematic uncertainties 1.9×10^{-17} and 2.3×10^{-17} in Hg^+ and Al^+, respectively, and for the frequency stability $4 \times 10^{-15}\tau^{-1/2}$ [4,5]. Further clock frequency of the $^1S_0 \leftrightarrow {}^3P_0$ transition in the same group Al^+ ion is measured independently trapping with the Mg^+ and Be^+ ions with the achievement of a relative statistical measurement uncertainty of 7.0×10^{-18} improved with the Mg^+ ion for a relative stability of $2.8\times10^{-15}\tau^{-1/2}$ and a fractional frequency difference of 1.8×10^{-17} [2,4,5]. In other set of experiments, a frequency instability of $9\times10^{-15}\tau^{-1/2}$ with reproducibility at the 6×10^{-16} level for the $[4f^{14}]6s\ ^1S_0 \rightarrow [4f^{14}]5d\ ^2D_{3/2}$ clock transition in $^{171}Yb^+$ [2,4,5] and absolute frequency difference between the ^{87}Sr and $^{87}Sr^+$ clocks up to 2.8×10^{-17} [4, 5] have been accomplished. In a categorically different experiment with the octopole transition $[4f^{14}]6s\ ^1S_0 \rightarrow [4f^{13}]6S^2\ ^2F_{7/2}$ in $^{171}Yb^+$, which is insensitive to the field-induced transitions and for possessing very long lifetime of its metastable

excited state (about 6 years), a fractional uncertainty of 7.1×10^{-17} with quantum projection noise $\sigma_y(5{,}000\ s) = 6 \times 10^{-17}$ has been reported [4, 5]. Among many other proposed ions, theoretical analysis also projects that both the $S \leftrightarrow D_{3/2,5/2}$ transitions in the Ba^+ and Ra^+ ions are also capable for realizing them as very high accurate frequency standards [6–8].

Highly Charged Ions

There have been recent proposals [9] to consider single but a highly charged ion as new atomic clock for bespeaking only minuscule systematics encountered by such ions. This owes to their strikingly shrunk orbitals which get least affected by the external perturbations. Since the radiative transition matrix elements are proportional to power expression of the atomic radius, coupling of atomic states with the external fields scales down for increasing charge of an ion. This is an obviously favorable condition for becoming an atomic clock and is the key factor why it is believed that the highly charged ions can be the unrivaled clocks. In principle, highly charged ions can be loaded in the ion traps by employing sympathetic cooling techniques along with another ion species, say Be^+ ions [9]. At sufficiently low temperatures the rates of undesirable charge-exchange processes between two ionic species become negligible. In this scheme, long-range elastic Coulomb collisions with continually laser-cooled Be^+ ions can drive these highly charged ions with temperature down to mK. It is possible to co-trap heavy highly charged ions with relatively light ions of low ionic charge depending upon the ratio of ion charge to its mass. In fact heavier cooling species like Mg^+ can also be used instead of Be^+ to improvise mass matching, hence the cooling efficiency. It is found that the transitions among the fine-structure levels of the ground state in ions having $[4d^{10}]f^{12}$ electronic configuration, e.g., $[4d^{10}]f^{12}\ ^3H_6 \rightarrow [4d^{10}]f^{12}\ ^3F_4$ transition, are the most felicitous ions for constructing such kind of frequency standards [9].

The feasibility of using magnetic-dipole (M1)-induced hyperfine transitions in the highly charged ions has also been looked into for possible atomic clocks with exceptional accuracy on the basis of anticipated negligibly small blackbody radiation (BBR), quadratic **Zeeman**, ac Stark, and **quadrupole** shifts in these ions [10]. The advantage of these clocks over other hyperfine clock transitions is that wavelengths of the corresponding transitions fall within the optical domain, while most of the other considered ground state-based hyperfine clock transitions lie in the microwave spectral region.

Other Possible Candidates

In spite that there is a less scope to discuss elaborately and connect the need of relativistic calculations at present, a brief discussion on proposed prospective nuclear and molecular clocks is presented for the sake of completeness and to give a broad outlook on this topic to the general readers.

Nuclear Clock

Nuclear systems, where the most common three fundamental interactions strongly interplay each other, are very interesting from the point of view of studying many fundamental sciences; however, their spectroscopy was not yet explored for metrology purposes owing to strenuous procedure to keep bare atomic nucleus isolated. In 2003, Peik and Tamm came up with an idea [11] proposing to use nuclear transition in the $^{229}Th^{3+}$ ion for the frequency standard. Radioactive ^{229}Th nucleus has an isomeric nuclear state at the unusually low excitation energy of 3.5 ± 1.0 eV above the ground state and its lifetime is about two to four times smaller than the half-life of its nucleus. This energy is comparable with the excitation energies of outer electrons in the atomic shell of $^{229}Th^{3+}$. The corresponding transition wavelength lies in the range of available tunable laser sources using which its frequency can be measured very precisely. However, the measurement of the corresponding frequency may encounter two important limitations. The under-considered transition will experience a significant second-order differential Zeeman shift ($\sim 70\,\text{kHz}/mT^2$), and their electric-quadrupole transitions could restrict the clock linewidths to ~ 1 Hz. It has been demonstrated that a pair of stretched hyperfine states within the $5F_{5/2}$ electronic ground level of both the nuclear ground and isomeric manifolds in $^{229}Th^{3+}$ could provide a clock linewidth of $\leq 100\,\mu$Hz and can offer unprecedented systematic shifts [12].

Molecular Clock

It is also proposed fairly recently by Schiller et al. [13] after carrying out calculations of external-field shift coefficients and analyzing at least 11 systematic effects in the one-electron molecular H^{+2} and HD^+ ions that frequency measurements in their ro-vibrational transitions may reach up to 2×10^{-17} uncertainty at the room temperature. They also argued that by considering measurement of composite M + 1 transitions (M being the number of systematic effects to be canceled) in different wavelength ranges, one can reduce the external-field effects. Following this proposal, Karr was quick to estimate light shifts induced by the probe laser using his calculated values of transition amplitudes, differential dynamic polarizability, hyperfine-structure constants, and clock interrogation times for the states accessible by the two-photon and quadrupole transitions [14]. From the estimation of quadrupole and Zeeman shifts, he showed that light shift is the main limiting factor in the case of two-photon transitions for both the H^{+2} and HD^+ ions and gave an estimated accuracy level close to 5×10^{-16} in the best possible case. However, he suggested that quadrupole transitions could be better as promising clock transitions with the estimated accuracies reaching beyond 1×10^{-16}.

Typical Systematics

Systematics in the frequency standard measurements are categorically of two types. One of them is due to the construction of the instruments (defined as instrumental systematics), while the other one is subjected to expose the systems

to the external or stray electromagnetic fields (defined as external systematics). Generally, instrumental systematics depend on size, shape and construction of the mechanical devices, while systematics caused by the external forces are guided by the atomic state properties and the strengths of the applied fields. Thus, the latter systematics can be controlled by suitably choosing clock transitions and by optimizing strengths of the electromagnetic fields. In the theoretical calculation prospects, atomic properties that are relevant to the external systematics are of immense interest. A brief discussion on both types of systematics is outlined below.

Instrumental Effects

With atomic clocks operating in the optical lattices, millions of atoms are trapped and interrogated simultaneously. The mechanism of trapping atoms with lasers brings a seemingly insurmountable challenge, and most of their instrumental systematics depend on the configurations of the devices. One of the major systematics is due to the Doppler's shifts which are reduced by the Doppler cooling technique given by Hänsch and Schawlow. In contrast, issues related with ion clocks are technically different. It follows from Maxwell's equations that it is not possible to trap a charged particle using three-dimensional potential well; however, they can be confined in the Penning and Paul traps. In Penning traps, three-dimensional confinement is achieved by the combination of a strong static magnetic field and a quadrupolar electrostatic field. Since trapped ions are strongly perturbed by the applied fields in the Penning traps, such traps are incompatible for the optical frequency standards, and hence, the Paul traps are commonly used for clocks. In these traps the quadrupole potential seen by the ions is

$$\phi(r, z, t) = (Q_{\mathrm{dc}} + Q_{\mathrm{ac}} \cos \Omega t)(r^2 - 2z^2), \tag{4}$$

with the dc and ac components of the potential Q_{dc} and Q_{ac}, respectively, for the angular frequency Ω (known as micromotion frequency). Appropriately selecting values of Q_{dc}, Q_{ac}, and Ω, motions of the ions can be stabilized in both the radial and axial directions. Under these conditions the motions of the trapped ions can be separated into two parts, a driven oscillatory motion at the trap drive frequency (micromotion) and a slower motion associated with the time-averaged confining potential (**secular motion**). The characteristic frequency of the secular motion is an order of magnitude smaller than Ω. The typical depth of the Paul traps is of the order of 10 eV, which is sufficient enough to capture ions created by the electron impact ionization or photo-ionization of neutral atoms emitted from a hot oven. The ions are laser cooled within the traps and confined to a region with dimensions less than the wavelengths of the light used to probe the reference transitions of the optical clocks. This is referred to as Lamb–Dicke regime, where the first-order Doppler effect is completely eliminated. The micromotion of the ion leads to amplitude modulation of the cooling laser fluorescence via the Doppler effect and allows to be monitored using radio-frequency photon-correlation techniques. In this approach, micromotion

is minimized by applying small dc voltages to additional compensation electrodes and also the second-order Doppler shift is greatly reduced by confining the ion tightly to the center of the trap. It is also necessary to operate the trap under ultrahigh vacuum condition to reduce the collisional perturbations.

External Effects

In order to determine the overall uncertainty budgets for the optical frequency standards with trapped atomic systems, it is a requisite to consider the effects of environmental perturbations such as magnetic, trapping electric, and applied-light fields on the trapped atoms or ions. External magnetic fields interact with the magnetic moments of the atomic states leading to linear Zeeman shifts of the atomic transition frequencies. In cases like clock transitions in the odd isotopes of alkali-like ions, which have half-integral nuclear spin I, it is possible to select a magnetic field-independent $m_F = 0 \leftrightarrow m_F = 0$ component of the clock transition. These transitions can still experience second-order Zeeman shifts; however, they can be controlled at an acceptable precision level by operating the frequency standard in a low magnetic field of around $1\,mT$. In addition, the quadrupole shifts due to the interaction between the electric-field gradient and the atomic electric-quadrupole moments can also be eliminated by averaging out the shifts to the transition frequency over three mutually orthogonal magnetic-field orientations. Nonetheless, knowing the functional form of the perturbation due to the quadrupole shift is still very useful for conducting the experiment facilely. Stark shifts of the clock transition frequency in a trapped atom or ion can arise in a variety of ways. Firstly, the micromotion and the thermal motion of the atom or ion within the trap cause it to experience a nonzero root mean square value of the electric field. The magnitude and stability of this field depend on the stray charges present within the trap. With careful micromotion compensation, it should be possible to reduce these effects to few parts in 10^{18} level. Secondly, there will be a blackbody Stark shift due to the temperature of the apparatus surrounding the atom or ion. All the cooling, repumper, and clearing-out laser beams can be switched off, while the clock transition is being probed in order to minimize the surrounding effects, but the probe laser light has to be on throughout the experimentation. Thus, the Stark shifts can still be present, but they can be reduced by manipulating intensities required to drive the clock transitions. However, for the kind of laser linewidths achieved these days in the clock frequency measurements, high laser intensity is required to drive these transitions at reasonable rates. Hence, the transition frequencies are measured as functions of laser power and extrapolated to zero power to determine the actual value. Few important systematics will be discussed here that require auxiliary measurements to find out them or can be estimated by combining strengths of the applied electromagnetic fields with high-accuracy calculations of relevant physical properties of the transitions in the clock candidates. Of course, it is important to consider a suitable many-body method for such theoretical calculations. Particularly, at least four reasons can be cited to justify that it is enviable interest to carry

out theoretical studies of the above properties especially for the atomic clocks: (i) as has been demonstrated before, it is feasible to select an atomic system to check its viability for the frequency standard prospectus with accurate knowledge of atomic spectroscopy; (ii) when the experimental results are either not available or carrying out precise measurements is extremely strenuous, theoretical results can be their substitutes; (iii) performing high- precision calculations at the level of present interest can be much economical than setting up their auxiliary experiments; and (iv) comparison between the measurements and calculations of spectroscopic properties in a system obtained from the clock studies could serve as tool to assess the potential of the employed many-body methods.

Quadratic Zeeman Shift

In the presence of static magnetic field, the spectral lines of an atomic system can be split into several components as known from the Zeeman effect. In most of the atomic clocks, linear Zeeman shifts are almost canceled out, while the quadratic Zeeman shifts still contribute to the uncertainties. The interaction operator to determine the quadratic Zeeman shift due to the hyperfine interaction subjected to the external magnetic field $\vec{\boldsymbol{B}} = |B|\hat{z}$ is given by

$$H_B = A_{\rm hf}\vec{I}.\vec{J} + g_J\mu_B\vec{J}.\vec{B} + g_I\mu_B\vec{I}.\vec{B}, \tag{5}$$

where g_J is the gyromagnetic constant of the electronic state of angular momentum J, g_I is the gyromagnetic constant of the nuclear state with spin I, and μ_B is the Bohr magnetron. This shift can be estimated if the hyperfine constants and g-factors for both the electronic and nuclear components of the system known. $A_{\rm hf}$ values can be extracted by measuring hyperfine splitting; however, such procedure is complicated in a fermionic system with $I > 1/2$ when the second- or higher-order hyperfine interactions can be significantly contributing. Due to limited scope, only few examples of comparison between the experimental and theoretical results are provided later. Theoretically, $A_{\rm hf}$ of a state is calculated by

$$A_{\rm hf} = \frac{\mu_n}{I}\frac{\langle J \left\| H_{\rm hf}^{\rm mag} \right\| J\rangle}{\sqrt{J(J+1)(2J+1)}}, \tag{6}$$

for the nuclear magnetic moment $\mu_n = \mu_I\mu_B$ where μ_I is the nuclear moment in units of μ_B and $\langle J \left\| H_{\rm hf}^{\rm mag} \right\| J\rangle$ is the reduced matrix element of the electronic component of the magnetic-dipole hyperfine interaction Hamiltonian $H_{\rm hf}^{\rm mag}$.

Similarly, the g_J-factor can be measured precisely or can be evaluated using the expression

$$g_J\langle J\rangle = -\langle JJ|m_e c\sqrt{2}ir\{\vec{a}\otimes C^{(1)}\}^{(1)}|JJ\rangle \tag{7}$$

with the Dirac matrix $\vec{a}$ and Racah operator C.

Two other approaches are adopted to obtain more precise values of g_J-factors of the bound electrons. The expression given in Eq. (7) is basically for determining the Dirac g_J-factor of a bound electron analogous to $g_D = 2$ for a free electron. However, QED correction yields

$$g_D = 2\left[1 + \frac{\alpha_e}{2\pi} - 0.328\frac{\alpha_e^2}{\pi^2} + \cdots\right] \approx 2 \times (1.01160). \tag{8}$$

Accounting full QED effects for bound electrons in the atomic systems with more than three electrons is kind of intractable. The QED corrections, however, can be approximated roughly using the interaction Hamiltonian [15]

$$\delta g_J = 0.001160\, \beta \Sigma_0^{(1)}, \tag{9}$$

where β and $\Sigma_0^{(1)}$ are the Dirac matrix and the z-component of the Dirac spinor Σ, respectively.

It is also appropriate to estimate only the bound-state relativistic correction to g_J as [16]

$$\Delta g_J = \langle(\beta - 1)2J - \beta L\rangle, \tag{10}$$

for the orbital angular momentum operator L. Then, the final g_J-value can be determined by subtracting the above correction from the g_D-value.

Using a classical vector coupling model, g_J can also be estimated reasonably as

$$g_J = 1 + \frac{J(J+1) - l(l+1) + s(s+1)}{2j(j+1)}, \tag{11}$$

where l and s are the orbital and spin angular momentums, respectively. Similarly, g_I can be extracted from the knowledge of g_J- and g-factor of the hyperfine state (g_F) using the relation

$$\begin{aligned} g_F = {} & g_J \frac{F(F+1) + J(J+1) - I(I+1)}{2F(F+1)} \\ & + g_I \frac{F(F+1) - J(J+1) + I(I+1)}{2F(F+1)} \end{aligned} \tag{12}$$

for the I- and J-coupled hyperfine angular momentum F.

Quadratic Stark Shift

The change in the energy of an atomic state $|\Psi_n\rangle$ with angular momentum and its z-component of the state as j_n and m_j placed in an external weak electric field $\vec{E} = |E|\hat{z}$ is given by

$$\Delta E_n(j_n, m_j) \simeq -\frac{1}{2}\alpha_n^{E1}\ E^2 \tag{13}$$

where α_n^{E1} is the static electric dipole (E1) polarizability of the state. If the electric field is of ac type with altering frequency ω, it yields

$$\alpha_n^{E1} = -\Sigma_{k\neq n}|\langle\Psi_n|D|\Psi_k\rangle|^2 \times \left[\frac{1}{(E_n - E_k) + \omega} + \frac{1}{(E_n - E_k) - \omega}\right], \tag{14}$$

where D is the E1 operator and E_i is the energy of the $|\Psi_i\rangle$ state. The sum over k denotes contributions from all possible E1 transitions. Using the tensor product formalism, it becomes

$$\begin{aligned}\alpha_n^{E1} &= \alpha_n^{E1}(0) + A\ \cos\theta_k \frac{m_j}{j_n}\alpha_n^{E1}(1) \\ &+ \left\{\frac{3\cos^2\ \theta_p - 1}{2}\right\}\left\{\frac{3m_j^2 - j_n(j_n+1)}{j_n(2j_n-1)}\right\} a_n^{E1}(2).\end{aligned} \tag{15}$$

Here A, θ_k, and θ_p are the degree of circular polarization angle, angle between the wave vector of the electric field σ and the $\hat{z}$-axis, and angle between the direction of polarization and $\hat{z}$-axis, respectively. Here it should be noted that $A = 0$ for the linearly polarized light and $A = 1(-1)$ for the right(left)-handed circularly polarized light. In the absence of magnetic field, $\cos\theta_k = \cos\theta_p = 1$. In the above expression $\alpha_n^{E1}(0)$, $\alpha_n^{E1}(1)$, and $\alpha_n^{E1}(2)$ are known as scalar, vector, and tensor polarizabilities, respectively, and are usually defined in terms of m_j independent factors (reduced matrix elements) to evaluate them conveniently and given by

$$\begin{aligned}\alpha_n^{E1}(0) &= \frac{(\Sigma_q\ 1)(\Sigma_{m_j}\ 1)}{3(2j_n+1)}\Sigma_{j_k\neq j_n}|\langle J_n\|d\|J_k\rangle|^2 \\ &\times\left[\frac{1}{(E_n - E_k) + \omega} + \frac{1}{(E_n - E_k) - \omega}\right],\end{aligned} \tag{16}$$

$$\begin{aligned}\alpha_n^{E1}(1) &= -\frac{(\Sigma_q\ 1)(\Sigma_{m_j}\ 1)}{3(2j_n+1)}\sqrt{\frac{54\ j_n(2j_n+1)}{(j_n+1)}}\Sigma_{j_k\neq j_n}(-1)^{j_n+j_k+1}\left\{\begin{array}{ccc} j_n & 1 & j_n \\ 1 & j_k & 1\end{array}\right\} \\ &\times|\langle J_n\|d\|J_k\rangle|^2 \times\left[\frac{1}{(E_n - E_k) + \omega} - \frac{1}{(E_n - E_k) - \omega}\right],\end{aligned} \tag{17}$$

and

$$\alpha_n^{E1}(2) = -2\frac{(\Sigma_q\ 1)(\Sigma_{m_j}\ 1)}{3(2j_n+1)}\sqrt{\frac{15\ j_n(2j_n+1)(2j_n-1)}{2(j_n+1)(2j_n+1)}}$$
$$\Sigma_{j_k\neq j_n}(-1)^{j_n+j_k+1}\begin{Bmatrix} j_n & 2 & j_n \\ 1 & j_k & 1 \end{Bmatrix}$$
$$\times|\langle J_n\|d\|J_k\rangle|^2\times\left[\frac{1}{(E_n-E_k)+\omega}+\frac{1}{(E_n-E_k)-\omega}\right]. \tag{18}$$

Here q corresponds to the z-component of rank of D. It is obvious that for the static (dc) electric fields, contributions due to the vector polarizability vanishes and also contributions from both $\alpha_n^{E1}(1)$ and $\alpha_n^{E1}(2)$ to the closed-shell configurations nullify.

In the hyperfine $|I\ j_n F_n m_F\rangle$ state the Stark shift is given, analogously, by

$$\Delta E_n(F_n, m_F) \simeq -\frac{1}{2}\alpha_{n,F}^{E1}\ E^2, \tag{19}$$

where $\alpha_{n,F}^{E1}$ is the corresponding dipole polarizability and can also be expressed in terms of scalar, vector, and tensor components as before. Since it is hindered to work with the hyperfine states for practicality, the hyperfine state polarizabilities are expressed in terms of the atomic polarizabilities by relating $\alpha_{n,F}^{E1}$ with α_n^{E1} in the IJ-coupling approximation as

$$\alpha_{n,F}^{E1}(0) = \alpha_n^{E1}(0), \tag{20}$$

$$\alpha_{n,F}^{E1}(1) = (-1)^{j_n+I+F}\sqrt{\frac{4\ F_n\ (2F_n+1)(2j_n+1)(j_n+1)}{9(F_n+1)j_n}}\begin{Bmatrix} F_n & J_n & I \\ j_n & F_n & 1 \end{Bmatrix}\alpha_n^{E1}(1) \tag{21}$$

and

$$\alpha_{n,F}^{E1}(2) = (-1)^{j_n+I+F}\sqrt{\frac{F_n\ (2F_n-1)(2F_n+1)(2j_n+3)(2j_n+1)(j_n+1)}{(2F_n+3)(F_n+1)j_n(2j_n-1)}}$$
$$\begin{Bmatrix} F_n & j_n & I \\ j_n & F_n & 2 \end{Bmatrix}\alpha_n^{E1}(2). \tag{22}$$

So with accurate knowledge of the atomic polarizabilities and strengths of the applied electric fields, Stark shifts for the frequency standards can be deduced.

Electric-Quadrupole Shift

Electric-quadrupole shifts occur due to interaction between the atomic electric-quadrupole moments with the external electric-field gradients that are generated by the electrodes of the trapping systems. The Hamiltonian describing the interaction of external electric-field gradient with the atomic quadrupole moment of an atomic state is given by

$$H_Q = \vec{\nabla} E^{(2)} \cdot \vec{\Theta}^{(2)} = \Sigma_{q=-2}^{2} (-1)^q \nabla E_q^{(2)} \Theta_{-q}^{(2)}, \tag{23}$$

where $\vec{\nabla} E^{(2)}$ is the tensor describing the gradient of the external electric field at the position of the atom and $\vec{\Theta}^{(2)}$ is the electric-quadrupole operator. In the principal-axis frame H_Q reads out

$$H_Q = -2\ A\ \Theta_0^{(2)'} + \sqrt{\frac{2}{3}} \epsilon A (\Theta_2^{(2)\prime} + \Theta_{-2}^{(2)\prime}). \tag{24}$$

The first-order correction due to H_Q in the hyperfine state $|Ij_nF_nm_F\rangle$ is given by

$$\begin{aligned} \Delta E_Q &= \langle Ij_nF_nm_F|H_Q|Ij_nF_nm_F\rangle \\ &= \frac{-2[3M_F^2 - F(F+1)]A\langle Ij_nF_n\|\Theta^{(2)}\|Ij_nF_n}{[(2F_n+3(2F_n+2)(2F_n+1)2F_n(2F_n-1)]^{1/2}} \\ &\quad \times [(3\cos^2\beta - 1) - \epsilon \sin^2\beta(\cos^2\alpha - \sin^2\alpha)], \end{aligned} \tag{25}$$

where α, β, and γ are the standard Euler angles that convert the actual principal-axis frame to the working laboratory frame (taking $\gamma = 0$), ϵ is known as asymmetry parameter and A is the strength of the field gradient of the applied direct current (dc) voltage. The reduced matrix element in the above expression is given by

$$\langle Ij_nF_n|\Theta^{(2)}||Ij_nF_n\rangle = (-1)^{I+j_n+F_n}(2F_n+1) \begin{Bmatrix} j_n & 2 & j_n \\ F_n & I & F_n \end{Bmatrix} \begin{pmatrix} j_n & 2 & j_n \\ -j_n & 0 & j_n \end{pmatrix}^2 \Theta(j_n). \tag{26}$$

for the atomic quadrupole moment $\Theta(j_n)$. By knowing accurate $\Theta(j_n)$ values of the atomic states involved in a clock transition, differential quadrupole shift of a transition can be obtained. The quadrupole moment of an atomic level with angular momentum j_n is evaluated by

$$\Theta(j_n) = \langle j_n j_n|\Theta_0^{(2)}|j_n j_n\rangle \tag{27}$$

with $\Theta_0^{(2)} = \frac{e}{2}\Sigma(3z^2 - r^2)$. In an experiment, $\Theta(j_n)$ is measured by altering static dc voltage and is a challenge to extract precisely. Several calculations of this quantity are also available with accuracies as par with the measurements and even in

some cases accuracies in the calculations have surpassed over their corresponding measurements.

Multipolar BBR Shifts

Shifts in the quantum energy levels owing to the atom residing in an environment at finite temperature at which an atom can emit radiations with wavelengths depending on the magnitude of temperature are known as BBR shift. Atomic clocks are advised to operate at the room temperature for the general use and the wavelengths emitted from the atoms at this temperature are typically in the infrared region of the electromagnetic spectrum. Accurate estimates of BBR shifts for the laboratory as well as those used in the spacecraft atomic clocks are inevitable as even in the most remote place of intergalactic space, an isolated atom is still subject to cosmic microwave background radiation (CMBR). Uncertainties in the estimated BBR shifts can impose limits to achieve the best atomic clocks. These shifts can be estimated using the multipole expansion of the electromagnetic field. Usually contribution due to the E1 channel is taken into account for its dominant contribution; however, contributions from other higher multipoles, especially from the M1 and E2 channels, can also be significant when the accuracy of clocks reach below 10^{-19} level. A general theory of BBR shift follows below.

The Hamiltonian describing interactions between the electrons in an atomic system with the external propagating electromagnetic field in the Coulomb gauge coupling is given by

$$V(r,\omega) = -c\vec{\alpha}\cdot\vec{A}(r,\omega) = -c(\vec{\alpha}\cdot\vec{\epsilon})e^{i\vec{k}\cdot\vec{r}}, \tag{28}$$

where ω is the angular frequency of the field and $\vec{k} = |k|\hat{k}$ and $\vec{\epsilon}$ are its wave vector and polarization direction, respectively. The expression for the BBR shift of an atomic energy level $|\Psi_n\rangle$ with energy E_n is given by

$$\Delta E_{\rm BBR} = \frac{1}{2}\Sigma_{m,\omega}|V_{nm}(r,\omega)|^2\left(\frac{E_n - E_m}{(E_n - E_m)^2 - \omega^2}\right), \tag{29}$$

for $V_{nm}(r,\omega)$ being the transition matrix of $V(r,\omega)$ between the states $|\Psi_n\rangle$ and $|\Psi_m\rangle$.

Carrying out multipolar expansion of $V(r,\omega)$ in terms of general moments $Q^{\lambda}_{LM}(\vec{k}\cdot\vec{r})$, it can give

$$\begin{aligned}(\vec{\alpha}\cdot\vec{\epsilon})e^{i\vec{k}\cdot\vec{r}} &= -\Sigma_{LM}\frac{(K^L)(i^{L+1+\lambda})}{(2L+1)!!}[\vec{Y}^{\lambda}_{LM}(\hat{k})\cdot\vec{\epsilon})]\sqrt{\frac{4\pi(2L+1)(L+1)}{L}}Q^{\lambda}_{LM}(\vec{k}\cdot\vec{r}) \\ &= -\Sigma_{LM,l}\frac{(K^L)(i^{L+1+\lambda})}{(2L+1)!!}Y^{\lambda}_{LM}(k_l)\sqrt{\frac{4\pi(2L+1)(L+1)}{L}}Q^{\lambda}_{LM}(r_l),\end{aligned} \tag{30}$$

where k_l is the component of $\vec{k}$ projecting toward the lth unit vector of $\vec{\epsilon}$, and $\lambda = 1$ and $\lambda = 0$ correspond to the electric and magnetic multipoles, respectively. Since emission from BBR is isotropic, each component of the electric and magnetic fields is related to the spectral energy density as

$$u(\omega, T) = \frac{3}{8\pi} E_l^2(\omega) = \frac{3}{8\pi} B_l^2(\omega) = \frac{1}{\pi^2 c^3} \frac{\omega^3}{e^{\omega/k_B T} - 1}. \tag{31}$$

Hence, averaging over polarizations and propagation directions in Eq. (29), we get [17, 18]

$$\Delta E_{\mathrm{BBR}}^{\lambda,L} = -\frac{\alpha k^B T}{2j_n + 1} \Sigma_{\substack{m \neq n \\ L,M}} \left| \langle \Psi_n \| Q_{LM}^{\lambda} \| \Psi_m \rangle \right|^2 \times F_L \left(\frac{E_n - E_m}{k_B T} \right), \tag{32}$$

where the universal function is defined as $F_L(y) = \frac{1}{\pi} \frac{L+1}{L(2L+1)!!(2L-1)!!} \times \int_0^\infty dx \left(\frac{1}{y+x} + \frac{1}{y-x} \right) \frac{x^{2L+1}}{e^x - 1}$ that is applicable to all atoms with argument y depending on the range of the atomic parameters. With $|y| \gg 1$, corresponding to the transition energy much larger than the $k_B T$ values as of our interest and for the dominant term $L = 1$, we have

$$\Delta E_{\mathrm{BBR}}^{\lambda} = -\frac{1}{2c_n^{Q_L^{\lambda}}} \left(\frac{8\pi^3 \alpha_e^3 (k_B T)^4}{45(2j_n + 1)} \right) \alpha_n^{Q_L^{\lambda}}(0), \tag{33}$$

where α_e is the fine-structure constant and $\alpha_n^{Q_L^{\lambda}}(0)$ is the scalar polarizability defined as

$$\alpha_n^{Q_L^{\lambda}}(0) = C_n^{Q_L^{\lambda}} \Sigma_{m \neq n} \frac{\left| \langle \Psi_n \| Q_{LM}^{\lambda} \| \Psi_m \rangle \right|^2}{E_n - E_m}, \tag{34}$$

with the appropriate angular coefficient $C_n^{Q_L^{\lambda}} = \frac{2}{\alpha_e^{2(\lambda-1)}(2L+1)(2j_n+1)}$ due to the radiative moment Q_L^{λ}. Usually large contribution comes from E1, followed by M1, then E2, and so on. On average over polarization, the BBR shifts from the first three important channels are given by

$$\Delta E_{\mathrm{BBR}}^{E1} = -\frac{1}{2} \frac{4\pi^3 \alpha_e^3}{15} (k_B T)^4 \alpha_n^{E1}(0) = -\frac{1}{2} \langle E_{E1}^2(\omega) \rangle \alpha_n^{E1}(0), \tag{35}$$

$$\Delta E_{\mathrm{BBR}}^{M1} = -\frac{1}{2} \frac{4\pi^3 \alpha_e^5}{15} (k_B T)^4 \alpha_n^{M1}(0) = -\frac{1}{2} \alpha_e^2 \langle B_{M1}^2(\omega) \rangle \alpha_n^{M1}(0) \tag{36}$$

and

$$\Delta E^{E2}_{\rm BBR} = -\frac{1}{2}\frac{8(\alpha\pi)^5}{189}(k_B T)^6 \alpha_n^{E2}(0) = -\frac{1}{2}\langle E^2_{E2}(\omega)\rangle \alpha_n^{E2}(0), \tag{37}$$

where $\alpha_n^{M1}(0)$ and $\alpha_n^{E2}(0)$ are the scalar M1 and E2 polarizabilities, respectively, and $\langle E^2_{E1}(\omega)\rangle$, B^2_{M1}, and $\langle E^2_{E2}(\omega)\rangle$ are the averaged E1-induced electric, M1- induced magnetic, and E2-induced electric fields, respectively. Conventionally, the above shifts are expressed as functions of temperature scaled with respect to the room temperature ($T = 300\,^\circ$K) as

$$\Delta E^{E1}_{\rm BBR}(300\,^\circ{\rm K}) = -\frac{1}{2}\left(831.9\frac{V}{m}\right)^2\left[\frac{T(K)}{300}\right]^4 \alpha_n^{E1}(0), \tag{38}$$

$$\Delta E^{M1}_{\rm BBR}(300\,^\circ{\rm K}) = -\frac{1}{2}(2.77\times 10^{-6}{\rm Tesla})^2\left[\frac{T(K)}{300}\right]^4 \alpha_n^{M1}(0), \tag{39}$$

and

$$\Delta E^{E2}_{\rm BBR}(300\,^\circ{\rm K}) = -\frac{1}{2}\left(7.2\times 10^{-3}\frac{V}{m}\right)^2\left[\frac{T(K)}{300}\right]^6 \alpha_n^{E2}(0). \tag{40}$$

Even for estimating the E1 BBR shift, the above formula does not take into account frequency distribution properly. Including the appropriate dynamic correction, the final expression for the BBR shift due to the E1 component yields

$$\Delta E^{E1}_{\rm BBR}(300\,^\circ{\rm K}) = -\frac{1}{2}\left(831.9\frac{V}{m}\right)^2\left[\frac{T(K)}{300}\right]^4 \alpha_n^{E1}(0)[1+\eta(\alpha_n^{E1},T)], \tag{41}$$

where

$$\eta(\alpha_n^{E1},T) = \frac{\left(\frac{80}{63}\right)\pi^2}{\alpha_n^{E1}(0)T}\Sigma_{m\neq n}\frac{|\langle j_n\|D\|j_m\rangle|^2 T^3}{(2j_n+1)(E_n-E_m)^3}\left(1+\frac{21\pi^2T^2}{5(E_n-E_m)^2}+\frac{336\pi^4T^4}{11(E_n-E_m)^4}\right).$$

Relativistic Many-Body Methods

As mentioned at several occasions, it is possible to estimate many important systematics for the atomic clocks and propose new clock candidates by the theoretical studies. This warrants for development of suitable many-body methods that are capable of considering the electron correlation effects more effectively. It is also evident from most of the undertaken examples that consideration of a suitable atomic

transition with one of the atomic states having almost-degenerate fine-structure partner has many merits, e.g., achieving high stability and accuracy, for the optical frequency standards. This urges for using relativistic mechanics in the theoretical studies. Moreover, almost all the candidates considered for frequency standards are relatively heavy atomic systems for which accurate calculations of their properties necessitates employing valid relativistic many-body methods. Consideration of a fully relativistic theory in its covariant form is impractical in the many-electron bound systems. Working with Dirac Hamiltonian for electrons along with nuclear potential and Coulomb repulsion between the electrons can suffice the present goal to accomplish calculations within the required accuracy. Nevertheless, corrections from higher relativistic effects at the leading orders, e.g., Breit interaction due to exchange of transverse photons and quantum electrodynamics (QED), can be incorporated to augment the calculated results.

General Approach

The Dirac–Coulomb (DC) Hamiltonian for an atomic system is given by

$$H_{\mathrm{DC}} = \Sigma_i [c\ \vec{a}_i \cdot \vec{p}_i + \beta_i\ c^2 + V_n(r_i)] + \Sigma_{\substack{i,j \\ i \geq j}} \frac{1}{r_{ij}} \tag{42}$$

(in atomic unit (au)), where $\vec{\alpha}$ and β are the Dirac matrices and $V_n(r)$ is the nuclear potential. This is a good approximation to describe the positive energy states of the Dirac theory. Weak coupling with the positron wave functions are usually neglected and also the rest mass energy of the electrons can be subtracted for the convenience. Thus, the working DC Hamiltonian reads out

$$H_{\mathrm{DC}} = \Sigma_i [c\ \vec{a}_i \cdot \vec{p}_i + (\beta_i - 1)\ c^2 + V_n(r_i)] + \Sigma_{\substack{i,j \\ i \geq j}} \frac{1}{r_{ij}}. \tag{43}$$

Again, it may not be appropriate to assume atomic nucleus as a point- like object for accurate calculations. On the other hand, there are not proper valid models available to describe the nuclear structure exactly. Among many, Fermi charge distribution model is more popular in which density of an electron within the atomic nucleus is described by

$$\rho_n(r) = \frac{\rho_0}{1 + e^{(r-b)/a}}, \tag{44}$$

where ρ_0 is the normalization factor, b is known as half-charge radius and $a = 2.3/(4 \ln 3)$ is related to the skin thickness of the nucleus. Considering this distribution, the nuclear potential can be obtained as

$$V_n(r) = \frac{Z}{\aleph r} \begin{cases} \frac{1}{b}\left(\frac{3}{2} + \frac{a^2\pi^2}{2b^2} - \frac{r^2}{2b^2} + \frac{3a^2}{b^2}P_2 + \frac{6a^3}{b^2 r}(S_3 - P_3^+)\right) & \text{for } a \geq b \\ \frac{1}{r}\left(1 + \frac{a62\pi^2}{b^2} - \frac{3a^2 r}{b^3}P_2^- + \frac{6a^3}{63b}(S_3 - P_3^-)\right) & \text{for } r \geq b, \end{cases} \tag{45}$$

for the factors $\aleph = 1 + \frac{a^2\pi^2}{b^2} + \frac{6a^3}{b^3}S_3$ with $S_k = \Sigma_{m=1}^{\infty}\frac{(-1)^{m-1}}{m^k}e^{-b/a}$ and $P_k^{\pm} = \Sigma_{m=1}^{\infty}\frac{(-1)^{m-1}}{m^k}e^{\pm m(r-b)/a}$. The b-parameter can be determined from $b = \sqrt{\frac{5}{3}r_{\rm rms}^2 - \frac{7}{3}a^2\pi^2}$ with the root mean square radius $r_{\rm rms}$, which can be estimated using the empirical formula $r_{\rm rms} = 0.836A^{1/3} + 0.57$ in fermi (fm) or can be taken from a standard nuclear data table.

We now turn to outline procedures to estimate corrections due to higher-order relativistic effects. Since the nuclear potential is the dominating contributing term in the above DC Hamiltonian, the QED effects are estimated by treating the nuclear potential as the strong external electromagnetic field seen by the bound electrons in an atomic system. Thus, an effective potential is defined accounting lower-order vacuum polarization (VP) and self-energy (SE) QED effects [19] in place of $V_n(r)$ in the atomic Hamiltonian and reads

$$V_{\rm eff}(r) = [1 + V_U(r) + V_{\rm WC}(r) + V_{\rm mg}(r) + V_{\rm el}(r)]V_n(r), \tag{46}$$

where $V_U(r) = \frac{2\alpha_e}{3\pi}\int_1^{\infty} dt\,\frac{\sqrt{t^2-1}}{t^2}\left(1 + \frac{1}{2t^2}\right)e^{-2{\rm ctr}}$ and $V_{\rm WC}(r) = -\frac{4\alpha_e}{9\pi}\frac{0.092\ Z^2\alpha_e^2}{1+(1.62cr)^4}$ are the Uehling and Wichmann–Kroll corrections, respectively, representing the lower-order VP effect and the SE effect is taken through the magnetic form-factor $V_{\rm mg}(r) = \frac{\alpha_e}{4\pi c}i\gamma\cdot\Delta\left[V_n(r)\left(\int_1^{\infty} dt\,\frac{1}{t\sqrt{t^2-1}}e^{-2{\rm ctr}} - 1\right)\right]$ and the electric form-factor $V_{\rm el}(r) = -\frac{\alpha_e}{\pi}\int_1^{\infty} dt\,\frac{1}{t\sqrt{t^2-1}}\left[\left(1 - \frac{1}{2t^2}\right)\left[\ln(t^2-1) + \ln\left(\frac{4c^2}{\lambda^2}\right)\right] - \frac{3}{2} + \frac{1}{t^2}\right]e^{-2{\rm ctr}}$ for the cutoff parameter chosen cautiously as $\lambda \sim (Z\alpha_e^2)c$.

The potential due to the Breit interaction between the electrons located at the ith and jth positions is given by

$$V_B(r_{ij}) = -\frac{1}{2r_{ij}}\{\alpha_i\cdot\alpha_j + (\alpha_i\cdot\hat{r}_{ij})(\alpha_j\cdot\hat{r}_{ij})\}. \tag{47}$$

Owing to the two-body nature of the Coulomb and Breit interactions, solving eigenvalue equation for the atomic Hamiltonian $H_{\rm at}$ (with only Coulomb or both Coulomb and Breit interactions), given by

$$H_{\rm at}\left|\Psi_n^{(0)}\right\rangle = E_n\left|\Psi_n^{(0)}\right\rangle, \tag{48}$$

with more than three electrons in an atomic system is infeasible. Instead, it is a usual practice to get the approximated solution to the above equation and then append corrections from the residual contributions gradually. This approximated solution is treated as a model space in the working Hilbert or Fock space accounting majority of the contributions from the Coulomb (and Breit) interaction(s) in the calculation of

the atomic wave functions. One of the most conducive and appropriate approaches to determine the approximated wave functions is to use the Hartree–Fock (Dirac–Fock (DF) in the relativistic framework) Hamiltonian (H_0). The residual interaction ($V_{\rm res} = H_{\rm at} - H_0$) can further improve the results by annexing contributions from the rest of the Hilbert or Fock space, referred to as orthogonal space, through a decent many-body method. Below we demonstrate few methods and try to inculcate one-to-one connections among these methods. For this purpose, we try to build up each many-body approach by commencing from the same DF wave function. To proceed further, we adopt the procedure of the generalized Bloch equation to explain the many-body methods systematically in a comprehensible and logical manner. In the many-body perturbation theory (MBPT), the exact wave function of an atomic state can be expressed as

$$|\Psi_n^{(0)}\rangle = \Omega_n^{(0)}|\Phi_n\rangle, \tag{49}$$

where $|\Phi_n\rangle$ is the model space (here DF wave function) and $\Omega_n^{(0)}$ is the wave operator which is responsible for incorporating contributions from the orthogonal space due to $V_{\rm res}$. Orthogonal space contributions can either be expressed in terms of the order of perturbation or in the form of excited configurations with respect to $|\Phi_n\rangle$. For the simple reason, we can go on with the perturbation series expansion approach first and then we can manifest the same in terms of the excited state configurations.

Two projection operators P and Q satisfying $|\Phi_n\rangle = P|\Psi_n^{(0)}\rangle$ and $Q = I - P$ for the identity operator I are defined for easy description, which follows $P = |\Phi_n\rangle\langle\Phi_n|$. In the perturbative approach, it yields

$$\Omega_n^{(0)} = \Omega_n^{(0,0)} + \Omega_n^{(1,0)} + \Omega_n^{(2,0)} + \cdots = \Sigma_k \Omega_n^{(k,0)}. \tag{50}$$

Notice that we use two superscripts, for the later use, among which the first one represents the number of $V_{\rm res}$ present in the calculations, while the second one with zero means there is no external source of perturbation taken into account. The amplitudes of the above wave operators are solved one by one in the sequence of order of perturbations involved with the wave operators using the following recursive relation:

$$\left[\Omega_n^{(k,0)}, H_0\right] P = Q V_{\rm res} \Omega_n^{(k-1,0)} P - \Sigma_{m=1}^{k-1} \Omega_n^{(k-m,0)} P V_{\rm res} \Omega_n^{(m-1,0)} P. \tag{51}$$

The energy of the state (E_n) can be evaluated using an effective Hamiltonian $H_n^{\rm eff} = PH\Omega_n^{(0)}P$ at different orders of perturbation with the expansion form of $\Omega_n^{(0)}$. That is, $E_n = \langle\Phi_n|H_n^{\rm eff}|\Phi_n\rangle$.

There are, specifically, two approaches adopted to evaluate dipole polarizability of an atomic state. In a conventional approach, the energy of the $|\Psi_n\rangle$ state of an atom placed in an isotropic electric field of strength ε changes as [20, 21]

$$E_n(\varepsilon) = E_n(0) - \frac{\alpha_n^{E_1}}{2}\varepsilon^2 - \cdots, \tag{52}$$

where $E_n(0)$ and $E_n(\varepsilon)$ are the total energies of the state in the absence and presence of the electric field, respectively. This approach requires mixed parity orbitals to calculate the atomic energy in the presence of the electric field, which are practically cumbersome using the spherical coordinate systems but are manifested using the molecular methods (in the Cartesian coordinate system).

Alternatively, the modified wave function ($|\Psi_n\rangle$) of the atomic system in the presence of an external weak perturbative source ($V_{\rm prt}$) can be approximated to first-order approximation as

$$|\Psi_n\rangle = \left|\Psi_n^{(0)}\right\rangle + \lambda \left|\Psi_n^{(1)}\right\rangle, \tag{53}$$

where λ is an arbitrary parameter (corresponds to ε in the evaluation of α_n^{E1}) representing the strength of the perturbation source. In this way, α_n^{E1} can be obtained by expressing [22–26]

$$\alpha_n^{E1} = \frac{\langle\Psi_n|D|\Psi_n\rangle}{\Psi_n|\Psi_n} \simeq \frac{\langle\Psi_n^{(0)}|D|\Psi_n^{(1)}\rangle}{\Psi_n^{(0)}|\Psi_n^{(0)}}, \tag{54}$$

by considering $V_{\rm prt} \equiv \varepsilon D\hat{z}$ for small values of ε.

It is commanding to obtain solution for $|\Psi_n^{(1)}\rangle$ by solving an inhomogeneous equation of the type

$$\left|(H_n^{\rm eff} - E_n)\right| \Psi_n^{(1)}\rangle = (E_n^1 - V_{\rm prt})|\Psi_n^{(0)}\rangle, \tag{55}$$

analogous to the approach proposed and implemented by Dalgarno [27]. In Bloch equation methodology, we can express $|\Psi_n^{(1)}\rangle = \Omega_n^{(1)}|\Phi_n\rangle$ such as $\Omega_n^{(1)} = \Sigma_k \Omega_n^{(k,1)}$ encompassing kth order of $V_{\rm res}$ and one order external perturbation $V_{\rm prt}$. The amplitudes of $\Omega_n^{(1)}$ are obtained from the following equation [26]:

$$\left[\Omega_n^{(k,1)}, H_0\right] P = Q[V_{\rm prt}\Omega_n^{(k,0)} + V_{\rm res}\Omega_n^{(k-1,1)}]P - \Sigma_{m=1}^{k-1}\left(\Omega_n^{(k-m,0)} P V_{\rm prt}\Omega_n^{(m,0)} P - \Omega_n^{(k-m,1)} P V_{\rm res}\Omega_n^{(m,0)}\right) P. \tag{56}$$

For the choice of reference state $|\Phi_n\rangle$ as the DF wave function and external perturbation operator $V_{\rm prt}$ being a one-body operator, the zeroth-order expressions for the wave operators can yield $\Omega_n^{(0,0)} = 1$, $\Omega_n^{(1,0)} = 0$, and $\Omega_n^{(0,1)} = \Sigma_{p,a}\frac{\langle p|V_{\rm prt}|a\rangle}{\epsilon_p - \epsilon_a}$ for the occupied a and unoccupied p-orbitals with energies ϵ_a and ϵ_p, respectively. In the double perturbative sources, up to $k = 0, 1, 2\cdots$ approximations in the wave operators are referred to MBPT(1) or DF, MBPT(2), MBPT(3), etc. methods, respectively.

Having said and done with the basic formalism of determining atomic wave functions in the many-body perturbative analysis, extending them to build up these wave functions containing all orders in $V_{\rm res}$ for both the cases, the absence and presence of external source, would be now much straightforward. This can be achieved by generalizing the above perturbative approaches after carefully formulating the wave operator Ω_n in a slight different form or assembling the coefficients from each order of perturbation expansion to compose various degrees of excitations. We discuss few important all-order many-body methods that are widely used in the studies of atomic clock pertinent physics; specifically, two all- order perturbative approaches known as **configuration interaction** (CI) and **coupled-cluster** (CC) methods in the absence of external field and their extensions to deal with weak external interaction up to the first-order perturbation.

Configuration Interaction (CI) Method

As mentioned above, the exact wave function $|\Psi_n^{(0)}\rangle$ can be expressed in terms of excited configurations with respect to $|\Phi_n\rangle$. Mathematically, this corresponds to

$$\begin{aligned}|\Psi_n^{(0)}\rangle &= |\Phi_n\rangle + \Sigma_I^{N_I} C_n^I |\Phi_n^I\rangle + \Sigma_{II}^{N_{II}} C_n^{II} |\Phi_n^{II}\rangle + \Sigma_{III}^{N_{III}} C_n^{III} |\Phi_n^{III}\rangle + \cdots \\ &\equiv \tilde{C}_n \left|\psi_n^{(0)}\right\rangle + \Sigma_I^{N_I} \tilde{C}_I \left|\psi_I^{(0)}\right\rangle + \Sigma_{II}^{N_{II}} \tilde{C}_{II} \left|\psi_{II}^{(0)}\right\rangle + \Sigma_{III}^{N_{III}} \tilde{C}_{III} \left|\psi_{III}^{(0)}\right\rangle + \cdots \end{aligned} \tag{57}$$

where $|\Phi_n^K\rangle$s are the determinants corresponding to DF wave functions for N_K number of possible kth excited states constructed from $|\Phi_n\rangle$ and C_n^*s are their corresponding mixing coefficients. Obviously it is only possible to generate N-tuple excitations for full expansion in a system having N-number of electrons. The above expansion is a direct consequence of the fact that each order in the correction from the perturbed wave functions can be expressed as linear combination of excited states. This also directly follows; the excited state determinants are nothing but the leading part of the excited atomic states. As a result, the calculated atomic state function (ASF) $|\Psi_n^{(0)}\rangle$ can be expressed as a linear combination of configuration state functions (CSFs) representing the trial atomic state functions (denoted by the notation $|\psi_n^{(0)}\rangle$). This approach of constructing atomic wave functions is known as CI method. In a special condition the above excited Slater determinants ($|\Phi_n^*\rangle$) can also be chosen as CSFs in a CI method. By diagonalizing the atomic Hamiltonian with respect to the CSFs, one can obtain the values of the C_k^* or $\tilde{C}_k^*$ coefficients. Although conceptually CI method looks simpler, computationally it is much difficult to account the higher-level excitations and often it has been truncated to only single and double excitations (referred to as CISD method) in the practical applications. It can be shown by considering only the double configurations that the corrections from $V_{\rm res}$ to the wave function calculations in a truncated CI method are proportional to $1/\sqrt{N}$ [20, 28]. This simply means for $N \to \infty$, the correlation contributions to the calculation diminish (famously known as size-consistency problem). Hence, truncated CI method may not be judicious to employ for accurate calculations of

atomic properties in the heavy systems (large N) conceptually and results from the truncated CI can be questionable for their validation. Nevertheless, results close to the experimental values can be achieved by taking only important CSFs from the energy-level configurations in a truncated CI method. Indeed upon the consideration of more Slater determinants in the construction of $|\psi_n^{(0)}\rangle$, it is possible to improve the quality of truncated CI results, and this approach is recognized as multi-configurational Dirac–Fock (MCDF) method.

In a similar approach, the first-order correction to the wave function due to the external field and matrix elements of an operator among different states can be obtained by mixing CSFs with appropriate parities and angular momentum as per the selection rules.

Coupled-Cluster (CC) Method

In the CC method, linear combination of the Slater determinants are carried out in a distinct manner so that atomic wave functions are contrived to form an exponential function. Starting with the same basic principle as in the CI method, we can rewrite

$$\begin{aligned}
|\Psi_n^{(0)}\rangle &= |\Phi_n\rangle + \Sigma_I^{N_I} C_n^I |\Phi_n^I\rangle + \Sigma_{II}^{N_{II}} C_n^{II} |\Phi_n^{II}\rangle + \Sigma_{III}^{N_{III}} C_n^{III} |\Phi_n^{III}\rangle + \cdots \\
&\equiv |\Phi_n\rangle + \Sigma_I^{N_I} T_I^{(0)} |\Phi_n\rangle + \Sigma_{II}^{N_{II}} T_{II}^{(0)} |\Phi_n\rangle + \Sigma_{III}^{N_{III}} T_{III}^{(0)} |\Phi_n\rangle + \cdots \\
&= |\Phi_n\rangle + T_1^{(0)} |\Phi_n\rangle + T_2^{(0)} |\Phi_n\rangle + T_3^{(0)} |\Phi_n\rangle + \cdots \\
&= e^{T_1^{(0)} + T_2^{(0)} + T_3^{(0)} + \cdots + T_N^{(0)}} |\Phi_n\rangle = e^{T^{(0)}} |\Phi_n\rangle, \qquad (58)
\end{aligned}$$

where $T_k^{(0)} = \Sigma_K^{N_K} T_K^{(0)}$ for $k = 1, 2, 3 \cdots$ represents the CC excitation operator with subscript k implying the kth level excitation carried out from $|\Phi_n\rangle$. The advantage of this method is of manyfold: (i) it is both conceptually and computationally simpler, (ii) truncated CC methods also satisfy both size-extensivity and size-consistency properties, (iii) owing to exponential form of the expression for the wave function, contributions from higher-level excitations to a certain extent also do appear through the nonlinear terms in a truncated CC method, etc.

Although we mentioned above about computational simplicity in the use of CC method, in actual practice it may not turn out to be factual. Because of the presence of the nonlinear terms and requirement of a sufficiently large Hilbert or Fock space to carry out accurate calculations of the atomic wave functions, intermediate computational strategy may be required conforming available computational resources and depending upon the size of the atomic system of our interest [20, 22, 26]. This can be judiciously accomplished by devising a proper plan before implementing the method. For example, a well-suited symmetry group and Kramers relations [29] can be adopted to reduce the computational cost when molecular orbitals are used. Since atomic orbitals are meticulously described in the spherical coordinate system, the use of reduced matrix elements instead of actual matrix elements would be more pertinent and can prevail extra computations for the azimuthal quantum numbers. This can be the most well-versed approach for states having closed-sell

configurations, but states of open-shell configurations cannot be dealt with this way. However, atomic states having one or two electrons in the valence orbitals and one or two electron less from closed-shell configurations can be computed using the reduced matrix elements by appending valence orbitals or removing electrons from the appropriate closed-shell configurations in the Fock-space approach. We discuss here a few such approaches but restricting to only one electron attachment or removal from the closed-shell configurations.

In the Fock-space CC formalism, wave functions of one valence ($v = n$) atomic states are expressed as

$$|\Psi_n\rangle = e^{T_0^{(0)}} \left\{1 + S_v^{(0)}\right\} |\Phi_v\rangle, \tag{59}$$

where S_v is a CC operator exciting the valence electron v along with the closed-core $|\Phi_v\rangle$. In a Fock-space approach, $|\Phi_v\rangle$ is constructed from the closed-core $|\Phi_0\rangle$ by appending the respective valence orbital v as $|\Phi_v\rangle = a_v^+|\Phi_0\rangle$. Likewise, wave functions for the states having less than one electron from the closed-shell configurations can be expressed as

$$|\Psi_a\rangle = e^{T_0^{(0)}} \left\{1 + R_a^{(0)}\right\} |\Phi_a\rangle, \tag{60}$$

where R_a is the CC operator responsible for annihilating a core electron ($a = n$) from the closed-core $|\Phi_n\rangle$ and exciting an electron from a reference state $|\Phi_a\rangle$ constructed as $|\Phi_a\rangle = a_a|\Phi_0\rangle$. In both of the above approaches, CC $T_0^{(0)}$ operator is responsible for accounting electron excitations from the closed-core $|\Phi_0\rangle$. In these expressions, superscript (0) is used to highlight that wave functions are still free from the external fields. It is called as CCSD method when only the single and double excitations are taken in, while it is known as CCSDT, CCSDTQ, etc. methods with the inclusion of triples, quadrupoles, and so on excitations, respectively [20]. Since computational complexity increases with the addition of higher-level excitations, CCSD approximation is the prevalent method in the atomic and molecular spectroscopy studies. Contributions from important triples can be included perturbatively through the CCSD method in the CCSD[T], CCSD(T), CCSDpT, and CCSDvT framework [20,25,30] in the same amount of computational requirements to uplift the results further. The matrix element of an operator O between the $|\Psi_f\rangle$ and $|\Psi_i\rangle$ states (for the expectation value $|\Psi_f\rangle = |\Psi_i\rangle$) can be determined by

$$\begin{aligned}\langle O\rangle_{fi} &= \frac{\langle\Psi_f|O|\Psi_i}{\sqrt{\langle\Psi_f|\Psi_f\rangle\langle\Psi_i|\Psi_i\rangle}} \\ &= \frac{\langle\Phi_f\left|\left\{1+\Omega_f^+\right\}\sigma\{1+\Omega_i\}\right|\Phi_i}{\sqrt{\langle\Phi_f\left|\left\{1+\Omega_f^+\right\}\bar{N}\{1+\Omega_i\}\right|\Phi_f\rangle\langle\Phi_i\left|\left\{1+\Omega_i^+\right\}\bar{N}\{1+\Omega_i\}\right|\Phi_i\rangle}},\end{aligned} \tag{61}$$

where Ω_n is either S_n or R_n for the attachment or detachment of an electron case, respectively, and $\bar{O} = e^{T_0^{(0)+}} O e^{T_0^{(0)}}$ and $\bar{N} = e^{T^{(0)+}} e^{T^{(0)}}$ are two non-truncated series in the above expression. For the closed-shell atomic states ($|\Psi_i\rangle = |\Psi_f\rangle = |\Psi_0\rangle$), it yields [20]

$$\langle O\rangle = \frac{\langle\Psi_0|O|\Psi_0\rangle}{\sqrt{\langle\Psi_0|\Psi_0\rangle}} = \langle\Phi_0|\bar{O}|\Phi_0\rangle_c \tag{62}$$

where the subscript c in the expression represents to only the connected terms. In many cases, only linear terms even in the CCSD method are considered (referred here as LCCSD method) for carrying out the calculations. The above non-truncated series can be computed at several steps [22–26], but it is advisable to use other CC methods like normal CC method with biorthogonal condition, extended CC method, etc. to fend off these non-truncated series [20].

Creating excited states from an atomic state of closed-shell configuration $|\Psi_0\rangle$ in a Fock-space approach is onerous, but an equation-of-motion CC (EOMCC) method is apropos to determine Lth level excited states $|\Psi_L(J,\pi)\rangle$ with total angular momentum J and parity π from $|\Psi_0\rangle$. In this approach, it follows [20, 31]

$$|\Psi_L(J,\pi)\rangle = \Omega_L(J,\pi)|\Psi_0\rangle = \Omega_L(J,\pi)e^{T_0^{(0)}}|\Phi_0\rangle = e^{T_0^{(0)}}\Omega_L(J,\pi)|\Phi_0\rangle. \tag{63}$$

By construction both the $T^{(0)}$ and Ω_L operators are similar in nature; hence, they can commute each other, but they commission operationally different roles.

In the presence of an external source, the first-order corrected wave function in the CC method can be expressed as [22–26]

$$|\Psi_n^{(1)}\rangle = e^{T^{(0)}}(T^{(1)})|\Phi_n\rangle \tag{64}$$

Similar to $T^{(0)}$ the perturbed CC operator $T^{(1)}$ excites electrons from $|\Phi_n\rangle$, but parity of $T^{(0)}$ is always even as it originates from the Coulomb (or Breit) operator, while the parity of $T^{(1)}$ can depend on the characteristic of $V_{\rm prt}$.

Random-Phase Approximation (RPA)

Though **random-phase approximation** is a subclass of CC method, technically it is derived from the DF method in a completely different approach. Its main advantage is that it can embody the core-polarization effects to all orders at the same time being cost effective. Its expression can be obtained from Eq. (56) by continuing k to infinite order for $\Omega^{(k,1)}$ while suppressing $\Omega^{(k,0)}$ in a self-consistent procedure. The derivation of the final expression is a repercussion of expanding the DF wave function $|\Phi_n\rangle$ to first order due to $V_{\rm prt}$ and generalizing it to infinite order. Thence, it only picks up the singly excited configurations from $|\Psi_n\rangle$ in case of polarizability

calculations owing to one-body form of the interaction operator $V_{\rm prt} \equiv D$. In the RPA approach, the first-order corrected wave operator $\Omega_n^{(1)} \equiv \Omega_n^{(\rm RPA)}$ is explicitly given by

$$\Omega_n^{(\rm RPA)} = \Sigma_{k=1}^{\infty} \Sigma_{pq,ab} \left\{ \frac{[\langle pb|V_{\rm res}|aq\rangle - \langle pb|V_{\rm res}|qa\rangle]\Omega_{b\to q}^{(k-1,1)}}{\epsilon_p - \epsilon_a} + \frac{\Omega_{b\to q}^{(k-1,1)^+}[\langle pq|V_{\rm res}|ab\rangle - \langle pq|V_{\rm res}|ba\rangle]}{\epsilon_p - \epsilon_a} \right\}, \tag{65}$$

where $a \to p$ implies singly excitation operation by the wave operator replacing an orbital a by p in $|\Phi_n\rangle$.

Demonstration of Few Theoretical Results

The main aim here is to testify the role of relativistic many-body methods for accurate calculations of the quantities that can be useful in various possible ways for the atomic clock studies, especially in the precise estimate of systematics. Proclaiming through only few examples, accurate calculations of $A_{\rm hf}$ constants, g_J-factors, static polarizabilities, and quadrupole moments of atomic states in a number of clock candidates will be demonstrated at different levels of approximations in the employed many-body methods.

$A_{\rm hf}$-Constants and g_J-Factors

Almost all heavy singly charged ions from the alkaline earth metal group in the periodic table (from Ca to Ra) are either considered or proposed for the atomic clocks because these ions have typically an S-state as the ground state and two excited metastable D-states. In Table 2, a summary of the calculated results for $A_{\rm hf}/(\mu_I/I)$ of the above states in the alkaline earth ions, Yb^+ and Hg^+, considering the DF, RCC, and MCDF methods is reported from [6, 8, 32–34]. It has been shown explicitly in these references that accurate calculations of these quantities in the ground and $D_{3/2}$ states require both pair-correlation and core-polarization effects to be considered to all orders, while the core-polarization effects are solely important for the correct evaluation of $A_{\rm hf}$ values in the $D_{5/2}$ states [33]. In fact, these values for the metastable $6D_{3/2}$ state in Ra^+ was first predicted by theory using the CC method and later it was verified by the experiment [35].

Among others, $^{171}Yb^+$ion is the most valuable ion for the atomic clock perspectives as its three transitions, the largest among all the elements, are considered for the frequency standards as it has three metastable states ($5d$-fine-structure levels with lifetimes of the order of few seconds and the first excited $4f^{13}\ ^2F_{7/2}$ state

Table 2 Comparison between the calculated $A_{\rm hf}$ values divided by I/μ_I from the DF, CC, and MCDF methods against their available experimental results in the alkaline earth ions, Yb^+ and Hg^+. The MCDF results are mostly taken from [32], while other results are compiled from the electron attachment method of [6, 8, 33, 34] and from the references therein. Here n represents the principal quantum number of the ground state of the corresponding ion

Method	Ca^+	Sr^+	Ba^+	Ra^+	Yb^+	Hg^+
ns state						
DF	1,561.62	3,039.53	4,741.67	14,913.89	9,637.42	34,380.31
CC	2,142.01	4,123.06	6,517.57	19,689.37	13,166.69	41,592.88
MCDF						
Experiment	2,142.02	4,120.04	6,429.83	18,772	12,807.14	40,035.74
$(n-1)d_{3/2}$ state						
DF	88.74	128.30	205.38	293.53	286.67	
CC	125.64	187.64	305.47	441.67	406.14	
MCDF	125.56		308.83		405.55	
Experiment	125.64		303.77	430.67	435.51	
$(n-1)d_{5/2}$ state						
DF	9.18	53.56	+82.58	107.60	109.47	
CC	9.56	−8.83	−19.19	−138.36	−69.88	
MCDF	12.86		+15.03		−12.75	
Experiment	10.34	8.95	−19.25		−64.42	

with a lifetime about 6 years). The octopole transition $6s \rightarrow 4f_{7/2}$ of this ion has the uttermost narrow linewidth which is advantageous for making prospective primary frequency standard. Except the $4f_{7/2}$ state, the other three states (including the ground state) of $^{171}Yb^+$ have three valence orbitals, but they can be treated as states with one valence configurations with the closed-core $[4f^{14}]$. Also, both the ground and $4f_{7/2}$ states of Yb^+ can be regarded as one electron less than the closed-core $[4f^{14}]6s^2$. Analogously both the states of the clock transition $[5d^{10}]6s\ ^2S_{1/2} \rightarrow 5d^9 6s^2\ ^2D_{5/2}$ in Hg^+ can be calculated in the similar approach. Comparison between the $A_{\rm hf}$ values obtained using different many-body methods and with the available experimental results can be found from Tables 2 and 3. It can be seen from these tables that when $A_{\rm hf}$ of the ground state of Yb^+ is calculated using an electron detachment than attachment approach, the result becomes closer with the experimental value. This is because in the former case the core orbitals see effect due to the valence electron and in the latter approach they come only through the CC operators [36]. It is indeed a quite remarkable observation, though not much difference found in the calculations for Hg^+ (Nandy DK, Private communication).

Lindroth and Ynnerman had critically evaluated Δg_J values of the ground states of Li, Be^+, and Ba^+ to quite a high accuracy using the CC method in comparison with the experimental values [16]. In Table 4, preliminary calculations of g_J and δg_J values have been demonstrated of the ground and metastable states of Ca^+ using the CC method at different levels of approximations and considering the higher-order relativistic corrections, while the semiclassical formula gives these

Table 3 Comparison between the calculated $A_{\rm hf}$ values divided by I/μ_I from the MCDF [32] method and electron detachment theory with DF and CC methods [36] against their available experimental values

	Yb^+		Hg^+	
	$[4f^{14}]6s$ state	$4f^{13}6s^2\ ^2F_{7/2}$ state	$[5d^{10}]6s$ state	$5d^96s^2\ ^2D_{5/2}$ state
DF	7,318.10	878.78	30,050.78	902.43
CC	12,871.96	1,016.87	41,608.82	1,080.29
MCDF				952.29
Experiment	12,807.14	916.60		

Table 4 Preparatory calculations of g_J-factors and their δg_J corrections (given within the parentheses) in the states of interest for the clock transitions in Ca^+ from DF and different CC methods and comparison with the semiclassical and available experimental results

Method	$4s\ ^2S_{1/2}$ state	$3d\ ^2D_{3/2}$ state	$3d\ ^2D_{5/2}$ state
DF	1.99995307831	0.799922324299	1.19991738748
	(+0.00284130797)	(−0.000898052128)	(+0.00112057581)
LCCSD	1.99347972816	0.796582101999	1.19726585871
	(+0.00282940412)	(−0.000898459207)	(+0.00111787848)
CCSD	1.99783376440	0.802135504861	1.20140651205
	(+0.00283303463)	(−0.000893389416)	(+0.00111891602)
CCSD(T)	1.99941009011	0.802210930664	1.20151553526
	(+0.00283303788)	(−0.000893393408)	(+0.00111892083)
CCSDvT	1.99942182027	0.802214948013	1.20151548994
	(+0.00283304004)	(−0.000893394022)	(+0.00111893179)
Δ Breit	−0.00000546639	+0.00004031931	+0.00000746347
	(−0.00000000443)	(−0.00000000687)	(+0.00000000212)
Δ QED	−0.00000006897	+0.00000000808	+0.00000000002
	(−0.00000000014)	(+0.00000000004)	(+0.00000000001)
Final	2.00224932	0.801361875	1.20264189
Semiclassical	2.0	0.8	1.2
Experiment [37]	2.00225664(9)		

values as 2, 0.8, and 1.2 for the respective states. In the above table, the ground state value is also compared with its experimental result [37] up to the 4th decimal places. It, thus, clearly advocates in favor of the potential of the CC method to produce results accurately.

Static Polarizabilities (In Au)

Two elegant methods have already been described to evaluate α_n^{E1} values in the atomic systems. To demonstrate the first kind, these values for the $3s^2\ ^1S_0$ and $3s3p\ ^3P_0$ states in Al^+ along with excitation energy between these two states are given in Table 5 by taking two arbitrary values of the electric field as $\varepsilon = 0.001$ au

Table 5 An example of amelioration of α_n^{E1} values for the states of clock transition in Al^+ by estimating second- order energy difference with arbitrary electric fields in the finite gradient method using an EOMCC approach [21]. The final results are also compared with a sum-over-states approach in the CI+MBPT method framework [38]

Method	Excitation energy	$3s^2\,{}^1S_0$	$3s3p^3\,P_0$	Differential
CCSD	37,186 ± 52	24.251 ± 0.044	24.656 ± 0.088	0.406 ± 0.042
+Triples	146 ± 33	−0.126 ± 0.011	−0.061 ± 0.015	0.065 ± 0.026
+Quadrupoles	2 ± 4	−0.002 ± 0.005	−0.001 ± 0.002	0.003 ± 0.007
+Breit	−6 ± 6	0.015 ± 0.0015	0.018 ± 0.018	0.003 ± 0.003
Total	37,326 ± 95	24.137 ± 0.075	24.614 ± 0.123	0.477 ± 0.078
CI+MBPT		24.048	24.543	0.495

Table 6 Ground state α_n^{E1} values of few alkaline earth atoms, Yb and Hg, from different many-body methods

Method	Ca	Sr	Ba	Yb	Hg	Al^+	In^+
DF	122.90	156.83	218.88	122.21	40.95	19.514	25.734
MBPT(2)	151.70	188.98			34.98		
MBPT(3)	132.80	163.13			22.98	21.752	18.374
RPA					44.98	26.289	29.570
LCCSD					33.91	26.118	25.360
CCSD	157.03	186.98	268.19	144.59	34.98	24.299	24.246
CCSD(T)					33.95		
CCSDpT					34.07	24.26	24.11
+ Breit	156.83	186.80			34.16		
+ QED	∼0.0	186.78			34.27		
CI+MBPT [38]						20.048	24.01
Experiment	169(17)	186(15)	268(22)	142(36)	33.91(34)		

and $\varepsilon = 0.002$ au in an EOMCC method with various levels of approximations [21]. These values are also compared with another latest calculation using the CI+MBPT method in a sum-over-states approach [38].

There have also been α_n^{E1} values evaluated for the ground states of alkaline earth atoms, Al^+, In^+, Yb, and Hg, and of the ground and excited states involved in the clock transitions of alkaline earth ions by calculating the first-order perturbed wave functions due to the E1 operator using the DF and CC methods. In Table 6, results for the ground states of the above systems are listed from these calculations and experiments [23, 26, 30, 39]. Similarly, results for the ions are given and compared with the experimental values [22–25] in Table 7. Polarizabilities due to the E1, M1, and E2 operators in these ions and only due to the E1 operator in few alkali atoms are also determined precisely using the sum-over-states approaches by dividing electron correlation effects into three classes [8, 18, 40–42] as contributions: due to the valence correlation, due to correlations among the occupied electrons, and due to the core-valence correlation which are denoted by $\alpha_n^{O,v}$, $\alpha_n^{O,c}$, and $\alpha_n^{O,cv}$

Table 7 Evaluation of results in the clock transitions of alkaline earth ions using the DF and RCC methods and comparison with the available measured values taken from [22–25]

	Ca^+			Sr^+			Ba^+			Ra^+		
Method	$4s_{1/2}$	$3d_{3/2}$	$3d_{5/2}$	$5s_{1/2}$	$4d_{3/2}$	$4d_{5/2}$	$6s_{1/2}$	$5d_{3/2}$	$5d_{5/2}$	$7s_{1/2}$	$6d_{3/2}$	$6d_{5/2}$
DF	42.76	40.66	39.71	56.72	64.83	60.82	81.99	40.03	38.96	73.18	81.36	63.90
CCSD(T)	73.0	28.5	29.5	88.29	61.43	62.87	124.26	48.81	50.67	104.54	83.71	82.38
Experiment	70.89(15)			93.3(9)			123.88(5)					
	75.3(4)											
	72.5(19)											

for the corresponding operator O, respectively [18]. These values, with individual contributions, are given in Table 8.

Using the above static polarizabilities, the BBR shift due to the E1 operator for the clock transition $3s^2\ ^1S_0 \rightarrow 3s3p\ ^3P_0$ in Al^+ yields $\Delta v = -0.0041$ Hz, while they are about 0.37, 0.22, 0.64, and 0.19 in Hz for the $ns \rightarrow (n-1)d_{5/2}$ clock transitions in the Ca^+, Sr^+, Ba^+, and Ra^+ ions, respectively, with n as the principal quantum number of the ground states of the respective ion.

In a similar approach to above sum-over-states approach, dynamic α_n^{E1} polarizabilities for both the ground and first two excited states are evaluated using both the linearly and circularly polarized lights in the alkali atoms [41, 42]. To illustrate magic wavelengths obtained from these calculations, the dynamic polarizabilities of the $5s$, $5p_{1/2}$, and $5p_{3/2}$ states in the Rb atom are plotted against the wavelengths of applied lights with both the linearly and left-handed circularly polarizations in Fig. 1a–d. **Magic wavelengths** for the $ns\ ^2S_{1/2} \rightarrow np\ ^2P_{1/2,3/2}$ transitions in the Rb atom due to these lights are indicted by pointing arrows at the crossings of the polarizability values of the above states.

Quadrupole Moments

Itano has reported quadrupole moments of many of the abovementioned ions and for Yb^+ by performing MCDF calculations [32]. These values are compared with the CC and experimental results [6, 8, 36, 43] in Table 10. As was in the A_{hf} results, the MCDF results for the Θ values in the nd states of the Alkaline earth ions are also off than their respective experimental values, whereas CC methods were able to produce them more reliably. A detailed analysis to the passage of electron correlation effects and role of higher relativistic corrections for precise estimate of Θ values in the Ca^+ metastable states is also demonstrated in Table 9. It can be noticed that correlation effects improve the results from the DF values significantly and occurrence of strong cancelations is found among the higher-order correlation effects. Though it appears that relativistic corrections are small in the determination of Θ values, their contributions, especially from the Breit interaction, are useful in order to reduce the uncertainty in the calculations. As can also be seen, some

Table 8 Determination of multipolar polarizabilities in the alkali atoms and alkaline earth ions using sum-over-states approaches as have been reported in [18,40–42]

Ion	$\alpha_n^{O,v}$	$\alpha_n^{O,c}$	$\alpha_n^{O,cv}$	α_n^{O}	$\alpha_n^{O,v}$	$\alpha_n^{O,c}$	$\alpha_n^{O,cv}$	α_n^{O}
	$ns\ ^2S_{1/2}$ state				$(n-1)d\ ^2D_{5/2}$ state			
α_n^{M1} result								
Ca^+	0.9×10^{-5}	5.0×10^{-5}	-1.4×10^{-8}	5.9×10^{-5}	−957	5.0×10^{-5}	0.0	−957
Sr^+	2.8×10^{-6}	1.0×10^{-3}	-2×10^{-8}	1×10^{-3}	−208	1×10^{-3}	-2.5×10^{-7}	−208
Ba^+	1.0×10^{-5}	3.5×10^{-4}	∼0.0	3.6×10^{-6}	−72.63	3.5×10^{-4}	-1.4×10^{-5}	−71.63
Ra^+	1.0×10^{-4}	6.3×10^{-3}	∼0.0	6.4×10^{-3}	−35.23	6.3×10^{-3}	-2.7×10^{-5}	−35.23
α_n^{E2} result								
Ca^+	900.64	6.15	0.0	906.79	−3,712.33	6.15	0.18	−3,706.00
Sr^+	1,351.50	14.50	-1.7×10^{-8}	1,366	−1,746.74	14.50	0.24	−1,746.74
Ba^+	4,184.80	48.51	−0.15	4,233.16	−1,357.03	48.51	−0.87	−1,309.39
Ra^+	2,478.98	64.37	−0.11	2,543.22	−1,220.64	64.37	−0.58	−1,156.85

Atom	$\alpha_n^{E1,v}$	$\alpha_n^{E1,c}$	$\alpha_n^{E1,cv}$	α_n^{E1}	Expt	$\alpha_n^{E1,v}$	$\alpha_n^{E1,c}$	$\alpha_n^{E1,cv}$	α_n^{E1}	$\alpha_n^{E1,v}$	$\alpha_n^{O,c}$	$\alpha_n^{E1,cv}$	α_n^{E1}
	$ns\ ^2S_{1/2}$ state					$np\ ^2P_{1/2}$ state				$np\ ^2P_{3/2}$ state			
Li	163.8	0.22	∼ 0.0	164.2	164.2(11)	126.75	0.22	∼ 0.0	126.97	126.76	0.22	∼ 0.0	126.76
Na	161.48	0.9	∼ 0.0	162.4	162.7(8)	360.1	0.9	∼ 0.0	361.0	361.1	0.9	∼ 0.0	362.0
K	284.36	5.5	−0.13	289.8	290.58(1.42)	599.8	5.5	∼ 0.0	605.0	610.8	5.5	∼ 0.0	616.0
Rb	309.46	9.1	−0.26	318.3	319.9(6.1)	801.4	9.1	∼ 0.0	810.5	858.9	9.1	∼ 0.0	868.0

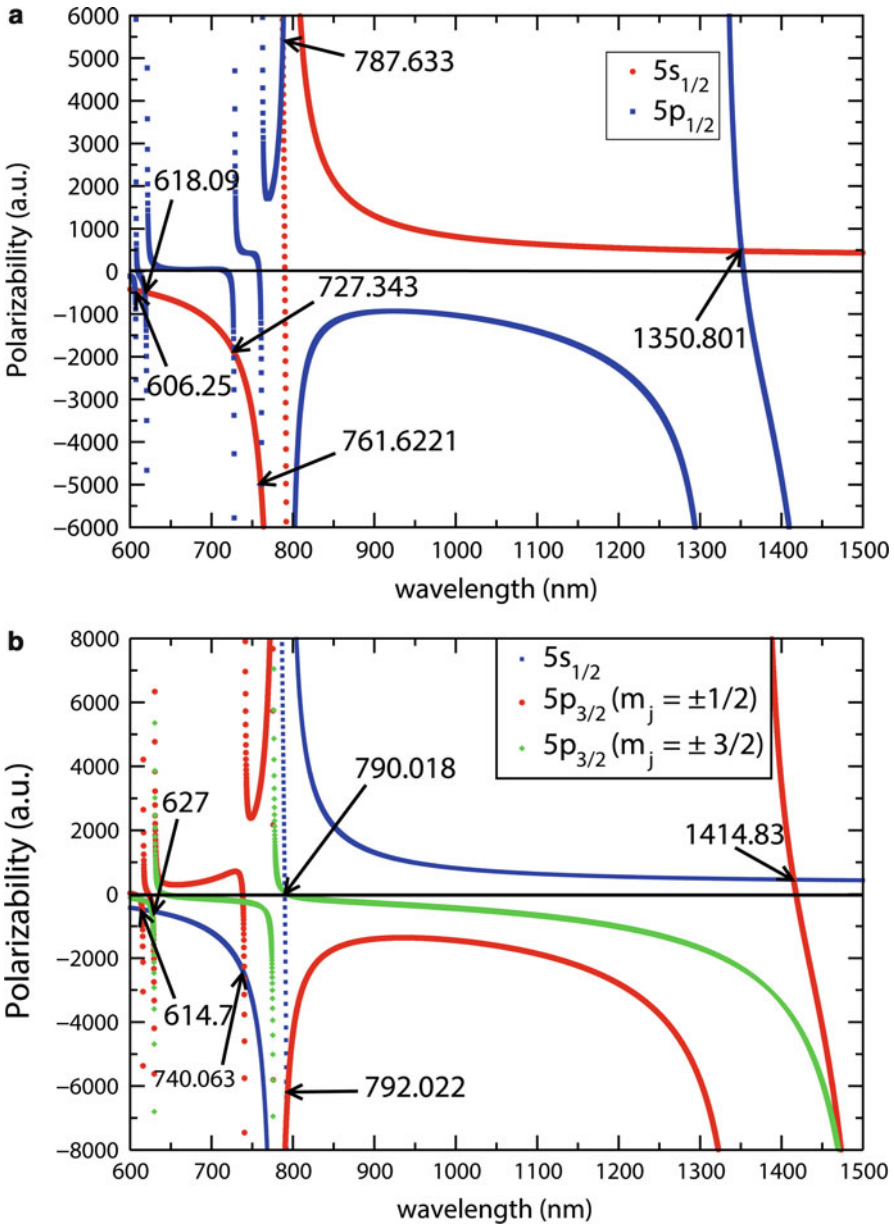


Fig. 1 (continued)

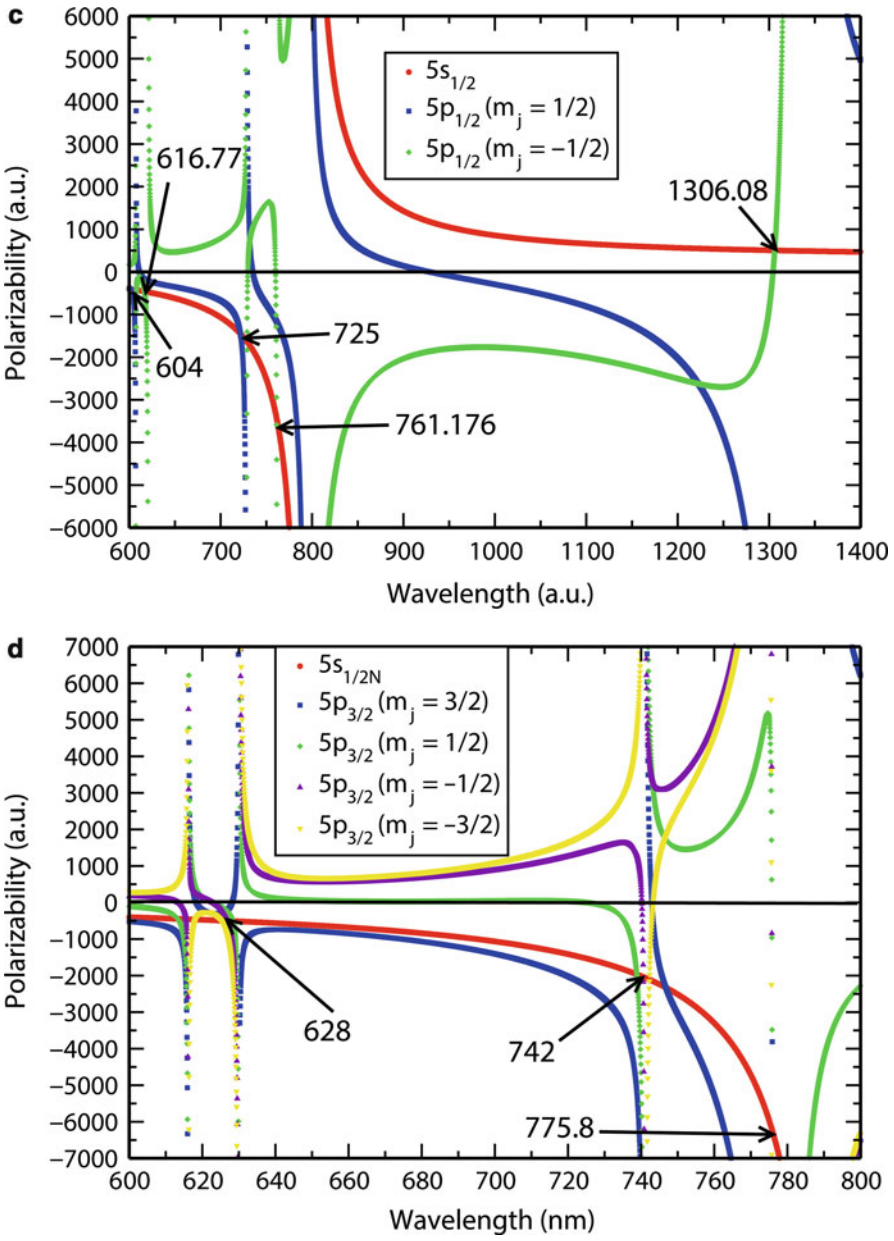


Fig. 1 Dynamic polarizabilities of the $5s$, $5p_{1/2}$, and $5p_{3/2}$ states in Rb atom corresponding to the linearly and left-handed circularly polarized lights. Magic wavelengths for (**a**) the $5s \rightarrow 5p_{1/2}$ transition with the linearly polarized light, (**b**) the $5s \rightarrow 5p_{3/2}$ transition with the linearly polarized light, (**c**) the $5s \rightarrow 5p_{1/2}$ transition with the circularly polarized light, and (**d**) the $5s \rightarrow 5p_{3/2}$ transition with the circularly polarized light, shown pointing by *arrows* at the crossing points

Table 9 Theoretical and experimental Θ values of states involved in the clock transitions of alkaline earth and Yb ions. A detailed analysis for the Ca^+ results given at different levels of approximations in the CC method and contributions from higher-order relativistic corrections is tabulated explicitly

	Ca^+		Sr^+		Ba^+		Ra^+			Yb^+	
Method	$3d_{3/2}$	$3d_{5/2}$	$4d_{3/2}$	$4d_{5/2}$	$5d_{3/2}$	$5d_{5/2}$	$6d_{3/2}$	$6d_{5/2}$	$5d_{3/2}$	$5d_{5/2}$	$4f_{7/2}$
DF	1.712	2.451	2.469	3.496	2.72	3.99	3.58	5.29	2.440	3.613	−0.259
MBPT(2)	1.288	1.848									
LCCSD	1.234	1.770848									
CCSD	1.309	1.877							2.068	3.116	−0.216
CCSD(T)	1.301	1.867	2.12	2.94	2.32	3.42	2.90	4.45			
CCSDvT	1.291	1.846848									
+Triples	−0.005	−0.004848									
+Breit	−0.003	−0.002848									
+QED	0.00002	0.00003848									
MCDF [32,44]	1.338	1.917	2.107	3.048	2.297	3.379			2.174	3.244	−0.22
Experiment		1.83(1)		2.6(3)					2.08(11)		−0.041(5)

of the calculations are as par with the available measurements, and in some cases they are even better than their experimental values. For many states, experimental results are not known but the calculated values can be reliably trusted because the employed methods have been proven to give rise this quantity in other systems very accurately. In fact, it is interesting to note that calculations of Θ in the $4f_{7/2}$ state of Yb^+ using the MCDF [44] and CC [36] methods report consistent values but its corresponding experimental result is found to be quite small. This, essentially, calls for further experimental and theoretical investigation to be sure about its correct value and to scrutinize reasons for which such large discrepancies are resulting. Nevertheless, both the experimental and theoretical results for the $5d_{3/2}$ state in Yb^+ agree with each other quite nicely and uncertainty in the theoretical value is estimated to be smaller [36]. In Fig. 2, the estimated quadrupole shifts for the clock transitions in Yb^+ using the calculated Θ values and compared with their corresponding experimental results are shown.

Clocks for Fundamental Physics

In the standard model (SM) of particle physics, there are several dimensionless "fundamental constants" such as gauge couplings whose constancies are not really predicted by the model itself, but they are accepted on some logical grounds without having sufficient scientific proof. However, this fact needs to be verified and any recognizable anomaly observed for all plausible circumstances with respect to variation in space or time that should be adequately explained by a fundamental theory. This persuades many to invent innovative notions of ideas to search for such signatures. One of the important and popular concepts is to presume dynamic

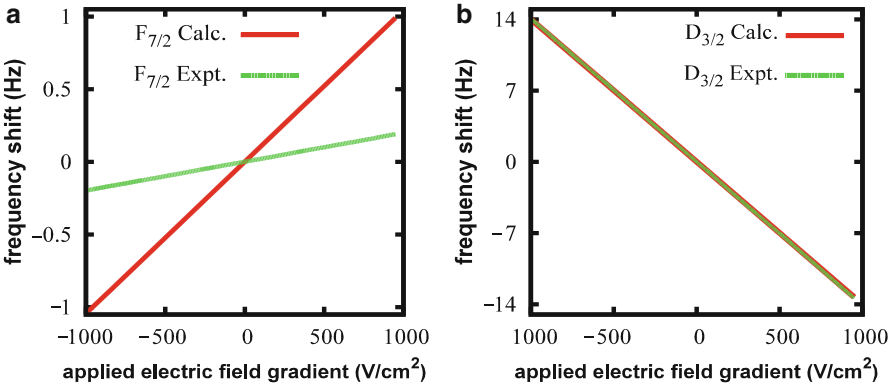


Fig. 2 Comparison between the quadrupole shifts estimated using Θ values from the measurement and calculations by the CC methods for (**a**) the $[4f^{14}6s]\,^2S_{1/2}(F=0) \rightarrow [4f^{13}6s^2]\,^2S_{7/2}(F=3)$ and (**b**) the $[4f^{14}6s]\,^2S_1(F=0) \rightarrow [4f^{14}5d]\,^2D_{3/2}(F=2)$ clock transitions in Yb^+

behavior of some of the physical constants, which should be probed by any either direct or indirect methods [45,46]. Since all the modern atomic clocks are capable of measuring atomic transition frequencies to ultrahigh-precision level, any observed discrepancies in that place of accuracy can be interrogated. It may be suggestive not to take values of the constants directly from their single observational methods; rather, they should be extracted from a series of theoretical and experimental studies concurrently. Determination of a self-consistent set of values of the fundamental constants giving best matching between theories and a defined set of experiments can address how much constancy the physical constants are really immanent within themselves. In fact, the test of variability of the physical constants does not require knowledge of their actual values to sufficient high precision in anticipation. A procedure for inferring any possible temporal variation of α_e from the clock frequency measurements has been elucidated here.

Let a physical quantity A be decomposed as $A = k_1 F_1 = k_2 F_2$ such as both k_1 and k_2 are two dimensionless quantities, while F_1 and F_2 are two functions of the physical constants among which F_1 is a function of the base units only (say) to some power. In this case the time variation of A (denoted by dA/dt) can be given by

$$\frac{d\ln A}{dt} = \frac{d\ln k_1}{dt} + \frac{d\ln F_1}{dt} = \frac{d\ln k_2}{dt} + \frac{d\ln F_2}{dt}, \tag{66}$$

such that, let us say, time variation of either dk_1/dt or dk_2/dt can be measured for which it is mandatory to have either $dF_1/dt = 0$ or $dF_2/dt = 0$. In this condition by measuring either dk_1/dt or dk_2/dt, it is possible to infer dA/dt. This is the basic underlying principle that can be adopted to probe $d\alpha_e/dt$. The ancillary attainment from the finding of $d\alpha_e/dt$ is that this information can be used for

extracting information on possible variation in quantum chromodynamics parameter $\Lambda_{\rm QCD}$ and electron (quark) mass $m_{e(q)}$ using the empirical relations [45,46]

$$\frac{\delta\Lambda_{\rm QCD}}{\Lambda_{\rm QCD}} \approx 34\frac{\delta\alpha_e}{\alpha_e} \tag{67}$$

and

$$\frac{\delta m_{e(q)}}{m_{e(q)}} \approx 70\frac{\delta\alpha_e}{\alpha_e}. \tag{68}$$

Since mass of the proton (m_p) is related with nuclear magnetic moment and is almost approximated to $m_p \approx 3\Lambda_{\rm QCD}$, so from $d\alpha_e/dt$ along with the above two relations, variation in the ratio m_p/m_e, strong to weak scale variation, can also be determined. Any possible signature in observing variation in the above physical quantities will be very useful to support physics describing by Grand Unified Theory (GUT) and other sophisticated models of particle physics and may also imply plausible violation of Einstein's equivalence principle [45,46].

The energy level in an atomic state of any multi-electron system is expressed as [47]

$$E_n \simeq c^2(Z\,\alpha_e)^2\left\{\frac{1}{2n^2} + \frac{(Z\alpha_e)^2}{2n^3}\left(\frac{1}{|k_n|} - \frac{3}{4n}\right)\right\}, \tag{69}$$

where $k_n = \pm\left(j_n + \frac{1}{2}\right)$ is the relativistic quantum number. Since the relativistic effects in an energy level close to the nucleus is large due to the high angular velocity of the electron and approximately given by

$$\Delta = -\frac{Z_a^2}{2}\frac{(Z\alpha_e)^2}{\nu^3}\left(\frac{1}{j_n + \frac{1}{2}} - \frac{Z_a}{Z\nu}\left[1 - \frac{Z_a}{4Z}\right]\right) \tag{70}$$

where ν is the effective principal quantum number and Z_a is the effective charge experienced by the electron in that energy level after accounting the screening effects due to the inner-core electrons. As the relativistic effects in the atomic energy levels scale in the order of α_e^2, the transition frequencies among the atomic levels are very sensitive to a small change in α_e and get enhanced with large atomic number Z and for a small value of ν. The advantage of considering atomic systems for probing variability of the physical constants is owing to the reason as any change due to a small variation in α_e in the atomic transition frequencies can be probed using the relativistic many-body methods by expressing a transition frequency (ω) as

$$\omega(x) \approx \omega_0 + q\,x, \tag{71}$$

where ω_0 is a referenced transition frequency with its corresponding fine-structure constant α_0 value, $x = \left(\frac{\alpha_e}{\alpha_0}\right) - 1$ is the Taylor coefficient of the first derivative of

ω, and $q = \frac{d\omega}{dx}\big|_{x=0}$ is known as the sensitivity coefficient due to variation in the α_e value. For the numerical estimate of the q-factor, it can be evaluated at the level of the first-order correction in α_e^2 using the expression

$$q \approx \frac{\omega(+x) - \omega(-x)}{2x} \tag{72}$$

for an arbitrary small value of x. By combining ultrahigh- precision frequencies measured in the optical frequency standards for a certain time interval with the calculated q-factors, it is possible to get some information on variation in α_e from a given reference. In the heavy atomic systems, the sensitivity parameters can be quite significant and can be of the order of one. Hence, it may be ideal to define another quantity (L_{α_e}) in the logarithmic form, to address the same issue, as

$$L_{\alpha_e} = \alpha_e \frac{\partial}{\partial \alpha_e} \ln[F_{\rm rel}(Z\alpha_e)] \tag{73}$$

where $F_{\rm rel}(Z\alpha_e)$ is the relativistic scaling factor in an atomic energy level. For the (optical) atomic transitions, it can be expressed as

$$L_{\alpha_e} = \frac{2q}{\omega_{\rm ref}}, \tag{74}$$

for the reference frequency $\omega_{\rm ref}$.

All the microwave clocks and some of the optical clocks, by the way, involve hyperfine transitions. The magnetic-dipole hyperfine structure constant $A_{\rm hf}$ of an atomic state can be analytically expressed as

$$A_{\rm hf} \propto \left[\frac{m_e e^4}{\hbar^2}\right] [\alpha_e^2 F_{\rm rel}(Z\alpha_e)] \left[\mu_B \frac{m_e}{m_p}\right], \tag{75}$$

where for the S-wave electron $F_{\rm rel}(Z\alpha_e) = \frac{3}{\sqrt{1-(Z\alpha_e)^2(3-(2Z\alpha_e)^2)}}$ and $L_{\alpha_e} = (Z\alpha_e)^2 \frac{11-(Z\alpha_e)^2}{(1-(Z\alpha_e)^2)(3-(Z\alpha_e)^2)}$. It can be shown that variation in α_e is related to variation in $F_{\rm rel}(Z\alpha_e)$ as

$$\frac{\delta F_{\rm rel}(Z\alpha_e)}{F_{\rm rel}(Z\alpha_e)} = K \frac{\delta \alpha_e}{\alpha_e}, \tag{76}$$

for the factor $K = \frac{(Z\alpha_e)^2}{(1-(Z\alpha_e)^2)(3-(2Z\alpha_e)^2)}$. Using this formula one obtains $K = 0.74$ for Cs, $K = 0.29$ for Rb, $K = 2.18$ for Hg^+, etc.; however, after expressing

$$A_{\rm hf} = \left(\frac{\alpha_e}{\alpha_0}\right)(A_{\rm hf}^{\rm ref} + qx) \tag{77}$$

Table 10 Preliminary results obtained by combining CC calculations with corresponding experimental frequency measurements from three important atomic clocks (Nandy DK, Private communication)

System	Transition	$K_{\rm rel}$	$\frac{1}{R}\frac{\partial R}{\partial t}$ (in year^{-1})
Yb^+	$[4f^{14}]6s \rightarrow 5d_{3/2}$	0.951	$(0.5 \pm 1.9) \times 10^{-16}$
Yb^+	$[4f^{14}]6s \rightarrow 4f^{13}6s^2\ ^2F_{7/2}$	−4.692	$(0.2 \pm 4.1) \times 10^{-16}$
Hg^+	$[5d^{10}]6s \rightarrow 5d^9 6s^2\ ^2D_{5/2}$	−2.88	$(3.7 \pm 3.9) \times 10^{-16}$

for the reference $A_{\rm hf}^{\rm ref}$ value, it is obtained as $K = 0.83$, $K = 0.34$, and $K = 2.28$ for Cs, Rb, and Hg^+, respectively [47].

The more straightforward relations that are convenient to use for inferring information on $d\alpha_e/dt$ from the combination of a probed transition (denoted by index 1) and a reference transition (referred by index 2) in the consideration of variety of transitions in an atomic system are given as [45,46]

$$\frac{\partial}{\partial t}\ln\left[\frac{f_{\rm hfs}^1}{f_{\rm hfs}^2}\right] \approx \frac{\partial}{\partial t}\ln\left[\frac{\mu_I^1}{\mu_I^1}\right] + \left(\frac{\partial F_{\rm rel}^1}{\partial \alpha_e} - \frac{\partial F_{\rm rel}^2}{\partial \alpha_e}\right)\frac{\partial \ln \alpha_e}{\partial t}, \tag{78}$$

$$\frac{\partial}{\partial t}\ln\left[\frac{f_{\rm atm}^1}{f_{\rm atm}^2}\right] \approx \frac{\partial}{\partial t}\ln\left(\frac{\partial F_{\rm rel}^1}{\partial \alpha_e} - \frac{\partial F_{\rm rel}^2}{\partial \alpha_e}\right)\frac{\partial \ln \alpha_e}{\partial t}, \tag{79}$$

and

$$\frac{\partial}{\partial t}\ln\left[\frac{f_{\rm hfs}^1}{f_{\rm atm}^2}\right] \approx \frac{\partial}{\partial t}\ln\left[\frac{\mu_I^1}{\mu_I^1} + (2 + L_{\rm hfs}^1 - L_{\rm atm}^2)\ln \alpha_e\right], \tag{80}$$

where the abbreviations hfs and atm represent, particularly, the hyperfine and atomic transitions, respectively. At least the following quantities

$$R = \frac{f_{\rm obs}}{f_{Cs}} \qquad \text{and} \qquad K_{\rm rel} = \frac{1}{F_{\rm rel}}\frac{\partial F_{\rm rel}}{\partial \alpha_e}$$

for three are clock transitions listed in Table 10 from the calculations using CC methods (Nandy DK, Private communication). Here Cs clock frequency standard is taken as the reference to compare with the observed values. Using these quantities, possible variation in α_e versus variation in nuclear magnetic moment μ_n inferred from the combinations of the above three clock transitions are demonstrated pictorially in Fig. 3.

Summary

Many basic concepts on working principles, needs, and roles of theoretical studies for atomic clocks are described for the general readers. It mainly highlights the essential criterion, present status, and future prospective of atomic clocks and

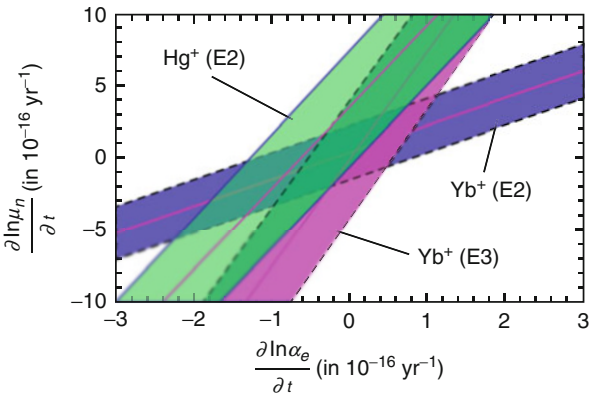


Fig. 3 Constraints on temporal variations in α_e and μ_n from the quadrupole (E2) clock transitions from Yb^+ and Hg^+ ions and octopole (E3) clock transition of Yb^+ ion with reference to Cs clock using the calculated q-parameters by CC method and available experimental data

mentions the candidates that are capable of replacing the present Cs primary frequency standard. Optical clocks based on the neutral atoms and a single trapped singly charged ions are discussed in great detail and their constraining factors to achieve high accuracy are demonstrated by considering a few cases. Uncertainties due to both the instrumental and external fields are also briefly given. In this context, the role of relativistic many-body methods for accurate estimate of uncertainties in some of the major systematics is discussed. Underlying differences between various many-body methods that are usually employed to calculate these properties are explained and few concrete examples are shown by giving results from these methods. The importance of higher-order relativistic effects is shown by evaluating their contributions explicitly in few cases. It is seen that in several cases, the theoretical calculations are more precise than the experimental results and in some situations both the experimental and theoretical results do not match. This, obviously, calls for developing more sophisticated many-body methods to validate these results more reliably. It is also emphasized how a priori theoretical studies are acting as the guiding factors for selecting new clock candidates. Finally, the need of studying atomic clocks, both theoretically and experimentally, for probing fundamental sciences is illustrated.

Acknowledgements The author is grateful to B. P. Das, D. K. Nandy, Y. Singh, and B. Arora for many fruitful discussions. The author also thanks J. Banerji for his critical reading of the chapter. Some of the preliminary calculations were carried out using the Physical Research Laboratory HPC cluster for the demonstration purposes.

References

1. Essen L, Parry JVL (1955) An atomic standard of frequency and time interval: a cæsium resonator. Nature 176:280–284. doi:10.1038/176280a0
2. Margolis HS (2009) Optical frequency standards and clocks. Contemp Phys 51:37–58. doi:10.1080/00107510903257616

3. Flambaum VV, Dzuba VA, Derevianko A (2008) Magic frequencies for cesium primary-frequency standard. Phys Rev Lett 101:220801-4. doi:10.1103/PhysRevLett.101.220801
4. Poli N, Oates CW, Gill P, Tino GM (2013) Optical atomic clocks. Rivista Del Nuovo Cimento 36:555–624. doi:10.1393/ncr/i2013-10095-x
5. Patrick G (2011) When should we change the definition of the second? Philos Trans R Soc A 369:4109–4130. doi:10.1098/rsta.2011.0237
6. Sahoo BK (2006) Relativistic coupled-cluster theory of quadrupole moments and hyperfine structure constants of 5d states in Ba+. Phys Rev A 74:020501(R)-4. doi:10.1103/PhysRevA.74.020501
7. Sherman JA, Trimble W, Metz S, Nagourney W, Fortson N (2005) Progress on indium and barium single ion optical frequency standards. Digest of the LEOS summer topical meetings (IEEE No. 05TH8797), e-print. arXiv:physics/0504013v2
8. Sahoo BK, Das BP, Chaudhuri R, Mukherjee D, Timmermans R, Jungmann K (2007) Investigations of Ra+ properties to test possibilities for new optical-frequency standards. Phys Rev A 76:040504(R)-4. doi:10.1103/PhysRevA.76.040504
9. Derevianko A, Dzuba VA, Flambaum VV (2012) Highly charged ions as a basis of optical atomic clockwork of exceptional accuracy. Phys Rev Lett 109:180801-180805. doi:10.1103/PhysRevLett.109.180801
10. Yudin VI, Taichenachev, Derevianko A (2014) Magnetic-dipole transitions in highly charged ions as a basis of ultraprecise optical clocks. Phys Rev Lett 113:233003-5. doi:10.1103/PhysRevLett.113.233003
11. Peik E, Tamm Chr (2003) Nuclear laser spectroscopy of the 3.5 eV transition in Th-229. Europhys Lett 61:181–186. doi:10.1209/epl/i2003-00210-x
12. Campbell CJ, Radnaev AG, Kuzmich A, Dzuba VA, Flambaum VV, Derevianko A (2012) Single-ion nuclear clock for metrology at the 19th decimal place. Phys Rev Lett 108:120802-5. doi:10.1103/PhysRevLett.108.120802
13. Schiller S, Bakalov D, Korobov VI (2014) Simplest molecules as candidates for precise optical clocks. Phys Rev Lett 113:023004-5. doi:10.1103/PhysRevLett.113.023004
14. Karr J-Ph (2014) H2+ and HD+: candidates for a molecular clock. J Mol Spectrosc 300:37–43. doi:10.1016/j.jms.2014.03.016
15. Akhiezer AJ, Berestetskii VB (1965) Quantum electrodynamics. Interscience, New York. Chap 8, Sec 20
16. Lindroth E, Ynnerman A (1993) Ab initio calculations of gj factors for Li, Be+, and Ba+. Phys Rev A 47:961–970. doi:10.1103/PhysRevA.47.961
17. Porsev SG, Derevianko A (2006) Multipolar theory of blackbody radiation shift of atomic energy levels and its implications for optical lattice clocks. Phys Rev A 74:020502-4. doi:10.1103/PhysRevA.74.020502
18. Arora B, Nandy DK, Sahoo BK (2012) Multipolar blackbody radiation shifts for single-ion clocks. Phys Rev A 85:012506-10. doi:10.1103/PhysRevA.85.012506
19. Flambaum VV, Ginges JSM (2005) Radiative potential and calculations of QED radiative corrections to energy levels and electromagnetic amplitudes in many-electron atoms. Phys Rev A 72:052115-13. doi:10.1103/PhysRevA.72.052115
20. Shavitt I, Bartlett RJ (2009) Many-body methods in chemistry and physics. Cambridge University Press, Cambridge
21. Kállay M, Nataraj HS, Sahoo BK, Das BP, Visscher L (2011) Relativistic general-order coupled-cluster method for high-precision calculations: Application to the Al+ atomic clock. Phys Rev A 83:030503-4(R). doi:10.1103/PhysRevA.83.030503
22. Sahoo BK (2008) An ab initio relativistic coupled-cluster theory of dipole and quadrupole polarizabilities: applications to a few alkali atoms and alkaline earth ions. Chem Phys Lett 448:144–149. doi:10.1016/j.cplett.2007.09.079
23. Sahoo BK, Das BP (2008)Relativistic coupled-cluster studies of dipole polarizabilities in closed-shell atoms. Phys Rev A 77:062516-5. doi:10.1103/PhysRevA.77.062516

24. Sahoo BK, Das BP, Mukherjee D (2009) Relativistic coupled-cluster studies of ionization potentials, lifetimes, and polarizabilities in singly ionized calcium. Phys Rev A 79:052511-9. doi:10.1103/PhysRevA.79.052511
25. Sahoo BK, Timmermans RGE, Das BP, Mukherjee D (2009) Comparative studies of dipole polarizabilities in Sr+, Ba+, and Ra+ and their applications to optical clocks. Phys Rev A 80:062506-10. doi:10.1103/PhysRevA.80.062506
26. Singh Y, Sahoo BK, Das BP (2013) Correlation trends in the ground-state static electric dipole polarizabilities of closed-shell atoms and ions. Phys Rev A 88:062504-11. doi:10.1103/PhysRevA.88.062504
27. Dalgarno A (1962) Atomic polarizabilities and shielding factors. Adv Phys 11:281–315. doi:10.1080/00018736200101302
28. Szabo A, Ostlund NS (1989) Modern quantum chemistry. Dover, New York
29. Peng D, Ma J, Liu W (2009) Int J Quantum Chem 109:2149–2167. doi:10.1002/qua.22078
30. Singh Y, Sahoo BK (2015) Rigorous limits on the hadronic and semileptonic CP-violating coupling constants from the electric dipole moment of Hg199. Phys Rev A 91, 030501-5. doi:10.1103/PhysRevA.91.030501
31. Nandy DK, Singh Y, Sahoo BK (2013) Development of a relativistic coupled-cluster method for one-electron detachment theory: application to Mn ix, Fe x, Co xi, and Ni xii ions. Phys Rev A 88:052512-12. doi:10.1103/PhysRevA.88.052512
32. Itano WM (2005) Quadrupole moments and hyperfine constants of metastable states of Ca+, Sr+, Ba+, Yb+, Hg+, and Au. Phys Rev A 73:022510-11. doi:10.1103/PhysRevA.73.022510
33. Sahoo BK, Sur C, Beier T, Das BP, Chaudhuri RK, Mukherjee D (2007) Enhanced role of electron correlation in the hyperfine interactions in D5/22 states in alkaline-earth-metal ions. Phys Rev A 75:042504-6. doi:10.1103/PhysRevA.75.042504
34. Sahoo BK, Das BP (2011) Parity nonconservation in ytterbium ion. Phys Rev A 84:010502-4(R). doi:10.1103/PhysRevA.84.010502
35. Versolato OO, Giri GS, Wansbeek LW, van den Berg JE, van der Hoek DJ, Jungmann K, Kruithof WL, Onderwater CJG, Sahoo BK, Santra B, Shidling PD, Timmermans RGE, Willmann L, Wilschut HW (2010) Laser spectroscopy of trapped short-lived Ra+ ions. Phys Rev A 82:010501-4. doi:10.1103/PhysRevA.82.010501
36. Nandy DK, Sahoo BK (2014) Quadrupole shifts for the Yb+171 ion clocks: Experiments versus theories. Phys Rev A 90:050503-5(R). doi:10.1103/PhysRevA.90.050503
37. Tommaseo G, Pfeil T, Revalde G, Werth G, Indelicato P, Desclaux JP (2003) The gJ-factor in the ground state of Ca +. Eur Phys J D 25:113–121. doi:10.1140/epjd/e2003-00096-6
38. Safronova MS, Kozlov MG, Clark CW (2012) Precision calculation of blackbody radiation shifts for optical frequency metrology. Phys Rev Lett 107:143006-5. doi:10.1103/PhysRevLett.107.143006
39. Singh Y, Sahoo BK (2014) Correlation trends in the polarizabilities of atoms and ions in the boron, carbon, and zinc homologous sequences of elements. Phys Rev A 90:022511-8. doi:10.1103/PhysRevA.90.022511
40. Sahoo BK (2014) Role of the multipolar black-body radiation shifts in the atomic clocks at the 10–18 uncertainty level. Pramana 83:255–263. doi:10.1007/s12043-014-0791-9
41. Arora A, Sahoo BK (2012) State-insensitive trapping of Rb atoms: linearly versus circularly polarized light. Phys Rev A 86:033416-12. doi:10.1103/PhysRevA.86.033416
42. Sahoo BK, Arora B (2013) Magic wavelengths for trapping the alkali-metal atoms with circularly polarized light. Phys Rev A 87:023402-9. doi:10.1103/PhysRevA.87.023402
43. Sur C, Latha KVP, Sahoo BK, Chaudhuri RK, Das BP, Mukherjee D (2006) Electric quadrupole moments of the D states of alkaline-earth-metal ions. Phys Rev Lett 96:193001-4. doi:10.1103/PhysRevLett.96.193001
44. Blythe PJ, Webster SA, Hosaka K, Gill P (2003) Systematic frequency shifts of the 467 nm electric octupole transition in 171Yb+. J Phys B: At Mol Opt Phys. 36:981–989. doi:10.1088/0953-4075/36/5/317

45. Uzan J-P (2003) The fundamental constants and their variation: observational and theoretical status. Rev Mod Phys 75:403–455. doi:10.1103/RevModPhys.75.403
46. Kolachevsky NK (2004) Laboratory search for time variation in the fine structure constant. Phys-Usp 47:1101–1108. doi:10.1070/PU2004v047n11ABEH001870
47. Dzuba VA, Flambaum VV, Webb JK (1999) Calculations of the relativistic effects in many-electron atoms and space-time variation of fundamental constants. Phys Rev A 59:230–237. doi:10.1103/PhysRevA.59.230

Relativistic Theories of NMR Shielding

21

Yunlong Xiao, Wenjian Liu, and Jochen Autschbach

Contents

Introduction . . . 658
Notation . . . 661
Theoretical Description of NMR Shielding . . . 661
The Effective Nuclear Spin Hamiltonian in Experimental Picture . . . 661
The Effective Nuclear Spin Hamiltonian in Theoretic Picture . . . 662
The Body-Fixed Molecular Hamiltonian . . . 663
Further Simplification of the First-Order Operators . . . 667
The Model Space V_{mod} . . . 668
Construction of the Effective Nuclear Spin Hamiltonian . . . 669
The Relativistic Diamagnetism . . . 672
The Gauge Origin Problem . . . 677
The MB-GIAO-DKS Approach . . . 678
General Formulation . . . 678
Application to Nuclear Shielding . . . 681
Two-Component Relativistic Theory of Nuclear Shielding . . . 682
The X2C Equations . . . 682
The X2C Approach for NMR Shielding . . . 685
The ZORA Approach for NMR Shielding . . . 688
Summary . . . 689
References . . . 690

Y. Xiao (✉) • W. Liu
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: xiaoyl@pku.edu.cn; liuwjbdf@gmail.com

J. Autschbach
Department of Chemistry, University at Buffalo, State University of New York, Buffalo, NY, USA
e-mail: jochena@buffalo.edu

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_17

Abstract

The theoretical picture of nuclear magnetic resonance (NMR) shielding is first discussed in depth through the body-fixed molecular Hamiltonian that treats electrons fully relativistically and nuclei quasi-relativistically. Various ways are then presented to capture the relativistic diamagnetism, which is otherwise missed by the standard four-component relativistic linear response theory of NMR shielding. With an explicit relativistic diamagnetism, the correct nonrelativistic limit of the four-component relativistic theory can be guaranteed even with a finite basis. The gauge origin problem can further be resolved by means of the idea of distributed gauge origins. The exact two-component counterpart can also be formulated in a simple way.

Keywords

Nuclear shielding • Relativistic molecular Hamiltonian • Four-component • Exact two-component • Diamagnetism • Magnetic balance • Gauge-including atomic orbital

Introduction

The nonrelativistic theory of nuclear magnetic resonance (NMR) shielding was formulated by Ramsey [1] in 1950. It can nowadays be applied at various levels of methodology for analyzing the electronic and molecular structures of medium-size to large molecules. Yet, since NMR shielding samples the electronic wave function in the vicinity of an NMR active nucleus, it is an intrinsic all-electron, relativistic property. The four-component relativistic theory of NMR shielding was formulated [2–4] between the 1960s and 1980s but was not much used during that period of time. This is not only because the theory was computationally too expensive but also because the theory itself was not yet put into an appropriate form [5, 6] (vide post). Because of this, various approximate two-component (A2C) relativistic theories became popular since the mid-1990s, including first-order perturbation theory [7–10], zeroth-order regular approximation (ZORA) [11,12], Douglas-Kroll-Hess (DKH)-type approximation [13, 14], and second-order regular approximation [15, 16] to the normalized elimination of the small component (NESC) approach [17]. Since the approximations inherent in such A2C theories all result from the atomic cores, it is clear that they cannot provide accurate absolute NMR shielding, especially of heavy atoms. For instance, it was found [18] that ZORA underestimates the isotropic shielding of ^{199}Hg, which is on the order of 10^4 ppm in typical mercury molecules, by over 1,000 ppm compared to four-component relativistic calculations. Nevertheless, such A2C theories work well for relative shielding (chemical shift δ), viz.,

$$\delta = \frac{\sigma^{\mathrm{ref}} - \sigma}{1 - \sigma^{\mathrm{ref}}} \simeq \sigma^{\mathrm{ref}} - \sigma. \tag{1}$$

Table 1 Summary of the acronyms

Acronym	Full name
NMR	Nuclear magnetic resonance, 2, 3, 4, 13, 26, 28, 29, 30, 31
LAB	Laboratory frame of reference, 4, 5, 6, 7, 8, 9, 11, 12, 13
BFF	Body-fixed frame of reference, 3, 6, 7, 8, 9, 10, 12, 14
DKS	Dirac-Kohn-Sham, 3, 20, 21, 23
4C	Four-component, 30
X2C	Exact two-component, 3, 23, 25, 26, 29, 30
NESC	Normalized elimination of the small component, 1, 26
SESC	Symmetrized elimination of the small component, 26
UESC	Unnormalized elimination of the small component, 25
A2C	Approximate two-component, 1, 3, 30
ZORA	Zeroth-order regular approximation, 1, 3, 29, 30
DKH	Douglas-Kroll-Hess approximation, 1, 3, 29, 30
LRT	Standard linear response theory (without magnetic balance), 18
ODA	Orbital decomposition approach, 16, 17, 18
EFUT	External field-dependent unitary transformation, 17, 18, 19, 22
FFUT	Full field-dependent unitary transformation, 16
MB	Magnetic balance, 3, 19, 20, 30
RMB	Restricted magnetic balance, 17, 18
GIAO	Gauge-including atomic orbital, 3, 18, 19, 20, 22, 30
ppm	Part per million, 1, 13
PES	Positive energy states, 15, 25, 28
NES	Negative energy states, 14, 17, 23, 28
$\vec{A}^{10}$	Vector potential of the uniform external magnetic field, 14, 16, 17, 18
$\vec{A}^{01}$	Vector potential of the nuclear magnetic dipole moment, 14, 17, 18

Here, σ^{ref} is the shielding of a reference with the same nuclear isotope of interest (^{199}Hg, say). Due to the subtraction of the shielding constants, core orbital contributions to the shielding, along with their relativistic effects, essentially cancel in the chemical shift. This is of course the reason why chemical shifts are sensitive to the chemical environment of a nucleus – their trends are determined by the valence orbitals. Yet, relativistic effects can be transferred via valence orbitals and chemical bonds from a heavy atom to a neighboring heavy or light atom. The nomenclature HAHA for relativistic ("heavy atom") effects on the heavy atom's NMR parameters, and HALA for relativistic effects induced by a heavy atom on a light atom's NMR parameters [19], has become quite common in the field. Despite the cancelation of core orbital contributions, both HAHA and HALA effects on chemical shifts can be very large, and therefore, relativistic methods should be used for NMR calculations of heavy element compounds.

The importance of valence orbitals for NMR shifts (and also NMR J-coupling [20]) implies that relativistic pseudopotentials for heavy atoms can also be employed effectively for chemical shift calculations of the HALA type [21]. In this case, an all-electron relativistic method is not vital – although potentially more accurate.

However, without further modifications, pseudopotentials cannot be used for heavy atom chemical shifts or J-coupling involving a heavy atom. Analysis has shown that the near-nucleus nodal structure of the valence orbitals is chiefly responsible for HAHA effects [22]. That is, the effects are valence shell effects in terms of orbitals but core and semi-core effects in terms of where they originate spatially. If the all-electron nodal structures of the valence orbitals are reconstructed from the pseudo-orbitals of a pseudopotential calculation, it is indeed possible to calculate heavy atom chemical shifts and relativistic effects thereupon properly [23]. However, at the time of writing this chapter, such methods are in limited use for heavy element chemical shift calculations. Likewise, it is possible to use frozen core orbitals for NMR calculations [22] (but of course not for absolute shielding). However, the valence orbital basis sets have to be very flexible in the core regions in order to recover the relativistic effects properly, with the consequence that the gain in computational efficiency is limited.

Considering the paramount role of valence orbitals for chemical shifts and J-coupling, it is then not surprising that variational A2C methods such as ZORA and DKH2 have been in use for a long time for NMR calculations [20, 24]. Such methods are often used in combination with density functional theory (DFT) for treating electron correlation. In this realm of application, with efficient but not necessarily the most accurate approximations for electron correlation, basis sets of moderate quality, and other aspects of the computational model to worry about (e.g., solvent effects), variational A2C approaches can be considered "good enough" for most purposes. Yet, there exist good reasons for treating relativistic effects more accurately. First, the NMR shielding is a fundamental physical observable. If absolute shielding constants can be calculated accurately, accurate chemical shift will be obtained naturally instead of as a result of error compensation. Second, a more accurate relativistic theory is not necessarily more expensive than the A2C ones. It will be shown here that, when formulated properly, both four-component [6] and exact two-component (X2C) [25, 26] relativistic theories of NMR shielding are operationally very much the same as the A2C ones, but more accurate, such as to render the latter in principle obsolete. Indeed, the existing A2C codes can sometimes be easily retrofitted for X2C calculations [27].

This chapter is composed in the following way. The notations to be adopted are first outlined in section "Notation." Section "Theoretical Description of NMR Shielding" is devoted to the proper theoretical definition of NMR shielding. To the best of our knowledge, this has not been documented before in the literature. The four-component relativistic theory of NMR shielding is reviewed in section "The Relativistic Diamagnetism," where we focus on the question why the diamagnetic term appearing in the nonrelativistic theory does not show itself in the four-component treatment and how to regain it appropriately. Then, the gauge origin problem is discussed in section "The Gauge Origin Problem." The working equations of NMR shielding within the framework of Dirac-Kohn-Sham (DKS) are given in section "The MB-GIAO-DKS Approach." The X2C theory is reviewed in section "Two-Component Relativistic Theory of Nuclear Shielding." Finally, the summary is given in section "Summary."

Notation

The System International-based atomic units are used throughout, in which $\hbar = M_e = e = 1$ and the speed of light c is approximately 137. The charge and mass of nucleus K are denoted by Z_K and m_K, respectively. Unless designated explicitly, the spatial coordinates of the nuclei ($\vec{R}_K$) and electrons ($\vec{r}_i$) refer to the body-fixed frame (BFF) of reference, the origin of which is placed at the nuclear center of mass (NCM). The spatial displacements are defined as $\vec{R}_{KL} = \vec{R}_K - \vec{R}_L$ or $\vec{r}_{iK} = \vec{r}_i - \vec{R}_K$. The three Cartesian components of a vector are denoted by u, v, and w. The Pauli representation is employed for both nuclear and electronic spins. For nonlinear molecules, the inertia tensors I^0, I', and I'' are defined in Eqs. (33), (29), and (34) in Ref. [28], respectively. For linear molecules, the inertia numbers $\mathcal{I}^0$, $\mathcal{I}'$, and $\mathcal{I}''$ are defined in Eqs. (27), (26), and (28) in Ref. [29], respectively. The occupied and virtual orbitals are denoted by $\{|i\rangle, |j\rangle\}$ and $\{|a\rangle, |b\rangle\}$, respectively, or by $\{|p\rangle, |q\rangle\}$ collectively. The Einstein summation convention over repeated indices is always employed.

Theoretical Description of NMR Shielding

The Effective Nuclear Spin Hamiltonian in Experimental Picture

NMR shielding is a parameter for defining an effective (mixed second-order) nuclear spin Hamiltonian $H_{\text{eff}}^{\text{NS}}$,

$$H_{\text{eff}}^{\text{NS}} = -\sum_K \mathcal{B}_{\text{eff},v}^{\text{LAB}(K)} \mu_{K,v}, \quad \vec{\mu}_K = g_K \mu_N \vec{S}_K, \tag{2}$$

$$= -\sum_K (B_{\text{ext},v}^{\text{LAB}} + \mathcal{B}_{\text{ind},v}^{\text{LAB}(K)}) \mu_{K,v}, \tag{3}$$

$$\mathcal{B}_{\text{ind},v}^{\text{LAB}(K)} = -B_{\text{ext},u}^{\text{LAB}} \sigma_{K,uv}^{\text{LAB}}. \tag{4}$$

Here, $\vec{\mu}_K$ is the magnetic momentum operator of nucleus K, with the spin angular momentum $\vec{S}_K$ quantized in the laboratory (LAB) frame of reference, while the quantity $\vec{\mathcal{B}}_{\text{ind}}^{\text{LAB}(K)}$ in Eq. (3) can be understood as the electronically induced magnetic field $\vec{B}_{\text{ind}}^{\text{LAB}(K)}(\{\vec{R}\})$ further averaged over the nuclear vibrational and rotational degrees of freedom. Consider a simple case where only the z-component of $\vec{B}_{\text{ext}}^{\text{LAB}}$ is nonvanishing and only nucleus K in the molecule has spin. Further assume $\mathcal{B}_{\text{eff},z}^{\text{LAB}(K)}$ is the only nonvanishing component and the spin states $|\Phi_i\rangle$ of nucleus K are eigenstates of both $S_{K,z}$ and $H_{\text{eff}}^{\text{NS}}$, viz.,

$$S_{K,z}|\Phi_i\rangle = M_i|\Phi_i\rangle, \quad H_{\text{eff}}^{\text{NS}}|\Phi_i\rangle = E_i|\Phi_i\rangle, \quad E_i = -g_K \mu_N M_i \mathcal{B}_{\text{eff},z}^{\text{LAB}(K)}. \tag{5}$$

It is seen that the energy levels E_i are proportional to the strength of the effective magnetic field and are equally separated for the absolute difference between the spin projections M_{i+1} and M_i is always one. Therefore, the transition frequency ν_L between any two adjacent states reads

$$\Delta E = g_K \mu_N \mathcal{B}_{\text{eff},z}^{\text{LAB}(K)} = h\nu_L. \tag{6}$$

As the nuclear magneton $\mu_N = \frac{e\hbar}{2m_p}$ is 1,836 times smaller than the Bohr magneton $\mu_B = \frac{e\hbar}{2M_e}$ and the magnetic field $\vec{B}_{\text{ext}}^{\text{LAB}}$ is very weak even in modern experimental setup [NB: this is why the induced field $\vec{\mathcal{B}}_{\text{ind}}^{\text{LAB}(K)}$ (3) has been truncated only to the linear term (4), with the proportionality $\sigma_{K,uv}^{\text{LAB}}$ being the nuclear shielding], the energy separation ΔE is in the radiofrequency domain, much smaller than the electronic excitation energies of a closed-shell molecule. That is, what is described here is a nucleus-electron composite system where the (quantized) nuclei are in the hyperfine excited spin states, while the electrons in the ground state. The shielding $\sigma_{K,uv}^{\text{LAB}}$ can be deduced from the measured transition frequencies ν_L or equivalently the strength of the magnetic field $\vec{\mathcal{B}}_{\text{eff}}^{\text{LAB}(K)}$. Note in passing that in this "experimental picture," the nuclei are considered as internal particles, whereas the electrons as external particles. To calculate the nuclear shielding $\sigma_{K,uv}^{\text{LAB}}$, we must go to a "theoretical picture," where the role of electrons and nuclear spins are reversed.

The Effective Nuclear Spin Hamiltonian in Theoretic Picture

To see how to construct the nuclear spin effective Hamiltonian $H_{\text{eff}}^{\text{NS}}$ (3), we first briefly review the concept of effective Hamiltonian. Consider the following Hamiltonian:

$$H = H^{00} + V^{10} + V^{01} + V^{11} + \cdots, \tag{7}$$

with the zeroth-order problem known as

$$H^{00}|p^{00}\rangle = E_p^{00}|p^{00}\rangle. \tag{8}$$

We further assume that there exist degeneracies in the zeroth-order space, but only a specific n_d-fold degenerate subspace V_{mod} is to be considered, i.e.,

$$E_p^{00} = \text{const.}, \quad p = 1, \ldots, n_d. \tag{9}$$

The degeneracy of the manifold V_{mod} will split in the presence of perturbations V^{10}, V^{01}, and V^{11}. The energies can be obtained exactly by solving the reduced eigenvalue problem

$$H_{\text{eff}}|q\rangle = E_q|q\rangle, \tag{10}$$

where the effective Hamiltonian H_{eff} is defined in V_{mod}, whose eigenvectors $|q\rangle$ are just linear combinations of $\{|p^{00}\rangle\}$. For the present purpose, it is the mixed second-order term of H_{eff} that is wanted, viz.,

$$H_{\text{eff}}^{\text{NS}} = H_{\text{eff}}^{11} = H_{\text{eff}}^{11,d} + H_{\text{eff}}^{11,p}, \tag{11}$$

$$H_{\text{eff}}^{11,d} = P^{00} V^{11} P^{00}, \tag{12}$$

$$H_{\text{eff}}^{11,p} = P^{00} V^{10} Q^{00} \frac{1}{E^{00} - H^{00}} Q^{00} V^{01} P^{00} + h.c., \tag{13}$$

where the project operators are defined as

$$P^{00} = \sum_{p \in V_{\text{mod}}} |p^{00}\rangle\langle p^{00}|, \tag{14}$$

$$Q^{00} = 1 - P^{00} = \sum_{q \notin V_{\text{mod}}} |q^{00}\rangle\langle q^{00}|. \tag{15}$$

Once the Hamiltonian (7) is specified (see section "The Body-Fixed Molecular Hamiltonian"), the desired effective Hamiltonian H_{eff}^{11} (11) can be constructed accordingly.

The Body-Fixed Molecular Hamiltonian

The point of departure is the following molecular Hamiltonian [28] that is written in the LAB frame and treats electrons fully relativistically and nuclei quasi-relativistically:

$$\begin{aligned} H^{\text{LAB}} = {} & T_{n0}^{\text{LAB}} + T_{n2}^{\text{LAB}} + T_{e}^{\text{LAB}} + H_{ee}^{\text{LAB}} + H_{ne}^{\text{LAB}} + H_{nn}^{\text{LAB}} \\ & + H_{eB}^{\text{LAB}} + H_{nB}^{\text{LAB}}, \end{aligned} \tag{16}$$

where the first two terms represent the respective nonrelativistic nuclear kinetic energy and its first-order relativistic correction, while the following four terms are the electronic kinetic energy, electron-electron, nucleus-electron, and nucleus-nucleus interactions, respectively. These terms have been documented before [28] and are hence not repeated here. The last two terms represent the interactions of the electronic current ($c\vec{\alpha}$) and nuclear spin ($\vec{\mu}_K$) with the uniform external magnetic field $\vec{B}_{\text{ext}}^{\text{LAB}}$, viz.,

$$H_{eB}^{\text{LAB}} = \frac{1}{2} \sum_{i=1} B_{\text{ext},u}^{\text{LAB}} (r_{i,v}^{\text{LAB}} - \tilde{R}_{g,v}^{\text{LAB}}) c\alpha_{i,w} \epsilon_{uvw} =: H_{10}^{\text{LAB}}, \tag{17}$$

$$H_{nB}^{\text{LAB}} = - \sum_{K=1} \mu_{K,u} B_{\text{ext},u}^{\text{LAB}} =: H_{11}^{\text{LAB}}, \tag{18}$$

where $\vec{\tilde{R}}_g^{\mathrm{LAB}}$ is the gauge origin of $\vec{B}_{\mathrm{ext}}^{\mathrm{LAB}}$ and ϵ_{uvw} the Levi-Civita permutation symbol. The third perturbation operator H_{01}^{LAB}, which is one of the terms in H_{ne}^{LAB} (see Eq. (92) in Ref. [28]), describes the interaction between the nuclear spin and the electronic current:

$$H_{01}^{\mathrm{LAB}} = \epsilon_{vv'w} \frac{\mu_{K,v} r_{iK,v'}^{\mathrm{LAB}}}{c^2 (r_{iK}^{\mathrm{LAB}})^3} c\alpha_w. \tag{19}$$

By using the identities

$$r_{i,v}^{\mathrm{LAB}} = R_{v'v} r_{iv'}^{\mathrm{BFF}} + R_{\mathrm{NCM},v}^{\mathrm{LAB}}, \tag{20}$$

$$R_{K,v}^{\mathrm{LAB}} = R_{v'v} R_{Kv'}^{\mathrm{BFF}} + R_{\mathrm{NCM},v}^{\mathrm{LAB}}, \tag{21}$$

the H_{10}^{LAB} (17) and H_{01}^{LAB} (19) operators can be rewritten as

$$H_{10}^{\mathrm{LAB}} = \frac{1}{2} \sum_{i=1} B_{\mathrm{ext},u}^{\mathrm{LAB}} (R_{v'v} r_{iv'}^{\mathrm{BFF}} + R_{\mathrm{NCM},v}^{\mathrm{LAB}} - \tilde{R}_{g,v}^{\mathrm{LAB}}) c\alpha_{i,w} \epsilon_{uvw}, \tag{22}$$

$$H_{01}^{\mathrm{LAB}} = \epsilon_{vv'w} \frac{\mu_{K,v} r_{iK,v''}^{\mathrm{BFF}}}{c^2 (r_{iK}^{\mathrm{BFF}})^3} R_{v''v'} c\alpha_w. \tag{23}$$

The real orthogonal cosine matrix R in Eq. (20) is the rotation of the spatial space from LAB to BFF:

$$R(\alpha, \beta, \gamma) = R_z(\gamma) R_y(\beta) R_z(\alpha), \tag{24}$$

$$R_z(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix},$$

$$R_z(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{25}$$

where α, β, and γ are the Euler angles (to be denoted collectively as s). Because of the relation

$$M_{\mathrm{nuc}} R_{\mathrm{NCM},u}^{\mathrm{LAB}} + \sum_i M_e (R_{\mathrm{NCM},u}^{\mathrm{LAB}} + R_{vu} r_{iv}^{\mathrm{BFF}}) = M_{\mathrm{mol}} R_{\mathrm{MCM},u}^{\mathrm{LAB}}, \tag{26}$$

the nuclear center of mass $\vec{R}^{\mathrm{LAB}}_{\mathrm{NCM}}$ in Eq. (20) should be understood as

$$R^{\mathrm{LAB}}_{\mathrm{NCM},u} = R^{\mathrm{LAB}}_{\mathrm{MCM},u} - \frac{M_e}{M_{\mathrm{mol}}} \sum_i R_{vu} r^{\mathrm{BFF}}_{iv}, \tag{27}$$

where $\vec{R}^{\mathrm{LAB}}_{\mathrm{MCM}}$ is the molecular center of mass (MCM). One particular disadvantage of the Hamiltonian H^{LAB} (16) lies in that the electronic and nuclear degrees of freedom are strongly coupled, given very different energy scales. To minimize such couplings, the BFF molecular Hamiltonian H^{BFF} should be introduced [28]:

$$H^{\mathrm{BFF}} = \mu^{-1/4} U_{eS} H^{\mathrm{LAB}} U^{-1}_{eS} \mu^{1/4}, \tag{28}$$

where μ is the determinant of the reciprocal of the effective inertial tensor, a complicated function of normal coordinates (see Eq. (28) in Ref. [28]). Since it commutes with all the perturbation operators, H^{LAB}_{01}, H^{LAB}_{10}, and H^{LAB}_{11}, it is actually ineffective for the present purpose. The U_{eS} operator in Eq. (28) is defined as

$$U_{eS} = e^{i\gamma S^e_z} e^{i\beta S^e_y} e^{i\alpha S^e_z}, \quad \vec{S}^e = \sum_i \vec{S}_i, \quad \vec{S}_i = \frac{1}{2}\vec{\Sigma}_i, \quad \vec{\Sigma}_i = \begin{pmatrix} \vec{\sigma}_i & 0 \\ 0 & \vec{\sigma}_i \end{pmatrix}. \tag{29}$$

Under such a transformation, the Dirac matrices are transformed as

$$U_{eS} \alpha_u U^{-1}_{eS} = \alpha_v R_{vu}. \tag{30}$$

The U_{eS} transformations of H^{LAB}_{10} (22) and H^{LAB}_{01} (23) then read

$$\begin{aligned} H^{\mathrm{BFF}}_{10} &= U_{eS} H^{\mathrm{LAB}}_{10} U^{-1}_{eS} \\ &= \frac{1}{2} \sum_{i=1} B^{\mathrm{LAB}}_{\mathrm{ext},u} (R_{v'v} r^{\mathrm{BFF}}_{iv'} + R^{\mathrm{LAB}}_{\mathrm{NCM},v} - \tilde{R}^{\mathrm{LAB}}_{g,v}) c\alpha_{i,w'} R_{w'w} \epsilon_{uvw}, \end{aligned} \tag{31}$$

$$\begin{aligned} H^{\mathrm{BFF}}_{01} &= U_{eS} H^{\mathrm{LAB}}_{01} U^{-1}_{eS} \\ &= \epsilon_{vv'w} \frac{\mu_{K,v} r^{\mathrm{BFF}}_{iK,v''}}{c^2 (r^{\mathrm{BFF}}_{iK})^3} R_{v''v'} c\alpha_{w'} R_{w'w}. \end{aligned} \tag{32}$$

The second-order perturbation operator H^{LAB}_{11} (18) commutes with U_{eS}, so it remains unchanged:

$$H^{\mathrm{BFF}}_{11} = H^{\mathrm{LAB}}_{11}. \tag{33}$$

To further simplify H^{BFF}_{10} (31), we invoke an additional perturbation-dependent unitary transformation to H^{BFF} (28):

$$H = U_g H^{\mathrm{BFF}} U_g^{-1}, \tag{34}$$

$$U_g = e^{iT_g}, \tag{35}$$

$$T_g = -\frac{1}{2}\epsilon_{uvw} B^{\mathrm{LAB}}_{\mathrm{ext},u}(-R^{\mathrm{LAB}}_{\mathrm{NCM},v} + \tilde{R}^{\mathrm{LAB}}_{g,v} - R_{v'v} R^{\mathrm{BFF}}_{g,v'}) r_{iw'} R_{w'w}, \tag{36}$$

where $\vec{R}^{\mathrm{BFF}}_g$ stands for the common gauge origin with constant components in the BFF and has nothing to do with $\vec{\tilde{R}}^{\mathrm{LAB}}_g$ introduced in Eq. (17). It is the Hamiltonian H (34) that is to be used for Eq. (7). By construction, T_g (36) commutes with H^{BFF}_{01} (32). Therefore,

$$V^{01} = H^{\mathrm{BFF}}_{01} = \epsilon_{vv'w} \frac{\mu_{K,v} r^{\mathrm{BFF}}_{iK,v''}}{c^2 (r^{\mathrm{BFF}}_{iK})^3} R_{v''v'} c\alpha_{w'} R_{w'w}, \tag{37}$$

$$V^{11} = H^{\mathrm{BFF}}_{11} + [iT_g, H^{\mathrm{BFF}}_{01}] = H^{\mathrm{BFF}}_{11} = -\sum_{K=1} \mu_{K,u} B^{\mathrm{LAB}}_{\mathrm{ext},u}. \tag{38}$$

In contrast, the V^{10} operator in Eq. (7) becomes

$$V^{10} = H^{\mathrm{BFF}}_{10} + [iT_g, H^{\mathrm{BFF}}_{00}] \tag{39}$$

$$= H^{\mathrm{BFF}}_{10} + c\alpha_{iw'}[iT_g, p^{\mathrm{BFF}}_{iw'}] + \mathcal{O}(\frac{M_e}{M_{\mathrm{nuc}}}), \tag{40}$$

where the $\mathcal{O}(\frac{M_e}{M_{\mathrm{nuc}}})$ terms arise from the fact that T_g (36) involves $R(\alpha, \beta, \gamma)$ and $R^{\mathrm{LAB}}_{\mathrm{NCM},u}$, which do not commute with the derivatives $\frac{\partial}{\partial s}$, $\frac{\partial}{\partial Q_k}$, and $\frac{\partial}{\partial R^{\mathrm{LAB}}_{\mathrm{MCM}}}$ in H^{BFF}_{00}. The commutator in Eq. (40) can be calculated as

$$\left[iT_g, p^{\mathrm{BFF}}_{iw'}\right] = \frac{1}{2}\epsilon_{uvw} B^{\mathrm{LAB}}_{\mathrm{ext},u}(-R^{\mathrm{LAB}}_{\mathrm{NCM},v} + \tilde{R}^{\mathrm{LAB}}_{g,v} - R_{v'v} R^{\mathrm{BFF}}_{g,v'}) R_{w'w} + \mathcal{O}(\frac{M_e}{M_{\mathrm{nuc}}}), \tag{41}$$

where the $\mathcal{O}(\frac{M_e}{M_{\mathrm{nuc}}})$ term stems from the commutator $[R^{\mathrm{LAB}}_{\mathrm{NCM},u}, p^{\mathrm{BFF}}_{iu}]$, with $R^{\mathrm{LAB}}_{\mathrm{NCM},u}$ (27) expressed in terms of independent variables $R^{\mathrm{LAB}}_{\mathrm{MCM},u}$, α, β, γ, and r^{BFF}_{iv}. Neglecting all the $\mathcal{O}(\frac{M_e}{M_{\mathrm{nuc}}})$ terms, V^{10} (40) can be simplified to

$$V^{10} = \frac{1}{2}\sum_{i=1} B^{\mathrm{LAB}}_{\mathrm{ext},u} R_{v'v}(r^{\mathrm{BFF}}_{iv'} - R^{\mathrm{BFF}}_{gv'}) c\alpha_{i,w'} R_{w'w} \epsilon_{uvw}. \tag{42}$$

In sum, the U_g transformation (35) has effectively shifted the gauge origin from $\vec{\tilde{R}}^{\mathrm{LAB}}_g$ (a constant vector in the LAB frame) to $\vec{R}^{\mathrm{BFF}}_g$ (a constant vector in the BFF).

Further Simplification of the First-Order Operators

For a real orthogonal matrix O and two arbitrary vectors $\vec{U}$ and $\vec{V}$, we have the following identity:

$$\epsilon_{vuw'} U_v V_u O_{w'w} = \epsilon_{vuw} U_{v'} O_{v'v} V_{u'} O_{u'u}. \tag{43}$$

It follows that the spatial transformation of the cross product $\vec{U} \times \vec{V}$ is equivalent to first transform $\vec{U}$ and $\vec{V}$, separately, and then make the cross product. With the help of Eq. (43), the expression (37) can be simplified to

$$V^{01} = \sum_K D_{v'}^{01(K)} R_{v'v} \mu_{K,v}, \tag{44}$$

$$D_v^{01(K)} = \sum_i \frac{\left(\vec{r}_{iK}^{\,\mathrm{BFF}} \times c\vec{\alpha}_i\right)_v}{c^2 (r_{iK}^{\mathrm{BFF}})^3}, \tag{45}$$

where $\vec{D}^{01(K)}$ is independent of the Euler angles. Similarly, V^{10} (42) can be simplified to

$$V^{10} = R_{u'u} D_{u'}^{10} B_{\mathrm{ext},u}^{\mathrm{LAB}}, \tag{46}$$

$$D_u^{10} = \frac{1}{2} \sum_i (\vec{r}_{ig}^{\,\mathrm{BFF}} \times c\vec{\alpha}_i)_u, \tag{47}$$

where $\vec{D}^{10}$ is also independent of the Euler angles.

It is interesting to note that $D_v^{01(K)}$ (45) can formally be written as

$$D_v^{01(K)} = \frac{\partial \sum_i c\vec{\alpha}_i \cdot \vec{A}_i^{C,01(K)}}{\partial \mu_{K,v}^{C,01}}, \tag{48}$$

where $\vec{\mu}_K^{C,01}$ is a classical (C) nuclear magnetic momentum with constant components in BFF and $\vec{A}_i^{C,01(K)}$ is the corresponding vector potential

$$\vec{A}_i^{C,01(K)} = \frac{\vec{\mu}_K^{C,01} \times \vec{r}_{iK}^{\,\mathrm{BFF}}}{c^2 \left(r_{iK}^{\mathrm{BFF}}\right)^3}. \tag{49}$$

This represents the usual "theoretical picture," where the electrons are treated as internal quantum particles, whereas the clamped nuclei as external classical particles. Likewise, the $\vec{D}^{10}$ operator in Eq. (47) can formally be written as

$$D_u^{10} = \frac{\partial \sum_i c\vec{\alpha}_i \cdot \vec{A}_i^{A,10}}{\partial B_u^{A,10}}, \tag{50}$$

where $\vec{B}^{A,10}$ is an artificial (A) magnetic field with constant components in BFF and $\vec{A}_i^{A,10}$ is the corresponding vector potential with the gauge origin at $\vec{R}_g^{\mathrm{BFF}}$:

$$\vec{A}_i^{A,10} = \frac{1}{2}\vec{B}^{A,10} \times \vec{r}_{ig}^{\mathrm{BFF}}. \tag{51}$$

Given the natural quantum mechanical definitions of $D_v^{01(K)}$ (45) and D_u^{10} (47), the ad hoc introduction of Eq. (48) through a classical nuclear magnetic moment $\vec{\mu}_K^{C,01}$ and Eq. (50) through an artificial magnetic field $\vec{B}^{A,10}$ is not necessary, widely adopted in the literature though.

The Model Space V_{mod}

Apart from the perturbation operators, V^{01} (38), V^{01} (44), and V^{10} (46), the effective Hamiltonian H_{eff}^{11} (11) still requires a proper definition of the model space V_{mod} in which it is defined. Since $H^{00} = H_{00}^{\mathrm{BFF}}$, the zeroth-order rovibronic state $|i_{\mathrm{rve}}^{00}\rangle$ under concern can be obtained as

$$H^{00}|i_{\mathrm{rve}}^{00}\rangle = E_i^{00}|i_{\mathrm{rve}}^{00}\rangle. \tag{52}$$

It is further assumed that $|i_{\mathrm{rve}}^{00}\rangle$ is nondegenerate. The model space V_{mod} can then be built according to

$$V_{\mathrm{mod}} = \{|i_{\mathrm{rve}}^{00}\Phi_p\rangle\}, \quad p = 1, \cdots, n_d, \tag{53}$$

where the spin states $\{\Phi_p\}$ form a complete orthonormal basis for the given nuclear spin. Under the adiabatic approximation, the rovibronic state $|i_{\mathrm{rve}}^{00}\rangle$ is expressed as the product of an electronic state $|i_{\mathrm{ele}}^{00}\rangle$, a vibrational state $|i_{\mathrm{vib}}^{00}\rangle$, and a rotational state $|i_{\mathrm{rot}}^{00}\rangle$, viz.,

$$|i_{\mathrm{rve}}^{00}\rangle \approx |i_{\mathrm{rot}}^{00}\rangle \otimes |i_{\mathrm{vib}}^{00}\rangle \otimes |i_{\mathrm{ele}}^{00}\rangle. \tag{54}$$

The electronic state $|i_{\mathrm{ele}}^{00}\rangle$ with energy $E_{\mathrm{ele}}^{00}[i_{\mathrm{ele}}^{00}](Q_k)$ can first be obtained by solving the electronic Dirac-Coulomb-Breit equation with the clamped nuclei. The rotational state $|i_{\mathrm{rot}}^{00}\rangle$ with quantum number J can then be calculated under the rigid rotor approximation, while the vibrational state $|i_{\mathrm{vib}}^{00}\rangle$ can finally be obtained by solving the nuclear Schrödinger-like equation

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial(Q_k)^2} + T_{\mathrm{cor}} + E_{\mathrm{ele}}^{00}[i_{\mathrm{ele}}^{00}](Q_k) + V_J + V_{\mathrm{ad}}(Q_k)\right]|i_{\mathrm{vib}}^{00}\rangle = E_{ev}^{00}|i_{\mathrm{vib}}^{00}\rangle. \tag{55}$$

Here, T_{cor} is a correction due to the vibrational angular momentum $\vec{j}^{\mathrm{vib}}$. For nonlinear molecules $T_{\mathrm{cor}} = \frac{1}{2} j_u^{\mathrm{vib}} (I'^{-1})_{uv} j_v^{\mathrm{vib}}$ [see Eq. (63) in Ref. [28]], while

for linear molecules $T_{\rm cor} = \frac{1}{2}\mathcal{I}\simeq^{-1}[(j_x^{\rm vib})^2 + (j_y^{\rm vib})^2]$ [see Eq. (72) in Ref. [29]]. The term V_J arises from the centrifugal effect of the molecular rotation (the Q_k-dependent part of the rotational energy). For nonlinear molecules $V_J = \frac{1}{2}J_u[I'^{-1} - (I^0)^{-1}]_{uv}J_v$ [see Eq. (63) in Ref. [28]], while for linear molecules $V_J = \frac{1}{2}[\mathcal{I}\simeq^{-1} - (\mathcal{I}^0)^{-1}](J_x^2 + J_y^2)$ [see Eq. (72) in Ref. [29]]. The last term $V_{\rm ad}(Q_k)$ collects all the diagonal adiabatic corrections (including the mass polarization, Watson potential, inverse-inertia-weighted spin-orbit, spin-spin, and orbit-orbit interactions, etc.). Equation (55) is usually approximated as

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial(Q_k)^2} + E_{\rm ele}^{00}[i_{\rm ele}^{00}](Q_k) + V_J\right]|i_{\rm vib}^{00}\rangle = E_{ev}^{00}|i_{\rm vib}^{00}\rangle, \tag{56}$$

or

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial(Q_k)^2} + E_{\rm ele}^{00}[i_{\rm ele}^{00}](Q_k)\right]|i_{\rm vib}^{00}\rangle = E_{ev}^{00}|i_{\rm vib}^{00}\rangle \tag{57}$$

when the centrifugal effect V_J is further neglected.

Construction of the Effective Nuclear Spin Hamiltonian

With the available zeroth-order state $|i_{\rm rve}^{00}\rangle$ (54), as well as the perturbation operators V^{11} (38), V^{01} (44), and V^{10} (46), the mixed second-order nuclear spin Hamiltonian $H_{\rm eff}^{11}$ (11) can now be constructed. The diamagnetic term $H_{\rm eff}^{11,d}$ (12) reads

$$H_{\rm eff}^{11,d} = \sum_{p,q=1}^{n_d} |i_{\rm rve}^{00}\Phi_p\rangle\langle i_{\rm rve}^{00}\Phi_p|V^{11}|i_{\rm rve}^{00}\Phi_q\rangle\langle i_{\rm rve}^{00}\Phi_q| \tag{58}$$

$$= -\sum_{p,q=1}^{n_d} |i_{\rm rve}^{00}\Phi_p\rangle\langle i_{\rm rve}^{00}\Phi_p|\sum_{K=1} B_{{\rm ext},u}^{\rm LAB}\mu_{K,u}|i_{\rm rve}^{00}\Phi_q\rangle\langle i_{\rm rve}^{00}\Phi_q| \tag{59}$$

$$= -|i_{\rm rve}^{00}\rangle\langle i_{\rm rve}^{00}|\sum_{K=1} B_{{\rm ext},u}^{\rm LAB}\mu_{K,u}|i_{\rm rve}^{00}\rangle\langle i_{\rm rve}^{00}| \tag{60}$$

$$= -|i_{\rm rve}^{00}\rangle\sum_{K=1} B_{{\rm ext},u}^{\rm LAB}\mu_{K,u}\langle i_{\rm rve}^{00}|. \tag{61}$$

If we understand that $H_{\rm eff}^{11,d}$ is only defined in the model space, we can simply write it as

$$H_{\rm eff}^{11,d} = -\sum_{K=1} B_{{\rm ext},u}^{\rm LAB}\mu_{K,u}. \tag{62}$$

The paramagnetic term $H_{\rm eff}^{11,p}$ (13) can be calculated similarly, leading to

$$H_{\text{eff}}^{11,p} = \frac{\langle i_{\text{rve}}^{00}|R_{u'u}D_{u'}^{10}|a_{\text{rve}}^{00}\rangle\langle a_{\text{rve}}^{00}|\sum_K D_{v'}^{01(K)}R_{v'v}|i_{\text{rve}}^{00}\rangle}{E_i^{00}-E_a^{00}} B_{\text{ext},u}^{\text{LAB}}\mu_{K,v} + h.c.. \tag{63}$$

Comparing the sum of Eqs. (62) and (63) with Eq. (3) leads to

$$\sigma_{K,uv}^{\text{LAB}} = \frac{\langle i_{\text{rve}}^{00}|R_{u'u}D_{u'}^{10}|a_{\text{rve}}^{00}\rangle\langle a_{\text{rve}}^{00}|D_{v'}^{01(K)}R_{v'v}|i_{\text{rve}}^{00}\rangle}{E_i^{00}-E_a^{00}} + c.c. \tag{64}$$

Here, the intermediate rovibronic states $|a_{\text{rve}}^{00}\rangle$ are different from the state $|i_{\text{rve}}^{00}\rangle$ under concern. However, they can have the same or different electronic states. In the former situation, the element $\langle a_{\text{rve}}^{00}|D_v^{01(K)}|i_{\text{rve}}^{00}\rangle$ is zero, because $\langle i_{\text{ele}}^{00}|D_v^{01(K)}|i_{\text{ele}}^{00}\rangle$ is zero (which is obvious for $|i_{\text{ele}}^{00}\rangle$ is nondegenerate, while $D_v^{01(K)}$ is time odd). Therefore, $|i_{\text{rve}}^{00}\rangle$ and $|a_{\text{rve}}^{00}\rangle$ must have different electronic states. The denominator of Eq. (64) can hence be approximated as

$$E_i^{00}[i_{\text{rve}}^{00}] - E_a^{00}[a_{\text{rve}}^{00}] \approx E_{\text{ele}}^{00}[i_{\text{ele}}^{00}](Q_k^0) - E_{\text{ele}}^{00}[a_{\text{ele}}^{00}](Q_k^0) \tag{65}$$

$$\approx E_{\text{ele}}^{00}[i_{\text{ele}}^{00}](Q_k) - E_{\text{ele}}^{00}[a_{\text{ele}}^{00}](Q_k), \tag{66}$$

which amounts to neglecting the vibrational and rotational corrections to the electronic energy difference. Using Eq. (66), the following closure relations for the normal vibrational coordinates Q_k and Euler angles $s = \alpha, \beta, \gamma$,

$$\sum_{p_{\text{vib}}^{00}} |p_{\text{vib}}^{00}(Q_k)\rangle\langle p_{\text{vib}}^{00}(Q_k')| = \delta(Q_k - Q_k'), \tag{67}$$

$$\sum_{p_{\text{rot}}^{00}} |p_{\text{rot}}^{00}(s)\rangle\langle p_{\text{rot}}^{00}(s')| = \delta(s - s'), \tag{68}$$

as well as the fact that neither D_u^{10} nor $D_v^{01(K)}$ is dependent on the Euler angles, Eq. (64) can be brought into the following form:

$$\sigma_{K,uv}^{\text{LAB}} = \sigma_{u'v'}^{K} \int R_{u'u}R_{v'v}i_{\text{rot}}^{00\dagger}i_{\text{rot}}^{00}\sin\beta d\alpha d\beta d\gamma, \tag{69}$$

where the nuclear shielding σ_{uv}^K in BFF is defined as

$$\sigma_{uv}^K = \int dQ_k \sigma_{uv}^K(Q_k) i_{\text{vib}}^{00\dagger} i_{\text{vib}}^{00}, \tag{70}$$

$$\sigma_{uv}^K(Q_k) = \frac{\langle i_{\text{ele}}^{00}|D_u^{10}|a_{\text{ele}}^{00}\rangle\langle a_{\text{ele}}^{00}|D_v^{01(K)}|i_{\text{ele}}^{00}\rangle}{E_{\text{ele}}^{00}[i_{\text{ele}}^{00}](Q_k) - E_{\text{ele}}^{00}[a_{\text{ele}}^{00}](Q_k)} + c.c.. \tag{71}$$

The nuclear shielding at the equilibrium structure, $\sigma_{uv}^{K}(Q_k^0)$, is usually a good approximation to σ_{uv}^{K}. To get more accurate result, the vibrational correction Eq. (70) should be performed, followed by the rotational average (69).

Now it is worthy to discuss what is available from an experimental measurement. First of all, the off-diagonal components of $\sigma_{K,uv}^{\mathrm{LAB}}$ cannot be measured, because they can hardly affect the measurable strength of the effective magnetic field $\vec{\mathcal{B}}_{\mathrm{eff}}^{\mathrm{LAB}(K)}$. This can be seen as follows. Supposing $\vec{B}_{\mathrm{ext}}^{\mathrm{LAB}}$ is applied to the z direction (B_z), we have

$$\vec{\mathcal{B}}_{\mathrm{eff}}^{\mathrm{LAB}(K)} = (-\sigma_{K,zx}^{\mathrm{LAB}} B_z, -\sigma_{K,zy}^{\mathrm{LAB}} B_z, (1-\sigma_{K,zz}^{\mathrm{LAB}}) B_z), \tag{72}$$

$$|\vec{\mathcal{B}}_{\mathrm{eff}}^{\mathrm{LAB}(K)}| = B_z \sqrt{(\sigma_{K,zx}^{\mathrm{LAB}})^2 + (\sigma_{K,zy}^{\mathrm{LAB}})^2 + (1-\sigma_{K,zz}^{\mathrm{LAB}})^2} \tag{73}$$

$$\approx (1-\sigma_{K,zz}^{\mathrm{LAB}}) B_z. \tag{74}$$

Equation (74) arises from the fact that all the components of $\sigma_{K,uv}^{\mathrm{LAB}}$ are only on the order of ppm, several orders of magnitude smaller than one. Second, even the diagonal components $\sigma_{K,uu}^{\mathrm{LAB}}$ can hardly be resolved individually. Instead, only the isotropic shielding $\sigma_{K,\mathrm{iso}}^{\mathrm{LAB}}$ can be measured, which is just the average of the diagonal components $\sigma_{K,uu}^{\mathrm{LAB}}$:

$$\sigma_{K,\mathrm{iso}}^{\mathrm{LAB}} = \frac{1}{3}\sigma_{u'v'}^{K} \int d\alpha d\beta d\gamma \sin\beta i_{\mathrm{rot}}^{00\dagger} i_{\mathrm{rot}}^{00} R_{u'u} R_{v'u} \tag{75}$$

$$= \frac{1}{3}\sigma_{u'u'}^{K} \int d\alpha d\beta d\gamma \sin\beta i_{\mathrm{rot}}^{00\dagger} i_{\mathrm{rot}}^{00} \tag{76}$$

$$= \frac{1}{3}\sigma_{u'u'}^{K} = \sigma_{\mathrm{iso}}^{K}. \tag{77}$$

The effective Hamiltonian $H_{\mathrm{eff}}^{\mathrm{NS}}$ (3) can then be replaced by

$$H_{\mathrm{eff}}^{\mathrm{NS}} = -\sum_{K} B_{\mathrm{ext},v}^{\mathrm{LAB}} (1-\sigma_{K,\mathrm{iso}}^{\mathrm{LAB}}) \mu_{K,v}. \tag{78}$$

In sum, the calculation of NMR shielding involves three steps. The first step is to calculate $\sigma_{uv}^{K}(Q_k)$ (71) for a given geometry. Here, only the isotropic component is interested. In view of Eqs. (48) and (50), $\sigma_{uv}^{K}(Q_k)$ (71) is nothing but the mixed second-order derivative of the electronic energy

$$\sigma_{uv}^{K}(Q_k) = \frac{d^2 E_{\mathrm{ele}}[i_{\mathrm{ele}}^{00}](Q_k)}{dB_u^{A,10} d\mu_{K,v}^{C,01}} \Big|_{B_u^{A,10}=\mu_{K,v}^{C,01}=0} \tag{79}$$

in the presence of an artificial magnetic field $\vec{B}^{A,10}$ and a classical nuclear magnetic moment $\vec{\mu}_K^{C,01}$. The second step is to account for the vibrational (cf. Eq. (70)) and rotational (cf. Eq. (69)) corrections for a given rovibronic state. The third step is to make the thermal average according to the statistical distributions of rovibronic states. The first step is the most crucial step and will be discussed below in details. To simplify the notation, the superscripts "BFF," "A," and "C" in Eq. (79) will be omitted.

The Relativistic Diamagnetism

The fully relativistic theory (71) of the nuclear shielding $\sigma_{uv}^K(Q_k)$ has a profound feature: It is purely paramagnetic according to the usual convention. That is, the diamagnetic term $\langle i_{\text{ele}}^{00}|D_{uv}^{11(K)}|i_{\text{ele}}^{00}\rangle$ known from the nonrelativistic theory [1] is missing. This is because the Dirac operator is only linear instead of quadratic with respect to the mechanical momentum. If the paramagnetism is interpreted as the polarization of the electron cloud under the influence of an external magnetic field, the diamagnetism would correspond to a rigid motion of the electron cloud. One then has a funny situation: The electrons of a spherical atom are only polarized, but not moved, by the external magnetic field in the relativistic world, whereas it is the opposite that happens in the nonrelativistic world. In a simplified note, one can say that "Dirac's atoms are purely paramagnetic while Schrödinger's atoms are purely diamagnetic." Apart from such conceptual discontinuity, the absence of the relativistic diamagnetism also has additional consequences:

- The negative energy states (NES) as intermediate states in the shielding expression (71) have nonzero contributions even in the nonrelativistic limit (nrl). In other words, the no-pair approximation fails completely in this case.
- The basis set convergence is extremely slow: Exceedingly large basis sets with functions of very high angular momenta are needed to get a reasonable estimate of the shielding.
- The way to obtain experimental shielding has been based on the separation of diamagnetic and paramagnetic terms, which cannot be generalized to the relativistic realm if the relativistic diamagnetism cannot be identified.

To show how to recapture the relativistic diamagnetism, suffice it to consider a single Dirac electron moving in an electric field represented by the local scalar potential V and a magnetic field represented by the vector potential $\vec{A} = \vec{A}^{01} + \vec{A}^{10}$ defined in Eqs. (49) and (51), respectively. The very first try [2] for this was to attribute the paramagnetic contribution of NES to the diamagnetism:

$$\sigma_{uv}^{K(-)}(Q_k) = \sum_{E_a^{00}<-c^2} \frac{\langle i^{00}|D_u^{10}|a^{00}\rangle\langle a^{00}|D_v^{01(K)}|i^{00}\rangle}{E_i^{00} - E_a^{00}} + c.c. \tag{80}$$

$$\approx \sum_{E_a^{00}<-c^2} \frac{1}{2c^2}\langle i^{00}|D_u^{10}|a^{00}\rangle\langle a^{00}|D_v^{01(K)}|i^{00}\rangle + c.c. \tag{81}$$

$$\approx \sum_{E_a^{00}<-c^2} \frac{1}{2c^2}\langle i^{00}|D_u^{10}|p^{00}\rangle\langle p^{00}|D_v^{01(K)}|i^{00}\rangle + c.c. \tag{82}$$

$$= \sum_{E_a^{00}<-c^2} \frac{1}{2c^2}\langle i^{00}|D_u^{10}D_v^{01(K)} + h.c.|i^{00}\rangle. \tag{83}$$

The common denominator approximation has been used when going from Eqs. (80) to (81). Compared with this approximation, going from Eqs. (81) to (82) is more justified, because both D^{10} and $D^{01(K)}$ are off-diagonal operators, such that the artificially introduced positive energy states (PES) contribute only to $\mathcal{O}(c^{-2})$. The closure relation for a complete basis is finally invoked to arrive at Eq. (83). More detailed discussions along this line can be found from Ref. [30]. Such manipulation looks simple but is a very crude approximation.

The second try [31, 32] for getting the relativistic diamagnetism was based on the Gordon decomposition [33] of the electronic current. It is equivalent to using the following operator identities:

$$\tilde{D}^{10} = D^{10} + [D^{00}, \tau^{10}], \tag{84}$$

$$\tilde{D}^{01} = D^{01} + [D^{00}, \tau^{01}], \tag{85}$$

where

$$D^{00} = c\vec{\alpha}\cdot\vec{p} + (\beta - 1)c^2 + V, \tag{86}$$

$$\tau^{10} = -\frac{1}{2c}\beta\vec{\alpha}\cdot\vec{A}^{10}, \tag{87}$$

$$\tau^{01} = -\frac{1}{2c}\beta\vec{\alpha}\cdot\vec{A}^{01}. \tag{88}$$

Under the Coulomb gauge, the *diagonal* operators $\tilde{D}^{10}$ and $\tilde{D}^{01}$ read explicitly

$$\tilde{D}^{10} = \beta\vec{A}^{10}\cdot\vec{p} + \frac{1}{2}\beta\vec{\Sigma}\cdot\vec{B}^{10}, \tag{89}$$

$$\tilde{D}^{01} = \beta\vec{A}^{01}\cdot\vec{p} + \frac{1}{2}\beta\vec{\Sigma}\cdot\vec{B}^{01}. \tag{90}$$

Equations (84)/(85) shows that the off-diagonal operator D^{10}/D^{01} entering the shielding $\sigma_{uv}^{K}(Q_k)$ (71) can be reexpressed as the summation of a diagonal operator $\tilde{D}^{10}/\tilde{D}^{01}$ and a commutator. The commutator will then appear as diamagnetism whereas $\tilde{D}^{10}/\tilde{D}^{01}$ in place of D^{10}/D^{01} as paramagnetism. This approach was reformulated more elegantly by Kutzelnigg [34] through a unitary transformation of the field-dependent Dirac operator D:

$$D = D^{00} + D^{10} + D^{01}, \tag{91}$$

$$\tilde{D} = e^{-\tau} D e^{\tau}, \quad \tau = \tau^{10} + \tau^{01}, \tag{92}$$

$$= D^{00} + \tilde{D}^{10} + \tilde{D}^{01} + \tilde{D}^{11} + \cdots, \tag{93}$$

where $\tilde{D}^{10}$ and $\tilde{D}^{01}$ are already defined in Eqs. (84) and (85), respectively. The diamagnetic term $\tilde{D}^{11}$ has a rather complicated form and is hence not documented here.

At variance with the approximate nature of the Sternheim approach [2, 30], the Gordon decomposition [31, 32] or equivalently the Kutzelnigg unitary transformation [34] approach is exact and captures the relativistic diamagnetism in a natural manner. However, this latter approach suffers from severe numerical divergences, which were only noticed in an alternative formulation, i.e., full field-dependent unitary transformation (FFUT) [35]. It turns out that the concept of "orbital decomposition" [35, 36] provides a more generic route for the exact and divergence-free formulations. It goes as follows.

From the one-electron Dirac equation,

$$\begin{pmatrix} V & c\vec{\sigma}\cdot\vec{\pi} \\ c\vec{\sigma}\cdot\vec{\pi} & V - 2c^2 \end{pmatrix} \begin{pmatrix} |p^L\rangle \\ |p^S\rangle \end{pmatrix} = \epsilon_p \begin{pmatrix} |p^L\rangle \\ |p^S\rangle \end{pmatrix}, \quad \vec{\pi} = \vec{p} + \vec{A}^{10} + \vec{A}^{01}, \tag{94}$$

we readily obtain the following relation:

$$|p^S\rangle = Y_p |p^L\rangle, \tag{95}$$

$$Y_p = \frac{1}{2c} R_p(\vec{r}) \vec{\sigma}\cdot\vec{\pi}, \quad R_p(\vec{r}) = [1 + \frac{\epsilon_p - V(\vec{r})}{2c^2}]^{-1} \approx 1 \tag{96}$$

between the small and large components of the Dirac spinor $|p\rangle$ with energy ϵ_p. Regarding $\vec{A}^{10}$ as the primary perturbation, the first-order term of the spinor $|p\rangle$ reads

$$|p^{10}\rangle = \begin{vmatrix} p^{L,10} \\ p^{S,10} \end{vmatrix} = \begin{vmatrix} p^{L,10} \\ Y_p^{00} p^{L,10} \end{vmatrix} + \begin{vmatrix} 0 \\ Y_p^{10} p^{L,00} \end{vmatrix}, \tag{97}$$

where

$$Y_p^{00} = \frac{1}{2c}\vec{\sigma}\cdot\vec{p} + \mathcal{O}(c^{-3}), \tag{98}$$

$$Y_p^{10} = \frac{1}{2c}\vec{\sigma}\cdot\vec{A}^{10} + \mathcal{O}(c^{-3}), \tag{99}$$

according to Eqs. (95) and (96). Had only the first leading term of Y_p^{10} (99) been used, the second term in Eq. (97) would be known before solving the response equation for $|p^{10}\rangle$. Inspired by this, Eq. (97) can be rewritten in a more general form as

$$|p^{10}\rangle = |p_{\mathrm{m}}^{10}\rangle + |p_{\mathrm{r}}^{10}\rangle, \quad |p_{\mathrm{m}}^{10}\rangle = Z^{10}|p^{00}\rangle, \tag{100}$$

with the following Z^{10} and $|p_{\mathrm{m}}^{10}\rangle$:

$$Z_{\mathrm{ODA}}^{10} = \begin{pmatrix} 0 & 0 \\ \frac{\vec{\sigma}\cdot\vec{A}^{10}}{2c} & 0 \end{pmatrix}, \quad |p_{\mathrm{m,ODA}}^{10}\rangle = \left| \begin{matrix} 0 \\ \frac{\vec{\sigma}\cdot\vec{A}^{10}}{2c} \end{matrix} p^{L,00} \right\rangle. \tag{101}$$

In other words, the first-order orbital $|p^{10}\rangle$ is here decomposed into a known field-dependent term $|p_{\mathrm{m}}^{10}\rangle$ and a residual $|p_{\mathrm{r}}^{10}\rangle$, which contribute to the respective diamagnetic (i.e., an expectation value over the zeroth-order state) and paramagnetic (i.e., response) components of the mixed second-order energy E^{11}, viz.:

$$E^{11} = \langle i^{00}|D^{01}|i^{10}\rangle + c.c. \tag{102}$$

$$= \langle i^{00}|D^{01}Z^{10} + h.c.|i^{00}\rangle + (\langle i^{00}|D^{01}|i_{\mathrm{r}}^{10}\rangle + c.c.). \tag{103}$$

The residual $|p_{\mathrm{r}}^{10}\rangle$ is to be projected onto the space of zeroth-order orbitals, i.e.,

$$|p_{\mathrm{r}}^{10}\rangle = \sum_{q,\epsilon_q^{00}>0} C_{qp}^{10}|q^{00}\rangle + \sum_{q,\epsilon_q^{00}<0} C_{qp}^{10}|q^{00}\rangle. \tag{104}$$

It can readily be shown [36] that the above second term due to the NES contributes to the paramagnetic shielding at $\mathcal{O}(c^{-2})$ and therefore vanishes in the nrl. This term can safely be neglected if only chemical shift instead of accurate total shielding is to be targeted [35].

It should be emphasized here that, given the significant difference between Z^{10} (101) and the exact Y_p^{10} (99), the *Ansatz* (100) for $|p^{10}\rangle$ is *exact*: Any difference in between is to be covered by the expansion (104) for $|p_{\mathrm{r}}^{10}\rangle$. Even the choice of $Z^{10} = 0$ is exact. It is just that the diamagnetism does not show up explicitly in this case. Given the nonuniqueness of Z^{10}, several other choices are possible [6]. For instance, the anti-Hermitian part of Z_{ODA}^{10} reads

$$Z^{10}_{\mathrm{EFUT}} = Z^{10}_{\mathrm{ODA}} - (Z^{10}_{\mathrm{ODA}})^{\dagger} = \begin{pmatrix} 0 & -\frac{\vec{\sigma}\cdot\vec{A}^{10}}{2c} \\ \frac{\vec{\sigma}\cdot\vec{A}^{10}}{2c} & 0 \end{pmatrix} = \tau^{10}, \tag{105}$$

which leads to

$$|p^{10}_{\mathrm{m,EFUT}}\rangle = \left| \begin{matrix} -\frac{\vec{\sigma}\cdot\vec{A}^{10}}{2c} p^{S,00} \\ \frac{\vec{\sigma}\cdot\vec{A}^{10}}{2c} p^{L,00} \end{matrix} \right\rangle. \tag{106}$$

This approach has been dubbed as EFUT (external field-dependent unitary transformation) [35]. Although formulated differently, EFUT is equivalent to the Kutzelnigg unitary transformation [34] but with τ (92) replaced by τ^{10} (87) and also equivalent to the Gordon decomposition approach if only the "10" but not the "01" part of the current density is decomposed. It is precisely the r^{-2} behavior of $\vec{A}^{01}$ (49) that results in numerical divergences through the commutator $[c\vec{\alpha}\cdot\vec{p}, \tau^{01}]$ [35]. The so-called restricted magnetic balance (RMB) approach [37] can also be regarded as a particular decomposition of the first-order orbitals [38]; see the RMB Z^{10}_{m} operator in Table 2. It deserves to be mentioned that all the approaches, with different Z^{10} though, are completely equivalent. They are just different decompositions of a single value (total shielding) into the sum of two values (diamagnetic and paramagnetic terms). In practice, they perform rather similarly [38].

For completeness, we now give the explicit expressions of the nuclear shielding by the EFUT approach. The paramagnetic term is very similar to Eq. (71),

$$\sigma^{K,\mathrm{p}}_{uv}(Q_k) = \frac{\langle i^{00}_{\mathrm{ele}}|D^{10}_u(\mathrm{EFUT})|a^{00}_{\mathrm{ele}}\rangle\langle a^{00}_{\mathrm{ele}}|D^{01(K)}_v(\mathrm{EFUT})|i^{00}_{\mathrm{ele}}\rangle}{E^{00}_{\mathrm{ele}}(i^{00}_{\mathrm{ele}}, Q_k) - E^{00}_{\mathrm{ele}}(a^{00}_{\mathrm{ele}}, Q_k)} + c.c., \tag{107}$$

where

$$D^{10}_u(\mathrm{EFUT}) = \tilde{D}^{10}_u, \quad D^{01(K)}_v(\mathrm{EFUT}) = D^{01(K)}_v, \tag{108}$$

Table 2 The Z^{10}_{m} and Z^{10}_{g} operators[a]. *LRT* linear response theory with $Z^{10}_{\mathrm{m}} = 0$, *ODA* orbital decomposition approach, *EFUT* external field-dependent unitary transformation, *RMB* restricted magnetic balance, *GIAO* gauge-including atomic orbitals[b] (Reprinted with permission from [26] ©2012 American Institute of Physics)

Method	$(Z^{10}_{\mathrm{m}})_{11}$	$(Z^{10}_{\mathrm{m}})_{12}$	$(Z^{10}_{\mathrm{m}})_{21}$	$(Z^{10}_{\mathrm{m}})_{22}$	$(Z^{10}_{\mathrm{g}})_{11}$	$(Z^{10}_{\mathrm{g}})_{12}$	$(Z^{10}_{\mathrm{g}})_{21}$	$(Z^{10}_{\mathrm{g}})_{22}$
LRT-GIAO	0	0	0	0	$-i\lambda^{\mu}_g$	0	0	$-i\lambda^{\mu}_g K_p$
ODA-GIAO	0	0	K_A	0	$-i\lambda^{\mu}_g$	0	0	$-i\lambda^{\mu}_g K_p$
EFUT-GIAO	0	$-K_A K_p$	K_A	0	$-i\lambda^{\mu}_g$	0	0	$-i\lambda^{\mu}_g K_p$
RMB-GIAO	0	0	0	K_A	$-i\lambda^{\mu}_g$	0	0	$-i\lambda^{\mu}_g K_p$

[a] $K_A = \frac{\vec{\sigma}\cdot\vec{A}^{10}_{\mu}}{2c}$; $K_p = \frac{\vec{\sigma}\cdot\vec{p}}{2c}$

[b] Strictly speaking, $Z^{10}_{\mathrm{g}} = [exp(-i\lambda^{\mu}_g)]^{10} = -i\lambda^{\mu}_g$ (112). However, to unify the notation, we have redefined Z^{10}_{g} as $-i\lambda^{\mu}_g K_p$

while the diamagnetic term reads

$$\sigma_{uv}^{K,\mathrm{d}}(Q_k) = \langle i_{\mathrm{ele}}^{00} | D_{uv}^{11(K)}(\mathrm{EFUT}) | i_{\mathrm{ele}}^{00} \rangle, \tag{109}$$

where

$$D_{uv}^{11(K)}(\mathrm{EFUT}) = \frac{d^2\beta \vec{A}^{10} \cdot \vec{A}^{01,(K)}}{dB_u^{10} d\mu_{K,v}^{01}}. \tag{110}$$

The Gauge Origin Problem

It is well known that the gauge origin of the external magnetic field is not an observable and should hence not affect the calculated nuclear shielding. However, this is not the case when a finite basis set is employed. To see this, consider two gauge origins: $\vec{R}_g$ and $\vec{R}_{g'}$. The vector potentials of the same external magnetic field $\vec{B}^{10}$ are related by

$$\vec{A}_{g'}^{10} = \vec{A}_g^{10} + \vec{\nabla}\lambda_{g'}^g, \tag{111}$$

where

$$\lambda_{g'}^g = \frac{1}{2}[\vec{B}^{10} \times (\vec{R}_g - \vec{R}_{g'})] \cdot \vec{r}. \tag{112}$$

This is equivalent to making a unitary transformation of the Hamiltonian and the wave function:

$$D' = e^{-i\lambda_{g'}^g} D e^{i\lambda_{g'}^g}, \tag{113}$$

$$|p(\vec{R}_{g'})\rangle = e^{-i\lambda_{g'}^g} |p(\vec{R}_g)\rangle. \tag{114}$$

It is interesting to see that such equivalence holds for both the nonrelativistic and relativistic cases. To first order, the wave functions with gauge origins at $\vec{R}_{g'}$ and $\vec{R}_g$ are related by

$$|p^{10}(\vec{R}_{g'})\rangle = -i\lambda_{g'}^g |p^{00}\rangle + |p^{10}(\vec{R}_g)\rangle. \tag{115}$$

The appearance of the electronic coordinate $\vec{r}$ in $\lambda_{g'}^g$ (112) implies that a fully polarized field-free basis set is required to account for the difference between $|p^{10}(\vec{R}_{g'})\rangle$ and $|p^{10}(\vec{R}_g)\rangle$. For an atom, the position $\vec{R}_K$ of the nucleus is a natural choice for the gauge origin, as the unperturbed atomic wave function is then correct up to first order (i.e., $|p^{10}(\vec{R}_K)\rangle = 0$) in the absence of spin-orbit interactions. For an arbitrary gauge origin $\vec{R}_g$, a first-order atomic orbital $|\chi_\mu^{10}(\vec{R}_g)\rangle$ of atom K can be decomposed into a gauge term and an atom-centered term

$$|\chi_\mu^{10}(\vec{R}_g)\rangle = -i\lambda_g^K|\chi_\mu^{00}(\vec{R}_K)\rangle + |\chi_\mu^{10}(\vec{R}_K)\rangle, \tag{116}$$

which amounts to replacing $\vec{R}_\mathrm{g}$ with $\vec{R}_K$ and $\vec{R}_{\mathrm{g}'}$ with $\vec{R}_\mathrm{g}$ in Eq. (115). The atom-centered term $|\chi_\mu^{10}(\vec{R}_K)\rangle$ can further be decomposed into a magnetic term and a residual according to Eq. (100). This idea can be generalized to molecular orbitals, leading to

$$|p^{10}\rangle = |p_\mathrm{g}^{10}\rangle + |p_\mathrm{m}^{10}\rangle + |p_\mathrm{r}^{10}\rangle, \tag{117}$$

$$|p_\mathrm{g}^{10}\rangle = -i\lambda_g^\mu|\chi_\mu^{00}(\vec{R}_\mu)\rangle C_{\mu p}^{00}, \tag{118}$$

$$= B_v^{10} Z_\mathrm{g}^{10,v}|p^{00}\rangle, \quad Z_\mathrm{g}^{10,v}|\chi_\mu^{00}\rangle = -\frac{i}{2}\left(\vec{R}_{\mu\mathrm{g}} \times \vec{r}\right)_v |\chi_\mu^{00}\rangle, \tag{119}$$

$$|p_\mathrm{m}^{10}\rangle = Z_\mathrm{m}^{10}|\chi_\mu^{00}(\vec{R}_\mu)\rangle C_{\mu p}^{00}, \tag{120}$$

$$= B_v^{10} Z_\mathrm{m}^{10,v}|p^{00}\rangle, \tag{121}$$

where $\vec{R}_\mu$ stands for the center of the atomic function $|\chi_\mu^{00}\rangle$ and $C_{\mu p}^{00}$ are the coefficients of zeroth-order molecular orbitals

$$|p^{00}\rangle = |\chi_\mu^{00}\rangle C_{\mu p}^{00}. \tag{122}$$

It is understood that both the gauge factor $Z_\mathrm{g}^{10,v}$ and the magnetic balance $Z_\mathrm{m}^{10,v}$ act directly on the field-free atomic function $|\chi_\mu^{00}\rangle$ instead of the molecular orbital $|p^{00}\rangle$ as a whole. Equation (119) is just the gauge-including atomic orbital (GIAO) approach [39, 40]. When Z_EFUT^{10} (105) is used for Z^{10} in Eq. (120), we have

$$|p_\mathrm{m}^{10}\rangle = B_v^{10} Z_\mathrm{m}^{10,v}|p^{00}\rangle, \quad Z_\mathrm{m}^{10,v}|\chi_\mu^{00}\rangle = -\frac{1}{4c}\left(\beta\vec{r}_\mu \times \vec{\alpha}\right)_v |\chi_\mu^{00}\rangle. \tag{123}$$

Equations (119) and (123) together have been dubbed as magnetically balanced GIAO (MB-GIAO) [41]. As confirmed numerically [41], not only the gauge dependence has been removed completely, but also the basis set convergence has been accelerated greatly.

The MB-GIAO-DKS Approach

In this section, we consider the MB-GIAO-DKS approach for nuclear shielding.

General Formulation

The (infinite-order) DKS energy in the presence of two fields characterized by $\vec{\lambda}^{01}$ and $\vec{\lambda}^{10}$ reads

$$E = \langle\psi_i|h|\psi_i\rangle + E_{\mathrm{coul}}[\rho] + E_{\mathrm{xc}}[\rho, s], \tag{124}$$

where

$$h = D^{00} + D_u^{01}\lambda_u^{01} + D_v^{10}\lambda_v^{10} + \lambda_u^{01}h_{uv}^{11}\lambda_v^{10}, \tag{125}$$

$$E_{\mathrm{coul}}[\rho] = \frac{1}{2}\iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} d\vec{r}_1 d\vec{r}_2. \tag{126}$$

A noncollinear form of the exchange-correlation functional $E_{\mathrm{xc}}[\rho, s]$ must be adopted to describe properly the spin-orbit couplings for the response even of closed-shell systems [42–44]. Under the adiabatic local density approximation (ALDA), three definitions of the four-component "spin density" s have been proposed [26]:

$$s = |\vec{\mathcal{M}}|, \quad \vec{\mathcal{M}}(\vec{r}') = \langle\psi_i|\hat{M}\delta(\vec{r} - \vec{r}')|\psi_i\rangle, \quad \hat{M} = \begin{pmatrix} \vec{\sigma} & 0 \\ 0 & \eta\vec{\sigma} \end{pmatrix}. \tag{127}$$

All the three variants are invariant with respect to rotations in the real and spin spaces. Given the formal difference, they perform rather similarly in practice [26].

The DKS spinors ψ_p can generally be expanded in a perturbation-dependent basis $\{\chi_\mu\}$:

$$\psi_p = \chi_\mu C_{\mu p}, \tag{128}$$

$$\chi_\mu = \chi_\mu^{00} + \chi_\mu^{01,u}\lambda_u^{01} + \chi_\mu^{10,v}\lambda_v^{10} + \cdots. \tag{129}$$

The stationarity condition for the coefficients reads

$$FC = SC\epsilon, \quad C^\dagger SC = I, \tag{130}$$

$$F_{\mu\nu} = \langle\chi_\mu|F|\chi_\nu\rangle, \quad S_{\mu\nu} = \langle\chi_\mu|\chi_\nu\rangle. \tag{131}$$

Taking λ_v^{10} as the primary perturbation, the condition can be expanded to first order as

$$F^{00}C^{00} = S^{00}C^{00}\epsilon^{00}, \quad C^{00\dagger}S^{00}C^{00} = I, \tag{132}$$

$$F^{10,v}C^{00} + F^{00}C^{10,v} = S^{10,v}C^{00}\epsilon^{00} + S^{00}C^{10,v}\epsilon^{00} + S^{00}C^{00}\epsilon^{10,v}, \tag{133}$$

$$C^{00\dagger}S^{10,v}C^{00} + (C^{10,v})^\dagger S^{00}C^{00} + C^{00\dagger}S^{00}C^{10,v} = 0. \tag{134}$$

The expressions can be simplified by going to the basis $\{\xi_q = \chi_\mu C_{\mu q}^{00}\}$, viz.,

$$\tilde{F}^{00} = \epsilon^{00}, \quad \tilde{S}^{00} = I, \quad U^{00} = I, \tag{135}$$

$$\tilde{F}^{10,v} + \epsilon^{00}U^{10,v} = \tilde{S}^{10,v}\epsilon^{00} + U^{10,v}\epsilon^{00} + \epsilon^{10,v}, \tag{136}$$

$$\tilde{S}^{10,v} + U^{10,v\dagger} + U^{10,v} = 0, \tag{137}$$

in view of the relations

$$\tilde{F} = C^{00\dagger} F C^{00}, \quad \tilde{S} = C^{00\dagger} S C^{00}, \quad U = (C^{00})^{-1} C. \tag{138}$$

Having determined the coefficients $U^{10,v}$ by Eq. (136), the orbitals ψ_i can be constructed as

$$\psi_i = \xi_q U_{qi} = \psi_i^{00} + \psi_i^{10,v} \lambda_v^{10} + \cdots, \tag{139}$$

$$\psi_i^{00} = \xi_i^{00} = \chi_\mu^{00} C_{\mu i}^{00}, \tag{140}$$

$$\psi_i^{10,v} = \psi_{i,b}^{10,v} + \psi_{i,r}^{10,v}, \tag{141}$$

$$\psi_{i,\mathrm{b}}^{10,v} = \xi_i^{10,v} = \chi_\mu^{10,v} C_{\mu i}^{00} = Z^{10,v} \psi_i^{00}, \tag{142}$$

$$\psi_{i,\mathrm{r}}^{10,v} = \psi_q^{00} U_{qi}^{10,v}. \tag{143}$$

It is seen that the first-order orbital $\psi_i^{10,v}$ (141) contains two terms: the already "known" term $\psi_{i,b}^{10,v}$ (142) and the residual response $\psi_{i,\mathrm{r}}^{10,v}$ (143) that can well be represented in the basis of zeroth-order orbitals. In terms of the so-obtained first-order orbitals, the mixed second-order energy can be calculated as

$$E_{uv}^{11} = E_{uv}^{11,\mathrm{d}} + E_{uv}^{11,\mathrm{p}} \tag{144}$$

$$E_{uv}^{11,\mathrm{d}} = \langle \psi_i^{00} | h_{uv}^{11} | \psi_i^{00} \rangle + \langle \psi_i^{00} | h_u^{01} Z^{10,v} + h.c. | \psi_i^{00} \rangle, \tag{145}$$

$$E_{uv}^{11,\mathrm{p}} = \langle \psi_i^{00} | h_u^{01} | \psi_p^{00} \rangle U_{pi}^{10,v} + c.c. \tag{146}$$

$$= -\langle \psi_i^{00} | h_u^{01} | \psi_j^{00} \rangle \tilde{S}_{ji}^{10,v} + \left[\langle \psi_i^{00} | h_u^{01} | \psi_a^{00} \rangle U_{ai}^{10,v} + c.c. \right], \tag{147}$$

$$\tilde{S}_{pq}^{10,v} = \langle \psi_p^{00} | Z^{10,v} + h.c. | \psi_q^{00} \rangle, \tag{148}$$

where the first and second terms of $E_{uv}^{11,\mathrm{p}}$ (147) arise from the occupied to occupied and occupied to virtual transitions, respectively. Moreover, the second term arises from the unitary normalization condition (137).

The expressions so far are valid for both electric and magnetic fields. For the latter, the first-order density vanishes pointwise [41], such that the coupled-perturbed DKS (CPDKS) equation (136) can further be rearranged [38] to

$$[(\epsilon_a^{00} - \epsilon_i^{00}) \delta_{ij} \delta_{ab} + 2 K_{ai,bj}] U_{bj}^{10,v} = -G_{ai}^{10,v}, \tag{149}$$

where

$$K_{ai,bj} = (\psi_a^{00} \beta \Sigma_v \psi_i^{00} | \frac{\partial^2 \varepsilon}{\partial s^2} [\rho^{00}, s^{00}] \delta(\vec{r} - \vec{r}') | \psi_j^{00} \beta \Sigma_v \psi_b^{00}), \tag{150}$$

$$G_{ai}^{10,v} = \langle \psi_a^{00} | \tilde{h}_v^{10} + \tilde{V}_{\mathrm{ind},v}^{10} | \psi_i^{00} \rangle - \tilde{S}_{ai}^{10,v} \epsilon_i^{00}, \tag{151}$$

$$\tilde{h}_v^{10} = h_v^{10} + Z_v^{10\dagger} F^{00} + F^{00} Z_v^{10}, \tag{152}$$

$$\langle \psi_a^{00} | \tilde{V}_{\mathrm{ind},v}^{10} | \psi_i^{00} \rangle = \langle \psi_a^{00} \beta \Sigma_u \psi_i^{00} | \frac{\partial^2 \varepsilon}{\partial s^2} [\rho^{00}, s^{00}] \delta(\vec{r} - \vec{r}') | \mathcal{M}_u^{10,v} \rangle, \tag{153}$$

$$\begin{aligned} \mathcal{M}_u^{10,v}(\vec{r}') = & \{ \langle \psi_i^{00} | \beta \Sigma_u \delta(\vec{r} - \vec{r}') | \psi_{i,\mathrm{b}}^{10,v} \rangle + c.c. \} \\ & - \langle \psi_i^{00} | \beta \Sigma_u \delta(\vec{r} - \vec{r}') | \psi_j^{00} \rangle \tilde{S}_{ji}^{10,v}. \end{aligned} \tag{154}$$

Application to Nuclear Shielding

For nuclear shielding, the $Z^{10,v}$ operator in Eq. (142) is composed of both $Z_{\mathrm{g}}^{10,u}$ (119) and $Z_{\mathrm{m}}^{10,u}$ (123). Due to the lack of h_{uv}^{11}, only the second term of $E_{uv}^{11,\mathrm{d}}$ (145) gives rise to the diamagnetic term of the shielding $\sigma_{vu}^K(Q_k)$,

$$\sigma_{vu}^{K,\mathrm{d}}(Q_k) = \langle \psi_i^{00} | D_u^{01(K)} Z_v^{10} + h.c. | \psi_i^{00} \rangle \tag{155}$$

$$= \sigma_{vu}^{K,\mathrm{da}} + \sigma_{vu}^{K,\mathrm{db}} + \sigma_{vu}^{K,\mathrm{dc}}, \tag{156}$$

$$\sigma_{vu}^{K,\mathrm{da}} = \langle \chi_\nu^{00} | \frac{\beta(\delta_{uv} \vec{r}_K \cdot \vec{r}_\mu - r_{K,v} r_{\mu,u})}{2c^2 r_N^3} | \chi_\mu^{00} \rangle C_{\nu i}^{00*} C_{\mu i}^{00}, \tag{157}$$

$$\sigma_{vu}^{K,\mathrm{db}} = \langle \chi_\nu^{00} | -\frac{i}{2} \left[(\vec{R}_{\mu\nu}) \times \vec{r} \right]_v D_u^{01(K)} | \chi_\mu^{00} \rangle C_{\nu i}^{00*} C_{\mu i}^{00}, \tag{158}$$

$$\sigma_{vu}^{K,\mathrm{dc}} = \langle \chi_\nu^{00} | \frac{1}{4c} \left(\beta \vec{R}_{\mu\nu} \times \vec{\alpha} \right)_v D_u^{01(K)} | \chi_\mu^{00} \rangle C_{\nu i}^{00*} C_{\mu i}^{00}, \tag{159}$$

where $\sigma_{vu}^{K,\mathrm{da}}$ stems from the magnetic balance incorporated by EFUT [35], whereas both $\sigma_{vu}^{K,\mathrm{db}}$ and $\sigma_{vu}^{K,\mathrm{dc}}$ arise from the GIAO [41].

In view of Eq. (147), the paramagnetic term of $\sigma_{vu}^K(Q_k)$ reads

$$\sigma_{vu}^{K,\mathrm{p}}(Q_k) = \sigma_{vu}^{K,\mathrm{p1}} + \sigma_{vu}^{K,\mathrm{p2}}, \tag{160}$$

$$\sigma_{vu}^{K,\mathrm{p1}} = -\langle \psi_i^{00} | D_u^{01(K)} | \psi_j^{00} \rangle \tilde{S}_{ji}^{10,v} \tag{161}$$

$$\sigma_{vu}^{K,\mathrm{p2}} = \langle \psi_i^{00} | D_u^{01(K)} | \psi_a^{00} \rangle U_{ai}^{10,v} + c.c., \tag{162}$$

where

$$\begin{aligned} \tilde{S}_{qp}^{10,v} = & \langle \chi_\nu^{00} | -\frac{i}{2} \left[(\vec{R}_{\mu\nu}) \times \vec{r} \right]_v | \chi_\mu^{00} \rangle C_{\nu q}^{00*} C_{\mu p}^{00} \\ & + \langle \chi_\nu^{00} | \frac{1}{4c} \left(\beta \vec{R}_{\mu\nu} \times \vec{\alpha} \right)_v | \chi_\mu^{00} \rangle C_{\nu q}^{00*} C_{\mu p}^{00}. \end{aligned} \tag{163}$$

The coefficients $U_{ai}^{10,v}$ are determined by the CPDKS equation (149), where the force $G_{ai}^{10,v}$ reads explicitly

$$G_{ai}^{10,v} = \langle \psi_a^{00} | \tilde{h}_v^{10} | \psi_i^{00} \rangle + \langle \psi_a^{00} | \tilde{V}_{\mathrm{ind},v}^{10} | \psi_i^{00} \rangle - \tilde{S}_{ai}^{10,v} \epsilon_i^{00}. \tag{164}$$

The first term of Eq. (164) reads

$$\begin{aligned} \langle \psi_q^{00} | \tilde{h}_v^{10} | \psi_p^{00} \rangle = \langle \chi_\nu^{00} | \frac{\beta (r_\mu \times \vec{p})_v + \beta \Sigma_v}{2} - \frac{i}{2} \left[(\vec{R}_{\mu\nu}) \times \vec{r} \right]_v F^{00} \\ + \frac{1}{4c} \left(\beta \vec{R}_{\mu\nu} \times \vec{\alpha} \right)_v F^{00} | \chi_\mu^{00} \rangle C_{\nu q}^{00*} C_{\mu p}^{00}, \end{aligned} \tag{165}$$

while the second term of Eq. (164) is defined by Eq. (153) yet with the following spin magnetization density $\mathcal{M}_u^{10,v}$ as input:

$$\mathcal{M}_u^{10,v} = \mathcal{M}_{u,\mathrm{b1}}^{10,v} + \mathcal{M}_{u,\mathrm{b2}}^{10,v} + \mathcal{M}_{u,\mathrm{b3}}^{10,v} + \mathcal{M}_{u,\mathrm{b4}}^{10,v}, \tag{166}$$

$$\mathcal{M}_{u,\mathrm{b1}}^{10,v} = -\frac{i}{2} \chi_\nu^{00\dagger} \left((\vec{R}_{\mu\nu} \times \vec{r})_v \beta \Sigma_u \right) \chi_\mu^{00} C_{\nu i}^{00*} C_{\mu i}^{00}, \tag{167}$$

$$\mathcal{M}_{u,\mathrm{b2}}^{10,v} = -\frac{1}{4c} \chi_\nu^{00\dagger} \left[i \delta_{uv} \vec{R}_{\mu\nu} \cdot \vec{\alpha} - i (R_{\mu\nu})_u \alpha_v \right] \chi_\mu^{00} C_{\nu i}^{00*} C_{\mu i}^{00}, \tag{168}$$

$$\mathcal{M}_{u,\mathrm{b3}}^{10,v} = -\frac{1}{4c} \chi_\nu^{00\dagger} \gamma_5 \epsilon_{wuv} \left[(r_\mu)_w + (r_\nu)_w \right] \chi_\mu^{00} C_{\nu i}^{00*} C_{\mu i}^{00}, \tag{169}$$

$$\mathcal{M}_{u,\mathrm{b4}}^{10,v} = -\psi_i^{00\dagger} \beta \Sigma_u \psi_j^{00} \tilde{S}_{ji}^{10,v}. \tag{170}$$

When solving the CPDKS equation (149), the NES can safely be treated in an uncoupled manner [41] (i.e., setting $K_{ai,bj} = 0$ for $U_{bj}^{10,v}$ with $\epsilon_b < -2c^2$). The computational cost of Eq. (149) is then not much different from the two-component counterpart. The contributions of the NES to the nuclear shielding can even be obtained from atomic calculations in the spirit of "from atoms to molecule" [45,46]. If only chemical shift is wanted, the NES can be neglected completely for they are essentially irrelevant for the description of first-order valence orbitals.

Two-Component Relativistic Theory of Nuclear Shielding

The X2C Equations

We start with a general four-component relativistic mean-field equation

$$F \psi_i = \epsilon_i \psi_i \tag{171}$$

or in block form

$$\begin{pmatrix} F^{LL} & F^{LS} \\ F^{SL} & F^{SS} \end{pmatrix} \begin{pmatrix} \psi_i^L \\ \psi_i^S \end{pmatrix} = \begin{pmatrix} \psi_i^L \\ \psi_i^S \end{pmatrix} \epsilon_i. \tag{172}$$

The spinor ψ_i is to be expanded as

$$\psi_i = Z\psi_i^M, \quad \psi_i^M = \begin{pmatrix} \psi^L \\ \phi_i^L \end{pmatrix} = \begin{pmatrix} g_\mu \mathbf{C}_{\mu i}^L \\ g_\mu \mathbf{C}_{\mu i}^S \end{pmatrix}. \tag{173}$$

The large and pseudo-large components of the modified spinor ψ_i^M are assumed to share the same symmetry properties and can therefore be expanded in the same basis $\{g_\mu\}$. The Z operator in Eq. (173) takes the following form [26]:

$$Z = Z_g(Z_k + Z_m), \tag{174}$$

the zeroth and first orders of which read

$$Z^{00} = Z_k = \begin{pmatrix} 1 & 0 \\ 0 & \frac{\vec{\sigma}\cdot\vec{p}}{2c} \end{pmatrix}, \tag{175}$$

$$Z^{10} = Z_g^{10} + Z_m^{10}. \tag{176}$$

For the various choices of the Z^{10} operator (176), see Table 2 (a more complete list is given in Table 1 of Ref. [26]).

With the replacement (173), Eq. (171) can be rewritten as

$$F^M \psi_i^M = M\psi_i^M \epsilon_i, \quad F^M = Z^\dagger F Z, \quad M = Z^\dagger Z. \tag{177}$$

To expedite subsequent manipulations, we further rewrite the modified spinor ψ_i^M as

$$\psi_i^M = \begin{pmatrix} f_p \mathbf{A}_{pi} \\ f_p \mathbf{B}_{pi} \end{pmatrix}, \tag{178}$$

where

$$f_p = g_\mu \mathbf{C}_{\mu p}^{L00}, \quad \mathbf{A} = (\mathbf{C}^{L00})^{-1}\mathbf{C}^L, \quad \mathbf{B} = (\mathbf{C}^{L00})^{-1}\mathbf{C}^S. \tag{179}$$

The matrix representation of Eq. (177) in the f_p-basis is then

$$\mathbf{HC} = \mathbf{MC}\epsilon, \tag{180}$$

where

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix}, \quad \mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{pmatrix}, \tag{181}$$

and

$$(\mathbf{H}_{11})_{pq} = \langle f_p | Z_{11}^\dagger F^{LL} Z_{11} + Z_{11}^\dagger F^{LS} Z_{21} + Z_{21}^\dagger F^{SL} Z_{11} + Z_{21}^\dagger F^{SS} Z_{21} | f_q \rangle, \tag{182}$$

$$\begin{aligned}(\mathbf{H}_{12})_{pq} &= \langle f_p | Z_{11}^\dagger F^{LL} Z_{12} + Z_{11}^\dagger F^{LS} Z_{22} + Z_{21}^\dagger F^{SL} Z_{12} + Z_{21}^\dagger F^{SS} Z_{22} | f_q \rangle \\ &= (\mathbf{H}_{21})_{qp}^*,\end{aligned} \tag{183}$$

$$(\mathbf{H}_{22})_{pq} = \langle f_p | Z_{12}^\dagger F^{LL} Z_{12} + Z_{12}^\dagger F^{LS} Z_{22} + Z_{22}^\dagger F^{SL} Z_{12} + Z_{22}^\dagger F^{SS} Z_{22} | f_q \rangle, \tag{184}$$

$$(\mathbf{M}_{11})_{pq} = \langle f_p | Z_{11}^\dagger Z_{11} + Z_{21}^\dagger Z_{21} | f_q \rangle, \tag{185}$$

$$(\mathbf{M}_{12})_{pq} = \langle f_p | Z_{11}^\dagger Z_{12} + Z_{21}^\dagger Z_{22} | f_q \rangle = (\mathbf{M}_{21})_{qp}^*, \tag{186}$$

$$(\mathbf{M}_{22})_{pq} = \langle f_p | Z_{12}^\dagger Z_{12} + Z_{22}^\dagger Z_{22} | f_q \rangle. \tag{187}$$

To formulate the X2C counterpart of Eq. (180), we first introduce a formal relation between the small and large coefficients for the PES

$$\mathbf{B} = \mathbf{X}\mathbf{A}, \tag{188}$$

in terms of which Eq. (180) can be written as two coupled sets of equations only in the large-component space

$$\mathbf{L}_+ \mathbf{A} = \mathbf{M}_+ \mathbf{A} \boldsymbol{\epsilon}, \tag{189a}$$

$$\bar{\mathbf{L}}_+ \mathbf{A} = \bar{\mathbf{M}}_+ \mathbf{A} \boldsymbol{\epsilon}, \tag{189b}$$

where

$$\mathbf{L}_+ = \mathbf{H}_{11} + \mathbf{H}_{12}\mathbf{X}, \tag{190}$$

$$\mathbf{M}_+ = \mathbf{M}_{11} + \mathbf{M}_{12}\mathbf{X}, \tag{191}$$

$$\bar{\mathbf{L}}_+ = \mathbf{H}_{21} + \mathbf{H}_{22}\mathbf{X}, \tag{192}$$

$$\bar{\mathbf{M}}_+ = \mathbf{M}_{21} + \mathbf{M}_{22}\mathbf{X}. \tag{193}$$

Eliminating the energy dependence in Eq. (189) leads to

$$\mathbf{M}_+^{-1} \mathbf{L}_+ = \bar{\mathbf{M}}_+^{-1} \bar{\mathbf{L}}_+, \tag{194}$$

which is just the decoupling condition for block-diagonalizing Eq. (180) in one step [47, 48]. Equation (189a) for determining the large-component coefficients $\mathbf{A}$ is known as the unnormalized elimination of the small component (UESC) [17] albeit in the absence of magnetic fields. The Hamiltonian $\mathbf{L}_+$ (190) is noticeably

non-Hermitian. The Hermitian counterpart is readily obtained by left-multiplying Eq. (189b) with $\mathbf{X}^\dagger$ and adding the result to Eq. (189a), viz.,

$$\tilde{\mathbf{L}}_+\mathbf{A} = \tilde{\mathbf{S}}_+\mathbf{A}\boldsymbol{\epsilon}, \tag{195}$$

where

$$\tilde{\mathbf{L}}_+ = \mathbf{L}_+ + \mathbf{X}^\dagger\bar{\mathbf{L}}_+, \tag{196}$$

$$\tilde{\mathbf{S}}_+ = \mathbf{M}_+ + \mathbf{X}^\dagger\bar{\mathbf{M}}_+. \tag{197}$$

Although known as the normalized elimination of the small component (NESC) [17], Eq. (195) is nothing but the stationarity condition for the variation of the energy expectation value with respect to the large-component coefficients [47]. Another variant, i.e., the symmetrized elimination of the small component (SESC) [45, 46], can also be formulated but is not considered here. Note in passing that, featured by the relativistic metric (197), Eq. (195) is still in Dirac picture. A proper renormalization has to be introduced to go to the Schrödinger picture [49,50]. Again, it is not considered here, as it is immaterial for the understanding of the X2C theory of nuclear shielding.

The X2C Approach for NMR Shielding

Both Eq. (195) for the coefficients $\mathbf{A}$ and Eq. (194) for the decoupling matrix $\mathbf{X}$ are infinite order in the magnetic fields. For nuclear shielding, only the first-order terms $\mathbf{A}^{10}$ and $\mathbf{X}^{10}$ are needed. Expanding Eq. (195) up to first order leads to

$$\tilde{\mathbf{L}}_+^{00} = \boldsymbol{\epsilon}^{00}, \quad \mathbf{A}^{00} = \mathbf{I}, \quad \tilde{\mathbf{S}}_+^{00} = \mathbf{I}, \tag{198}$$

$$\tilde{\mathbf{L}}_+^{10} + \boldsymbol{\epsilon}^{00}\mathbf{A}^{10} = \tilde{\mathbf{S}}_+^{10}\boldsymbol{\epsilon}^{00} + \mathbf{A}^{10}\boldsymbol{\epsilon}^{00} + \boldsymbol{\epsilon}^{10}, \tag{199}$$

where

$$\tilde{\mathbf{L}}_+^{10} = \mathbf{L}_+^{10} + \mathbf{X}^{00\dagger}\bar{\mathbf{L}}_+^{10} + \mathbf{X}^{10\dagger}\bar{\mathbf{L}}_+^{00}, \tag{200}$$

$$\tilde{\mathbf{S}}_+^{10} = \mathbf{M}_+^{10} + \mathbf{X}^{00\dagger}\bar{\mathbf{M}}_+^{10} + \mathbf{X}^{10\dagger}\bar{\mathbf{M}}_+^{00}. \tag{201}$$

Likewise, the first order of Eq. (194) reads

$$(\mathbf{M}_+^{-1})^{10}\mathbf{L}_+^{00} + (\mathbf{M}_+^{-1})^{00}\mathbf{L}_+^{10} = \left(\bar{\mathbf{M}}_+^{-1}\right)^{10}\bar{\mathbf{L}}_+^{00} + \left(\bar{\mathbf{M}}_+^{-1}\right)^{00}\bar{\mathbf{L}}_+^{10}, \tag{202}$$

which can be rewritten as

$$\begin{aligned} &-(\mathbf{M}_+^{00})^{-1}\mathbf{M}_+^{10}(\mathbf{M}_+^{00})^{-1}\mathbf{L}_+^{00} + (\mathbf{M}_+^{-1})^{00}\mathbf{L}_+^{10} \\ &= -(\bar{\mathbf{M}}_+^{00})^{-1}\bar{\mathbf{M}}_+^{10}(\bar{\mathbf{M}}_+^{00})^{-1}\bar{\mathbf{L}}_+^{00} + \left(\bar{\mathbf{M}}_+^{-1}\right)^{00}\bar{\mathbf{L}}_+^{10} \end{aligned} \tag{203}$$

by means of the following identity for a matrix $\mathbf{W}$:

$$(\mathbf{W}^{-1})^{10} = -(\mathbf{W}^{00})^{-1}\mathbf{W}^{10}(\mathbf{W}^{00})^{-1}. \tag{204}$$

In view of the zeroth-order equation

$$(\mathbf{M}_+^{00})^{-1}\mathbf{L}_+^{00} = (\bar{\mathbf{M}}_+^{00})^{-1}\bar{\mathbf{L}}_+^{00} = \boldsymbol{\epsilon}^{00}, \tag{205}$$

Eq. (203) can further be simplified to

$$(\mathbf{M}_+^{-1})^{00}\left(\mathbf{L}_+^{10} - \mathbf{M}_+^{10}\boldsymbol{\epsilon}^{00}\right) = \left(\bar{\mathbf{M}}_+^{-1}\right)^{00}\left(\bar{\mathbf{L}}_+^{10} - \bar{\mathbf{M}}_+^{10}\boldsymbol{\epsilon}^{00}\right). \tag{206}$$

Expanding Eqs. (190)–(193) to first order and using the fact that both $\mathbf{M}_{12}^{00}$ and $\mathbf{M}_{21}^{00}$ are zero for all the Z operators under consideration (see Table 2), Eq. (206) can be rearranged [26] to

$$\mathbf{P}^{00}\mathbf{X}^{10} - \mathbf{X}^{10}\epsilon^{00} = \mathbf{Q}^{10}[\mathbf{A}^{10}, \mathbf{X}^{10}], \tag{207}$$

where

$$\mathbf{P}^{00} = (\mathbf{M}_{22}^{00})^{-1}\mathbf{H}_{22}^{00} - \mathbf{X}^{00}(\mathbf{M}_{11}^{00})^{-1}\mathbf{H}_{12}^{00}, \tag{208}$$

$$\begin{aligned}\mathbf{Q}^{10}[\mathbf{A}^{10}, \mathbf{X}^{10}] = {} & \mathbf{X}^{00}(\mathbf{M}_{11}^{00})^{-1}\left(\mathbf{H}_{11}^{10} + \mathbf{H}_{12}^{10}\mathbf{X}^{00} - \mathbf{M}_{11}^{10}\epsilon^{00} - \mathbf{M}_{12}^{10}\mathbf{X}^{00}\epsilon^{00}\right) \\ & - (\mathbf{M}_{22}^{00})^{-1}\left(\mathbf{H}_{21}^{10} + \mathbf{H}_{22}^{10}\mathbf{X}^{00} - \mathbf{M}_{21}^{10}\epsilon^{00} - \mathbf{M}_{22}^{10}\mathbf{X}^{00}\epsilon^{00}\right).\end{aligned} \tag{209}$$

The dependence of $\mathbf{Q}^{10}$ (209) on $\mathbf{A}^{10}$ and $\mathbf{X}^{10}$ stems from the induced exchange-correlation potential. Close inspections reveal that $\mathbf{P}^{00}$ (208) is of $\mathcal{O}(c^2)$ and $\mathbf{Q}^{10}$ (209) is of $\mathcal{O}(c^2)$ for $Z^{10} = 0$, whereas $\mathcal{O}(c^0)$ for all the cases with $Z^{10} \neq 0$. Therefore, $\mathbf{X}^{10}$ is of $\mathcal{O}(c^0)$ and $\mathcal{O}(c^{-2})$ for the respective two sets of schemes, clearly indicating the importance of incorporating the magnetic balance explicitly.

Equations (199) and (207) are themselves coupled equations and are also coupled together. To solve them, the following micro-macro iterative scheme can be invoked [26]:

- Macro-iteration = 0:

  1. Set (a) $\mathbf{X}^{10} = 0$ and $\mathbf{A}^{10} = 0$ or (b) $\mathbf{X}^{10} = 0$ and $\mathbf{A}^{10} = \mathbf{A}_{\text{uc}}^{10}$, with $\mathbf{A}_{\text{uc}}^{10}$ being the uncoupled solution of Eq. (199);
  2. Construct $\mathbf{Q}^{10}[\mathbf{A}^{10}, \mathbf{X}^{10}]$ (209) with a given pair of $\mathbf{A}^{10}$ and $\mathbf{X}^{10}$. If $\mathbf{Q}^{10}[\mathbf{A}^{10}, \mathbf{X}^{10}]$ is fixed, Eq. (207) will become a standard linear Sylvester equation for $\mathbf{X}^{10}$ and can be solved non-iteratively by the means of the LAPACK routines ZGEES and ZTRSYL;
  3. With the *fixed* $\mathbf{X}^{10}$ from step 2, construct and solve Eq. (199) for $\mathbf{A}^{10}$ iteratively (micro iterations);

- Macro-iteration = Macro-iteration + 1:
  Repeat steps 2 and 3, with the $\mathbf{X}^{10}$ from step 2 and the $\mathbf{A}^{10}$ from step 3 of the previous macro iteration, until the shielding tensor has converged.

The above scheme converges typically within one or two macro-cycles. Still, however, the computation is more expensive than the four-component counterpart (149). An accurate estimate of $\mathbf{X}^{10}$ must hence be made available from the outset. Among the various possibilities [26], the superposition of the atomic matrices $\mathbf{X}_A^{10}$, again in the spirit of "from atoms to molecule" [45,46], is most recommended. Here, an atomic $\mathbf{X}_A^{10}$ matrix is constructed from a spherical and unpolarized atomic/ionic configuration and back-transformed to the atomic orbital space,

$$\mathbf{X}_A^{10\prime} = \mathbf{C}_A^{L00}\mathbf{X}_A^{10}(\mathbf{C}_A^{L00})^{-1}. \tag{210}$$

The molecular $\mathbf{X}^{10}$ matrix in the f_p-basis is then obtained as

$$\mathbf{X}^{10} \approx (\mathbf{C}^{L00})^{-1}(\sum_A^{\oplus}\mathbf{X}_A^{10\prime})\mathbf{C}^{L00} = \mathbf{X}_A^{10\oplus}, \tag{211}$$

where $\mathbf{C}_A^{L00}$ and $\mathbf{C}^{L00}$ are the zeroth-order large-component coefficients of the atomic and molecular orbitals, respectively. This ansatz is not only very efficient but also very accurate (typically less than 0.05 % in error). By contrast, the one-electron approximation $\mathbf{X}_{1e}^{10}$ (which is widely adopted in electronic structure calculations) leads to much larger errors in the total shielding, still accurate enough for chemical shift though [26].

With the available $\mathbf{X}^{10}$ and $\mathbf{A}^{10}$, the nuclear shielding can be calculated as [25,26]

$$\sigma_{uv}^{K}(Q_k) = \sigma_{uv}^{K,d}(Q_k) + \sigma_{uv}^{K,p}(Q_k), \tag{212}$$

$$\sigma_{uv}^{K,p}(Q_k) = \sigma_{uv}^{K,p1}(Q_k) + \sigma_{uv}^{K,p2}(Q_k), \tag{213}$$

$$\sigma_{uv}^{K,p2}(Q_k) = \sigma_{uv}^{K,p2a}(Q_k) + \sigma_{uv}^{K,p2b}(Q_k), \tag{214}$$

where

$$\begin{aligned}\sigma_{uv}^{K,d}(Q_k) =& (\mathbf{H}_{11}^{uv(K)} + \mathbf{H}_{12}^{uv(K)}\mathbf{X}^{00} + \mathbf{X}^{00\dagger}\mathbf{H}_{21}^{uv(K)} \\ &+ \mathbf{X}^{00\dagger}\mathbf{H}_{22}^{uv(K)}\mathbf{X}^{00})_{ii},\end{aligned} \tag{215}$$

$$\begin{aligned}\sigma_{uv}^{K,p1}(Q_k) =& -(\mathbf{H}_{11}^{v(K)} + \mathbf{H}_{12}^{v(K)}\mathbf{X}^{00} + \mathbf{X}^{00\dagger}\mathbf{H}_{21}^{v(K)} \\ &+ \mathbf{X}^{00\dagger}\mathbf{H}_{22}^{v(K)}\mathbf{X}^{00})_{ij}\tilde{\mathbf{S}}_{+,ji}^{10,u},\end{aligned} \tag{216}$$

$$\begin{aligned}\sigma_{uv}^{K,p2a}(Q_k) =& (\mathbf{H}_{11}^{v(K)} + \mathbf{H}_{12}^{v(K)}\mathbf{X}^{00} + \mathbf{X}^{00\dagger}\mathbf{H}_{21}^{v(K)} \\ &+ \mathbf{X}^{00\dagger}\mathbf{H}_{22}^{v(K)}\mathbf{X}^{00})_{ia}\mathbf{A}_{ai}^{10,u} + \text{c.c.},\end{aligned} \tag{217}$$

$$\sigma_{uv}^{K,p2b}(Q_k) = \left((\mathbf{H}_{12}^{v(K)} + \mathbf{X}^{00\dagger}\mathbf{H}_{22}^{v(K)})\mathbf{X}^{10,u}\right)_{ii} + \text{c.c.}, \tag{218}$$

$$\mathbf{H}_{\theta\rho}^{uv(K)} = \frac{\partial^2 \mathbf{H}_{\theta\rho}}{\partial B_u^{10} \partial \mu_{K,v}^{01}}|_{\vec{B}^{10}=\vec{\mu}_K^{01}=0}, \quad \theta, \rho = 1, 2, \tag{219}$$

$$\mathbf{H}_{\theta\rho}^{v(K)} = \frac{\partial \mathbf{H}_{\theta\rho}}{\partial \mu_{K,v}^{01}}|_{\vec{B}^{10}=\vec{\mu}_K^{01}=0}. \tag{220}$$

As an expectation value, $\sigma_{uv}^{K,d}(Q_k)$ (215) can be recognized as the diamagnetic term. The paramagnetic term $\sigma_{uv}^{K,p}(Q_k)$ (213) consists of three contributions, i.e., $\sigma_{uv}^{K,p1}(Q_k)$ (216) involving summation over only the occupied orbitals resulting from the explicit dependence of the basis on the magnetic field, $\sigma_{uv}^{K,p2a}(Q_k)$ (217) involving summation over the virtual orbitals, and $\sigma_{uv}^{K,p2b}(Q_k)$ (218) involving $\mathbf{X}^{10}$. Note that, for any finite basis, the so-defined $\sigma_{uv}^{K,d}(Q_k)$, $\sigma_{uv}^{K,p1}(Q_k)$, and $\sigma_{uv}^{K,p2}(Q_k)$ agree with the four-component counterparts shown in Eqs. (155), (161), and (162), respectively, thereby justifying perfectly the name of "exact two-component." In other words, $\sigma_{uv}^{K,p}(Q_k)$ (213) is just a recombination of the contributions of the PES and NES to the four-component paramagnetic term. Also, the computational costs for solving the coupled-perturbed equations (199) (with the fixed $\mathbf{X}^{10}$ (211)) and (149) (with the uncoupled treatment of NES) are very much the same. Therefore, the four- and two-component approaches for NMR shielding (and other properties) have been made identical in all the aspects of simplicity, accuracy, and efficiency, a point observed already several years ago [46].

The ZORA Approach for NMR Shielding

The X2C approach presented in section "The X2C Approach for NMR Shielding" is simpler, but more accurate, than the low-order DKH approach [13, 14]: The former does not involve any new integrals other than those entering the four-component theory, whereas the latter involves additional complicated operators, the evaluations of which are nontrivial. Due to this reason, before the advent of X2C [25, 26], the ZORA approach [11, 12] has been the workhorse for chemical shift calculations of heavy element compounds. In the language of section "The X2C Equations," ZORA (with a common gauge origin) amounts to adopting the following Z operator:

$$Z_{\text{ZORA}} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{K\vec{\sigma}\cdot\vec{\pi}}{2c} \end{pmatrix}, \quad K = \frac{2c^2}{2c^2 - V}, \quad \vec{\pi} = \vec{p} + \vec{A}^{10} + \vec{A}^{01}, \tag{221}$$

and the following $\mathbf{X}$ matrix:

$$\mathbf{X}_{\text{ZORA}} = \text{I}. \tag{222}$$

Inserting Eqs. (221) and (222) into Eqs. (196) and (197) gives rise to

$$\mathbf{L}^{\mathrm{ZORA}} = \mathbf{H}_{11} + \mathbf{H}_{12} + \mathbf{H}_{21} + \mathbf{H}_{22} \tag{223}$$

$$= \mathbf{V} + \mathbf{T}^{\mathrm{ZORA}}, \quad T^{\mathrm{ZORA}} = \frac{1}{2}\vec{\sigma}\cdot\vec{\pi}\,K\vec{\sigma}\cdot\vec{\pi}, \tag{224}$$

$$\mathbf{S}^{\mathrm{ZORA}} = \mathbf{M}_{11} + \mathbf{M}_{22} \approx \mathbf{S}. \tag{225}$$

The effective one-electron operator in Eq. (94) has been assumed for the Fock operator (172) when going from Eqs. (223) to (224). The relativistic metric (225) is further approximated to the nonrelativistic one. So, the final field-dependent ZORA equation reads

$$\mathbf{L}^{\mathrm{ZORA}}\mathbf{A} = \mathbf{S}\mathbf{A}\boldsymbol{\epsilon}. \tag{226}$$

When solving Eq. (226), a (fixed) model potential $\tilde{V}$, instead of the true potential V, is usually adopted for the denominator of K (221), so as to circumvent the origin dependence of the energy scale. Under this approximation, the first order of Eq. (226) reads

$$\mathbf{L}_{10}^{\mathrm{ZORA}} + \boldsymbol{\epsilon}^{00}\mathbf{A}^{10} = \mathbf{A}^{10}\boldsymbol{\epsilon}^{00} + \boldsymbol{\epsilon}^{10}, \tag{227}$$

where

$$\mathbf{L}_{10}^{\mathrm{ZORA}} = \mathbf{V}^{10} + \mathbf{T}_{10}^{\mathrm{ZORA}}, \tag{228}$$

$$(\mathbf{T}_{10}^{\mathrm{ZORA}})_{pq} = \frac{1}{2}\langle f_p|\vec{\sigma}\cdot\vec{A}^{10}K\vec{\sigma}\cdot\vec{p} + h.c.|f_q\rangle. \tag{229}$$

Equation (227) is not much different from the coupled-perturbed X2C equation (199), provided that the approximation (211) is adopted for the molecular $\mathbf{X}^{10}$ matrix. As shown before [26], this approximation is very accurate and costs essentially nothing. Therefore, X2C has very little overhead over ZORA (or DKH2).

Summary

It is shown that the theoretical calculation of rovibronically resolved NMR shielding requires a body-fixed molecular Hamiltonian that treats both electrons and nuclei quantum mechanically. Such a Hamiltonian that treats electrons fully relativistically and nuclei quasi-relativistically has been adopted here. The shielding expression is then featured by only a single term, the paramagnetism. The underlying reason is that the Dirac operator is linear instead of quadratic with respect to the mechanical momentum. However, the "missing" diamagnetism can be recaptured by incorporating the magnetic balance in various equivalent ways through the generic idea of "orbital decomposition." The magnetic balance can further be adapted to distributed gauge origins, leading to, e.g., magnetically balanced gauge-including

atomic orbitals (4C-MB-GIAO), where each magnetically balanced atomic orbital has its own local gauge origin on its center. Such 4C-MB-GIAO approaches can be transformed to X2C-MB-GIAO via a single matrix block diagonalization. Under very mild approximations (which lead to errors of less than 0.05 % in the total shielding and no errors in the chemical shift), the 4C-MB-GIAO and X2C-MB-GIAO approaches can be made computationally identical. They also have little overhead over the A2C approaches. It is therefore expected that while the A2C approaches remain suitable for the realm of applications they were designed for, over time they will retire or be retrofitted for the 4C/X2C-MB-GIAO approaches.

Acknowledgements The research of this work was supported by grants from the NSFC (Project Nos. 21033001, 21273011, 21173006, 21473002, and 21290192) and NSF (Grant No. CHE-1265833 to JA).

References

1. Ramsey NF (1950) Magnetic shielding of nuclei in molecules. Phys Rev 78:699
2. Sternheim MM (1962) Second-order effects of nuclear magnetic fields. Phys Rev 128:676
3. Feiock FD, Johnson WR (1968) Relativistic evaluation of internal diamagnetic fields for atoms and ions. Phys Rev Lett 21:785
4. Pyykkö P (1983) On the relativistic theory of NMR chemical shifts. Chem Phys 74:1
5. Kutzelnigg W, Liu W (2009) Relativistic theory of nuclear magnetic resonance parameters in a Gaussian basis representation. J Chem Phys 131:044129
6. Xiao Y, Sun Q, Liu W (2012) Fully relativistic theories and methods for NMR parameters. Theor Chem Acc 131:1080
7. Nakatsuji H, Takashima H, Hada M (1995) Spin-orbit effect on the magnetic shielding constant using the ab initio UHF method. Chem Phys Lett 233:95
8. Fukui H, Baba T, Inomata H (1996) Calculation of nuclear magnetic shieldings. X. Relativistic effects. J Chem Phys 105:3175
9. Fukui H, Baba T, Inomata H (1997) Erratum: Calculation of nuclear magnetic shieldings. X. Relativistic effects [J Chem Phys 105 3175 (1996)]. J Chem Phys 106:2987
10. Kutzelnigg W (1999) Relativistic corrections to magnetic properties. J Comput Chem 20:1199
11. Wolff SK, Ziegler T, van Lenthe E, Baerends EJ (1999) Density functional calculations of nuclear magnetic shieldings using the zeroth-order regular approximation (ZORA) for relativistic effects: ZORA nuclear magnetic resonance. J Chem Phys 110:7689
12. Rodriguez-Fortea A, Alemany P, Ziegler T (1999) Density functional calculations of NMR chemical shifts with the inclusion of spin-orbit coupling in tungsten and lead compounds. J Phys Chem A 103:8288
13. Fukuda R, Hada M, Nakatsuji H (2003) Quasirelativistic theory for the magnetic shielding constant. I. Formulation of Douglas–Kroll–Hess transformation for the magnetic field and its application to atomic systems. J Chem Phys 118:1015
14. Melo JI, Ruiz de Azúa MC, Peralia JE, Scuseria GE (2005) Relativistic calculation of indirect NMR spin-spin couplings using the Douglas-Kroll-Hess approximation. J Chem Phys 123:204112
15. Kudo K, Maeda H, Kawakubo T, Ootani Y, Funaki M, Fukui H (2006) Relativistic calculation of nuclear magnetic shielding using normalized elimination of the small component. J Chem Phys 124:224106
16. Maeda H, Ootani Y, Fukui H (2007) Relativistic calculation of nuclear magnetic shielding tensor using the regular approximation to the normalized elimination of the small component. II. Consideration of perturbations in the metric operator. J Chem Phys 126:174102

17. Dyall KG (1997) Interfacing relativistic and nonrelativistic methods. I. Normalized elimination of the small component in the modified Dirac equation. J Chem Phys 106:9618
18. Autschbach J (2013) The role of the exchange-correlation response kernel and scaling corrections in relativistic density functional nuclear magnetic shielding calculations with the zeroth-order regular approximation. Mol Phys 111:2544
19. Edlund U, Lejon T, Pyykkö P, Venkatachalam TK, Buncel E (1987) Lithium-7, silicon-29, tin-119, and lead-207 NMR studies of phenyl-substituted Group 4 anions. J Am Chem Soc 109:5982
20. Autschbach J, Zheng S (2009) Relativistic computations of NMR parameters from first principles: theory and applications. Annu Rep NMR Spectrosc 67:1
21. Vaara J, Malkina OL, Stoll H, Malkin VG, Kaupp M (2001) Study of relativistic effects on nuclear shieldings using density-functional theory and spin–orbit pseudopotentials. J Chem Phys 114:61
22. Autschbach J (2010) Relativistic effects on magnetic resonance parameters and other properties of inorganic molecules and metal complexes. In: Barysz M, Ishikawa Y (eds) Relativistic methods for chemists. Challenges and advances in computational chemistry and physics, chap. 12, vol 10. Springer, Dordrecht, pp 521–598
23. Pickard CJ, Mauri F (2001) All-electron magnetic response with pseudopotentials: NMR chemical shifts. Phys Rev B 63:245101
24. Autschbach J (2013) In: Contreras RH (ed) High resolution nuclear magnetic resonance parameters for understanding molecules and their electronic structure. Science & technology of atomic, molecular, condensed matter & biological systems, vol 3. Elsevier, Amsterdam, pp 69–117
25. Sun Q, Liu W, Xiao Y, Cheng L (2009) Exact two-component relativistic theory for nuclear magnetic resonance parameters. J Chem Phys 131:081101
26. Sun Q, Xiao Y, Liu W (2012) Exact two-component relativistic theory for NMR parameters: general formulation and pilot application. J Chem Phys 137:174105
27. Autschbach J, Peng D, Reiher M (2012) Two-component relativistic calculations of electric-field gradients using exact decoupling methods: spin–orbit and picture-change effects. J Chem Theor Comput 8:4239
28. Xiao Y, Liu W (2013) Body-fixed relativistic molecular Hamiltonian and its application to nuclear spin-rotation tensor. J Chem Phys 138:134104
29. Xiao Y, Liu W (2013) Body-fixed relativistic molecular Hamiltonian and its application to nuclear spin-rotation tensor: linear molecules. J Chem Phys 139:034113
30. Aucar GA, Saue T, Visscher L, Jensen HJAa (1999) On the origin and contribution of the diamagnetic term in four-component relativistic calculations of magnetic properties. J Chem Phys 110:6208
31. Pyper NC, Zhang ZC (1988) Exact relativistic analogues of the non-relativistic hyperfine structure operators: I. Theory. Mol Phys 64:933
32. Szmytkowski R (2002) Larmor diamagnetism and Van Vleck paramagnetism in relativistic quantum theory: the Gordon decomposition approach. Phys Rev A 65:032112
33. Pyper NC (1983) The relativistic theory of the chemical shift. Chem Phys Lett 96:204
34. Kutzelnigg W (2003) Diamagnetism in relativistic theory. Phys Rev A 67:032109
35. Xiao Y, Liu W, Cheng L, Peng D (2007) Four-component relativistic theory for nuclear magnetic shielding constants: critical assessments of different approaches. J Chem Phys 126:214101
36. Xiao Y, Peng D, Liu W (2007) Four-component relativistic theory for nuclear magnetic shielding constants: the orbital decomposition approach. J Chem Phys 126:081101
37. Komorovsky S, Repisky M, Malkina OL, Malkin VG, Malkin Ondík I, Kaupp M (2008) A fully relativistic method for calculation of nuclear magnetic shielding tensors with a restricted magnetically balanced basis in the framework of the matrix Dirac–Kohn–Sham equation. J Chem Phys 128:104101
38. Cheng L, Xiao Y, Liu W (2009) Four-component relativistic theory for NMR parameters: unified formulation and numerical assessment of different approaches. J Chem Phys 130:144102

39. London F (1937) Théorie quantique des courants interatomiques dans les combinaisons aromatiques. J Phys Rad 8:397
40. Wolinski K, Hinton JF, Pulay P (1990) Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. J Am Chem Soc 112:8251
41. Cheng L, Xiao Y, Liu W (2009) Four-component relativistic theory for nuclear magnetic shielding: Magnetically balanced gauge-including atomic orbitals. J Chem Phys 131:244113
42. Gao J, Liu W, Song B, Liu C (2004) Time-dependent four-component relativistic density functional theory for excitation energies. J Chem Phys 121:6658
43. Gao J, Zou W, Liu W, Xiao Y, Peng D, Song B, Liu C (2005) Time-dependent four-component relativistic density-functional theory for excitation energies. II. The exchange-correlation kernel. J Chem Phys 123:054102
44. Peng D, Zou W, Liu W (2005) Time-dependent quasirelativistic density-functional theory based on the zeroth-order regular approximation. J Chem Phys 123:144101
45. Liu W, Peng D (2006) Computation of dipole, quadrupole, and octupole surfaces from the variational two-electron reduced density matrix method. J Chem Phys 125:044102 ; (E) 125:149901 (2006)
46. Peng D, Liu W, Xiao Y, Cheng L (2007) Making four-and two-component relativistic density functional methods fully equivalent based on the idea of “from atoms to molecule”. J Chem Phys 127:104106
47. Kutzelnigg W, Liu W (2005) Quasirelativistic theory equivalent to fully relativistic theory. J Chem Phys 123:241102
48. Liu W (2010) Ideas of relativistic quantum chemistry. Mol Phys 108:1679
49. Liu W, Peng D (2009) Exact two-component Hamiltonians revisited. J Chem Phys 131:031104
50. Liu W (2014) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59

Relativistic Theory of Nuclear Spin-Rotation Tensor

22

Yunlong Xiao, Wenjian Liu, and Kenneth Ruud

Contents

Introduction . . . 694
Conventions . . . 697
Relativistic Theory of the NSR Tensor . . . 698
The Molecular Hamiltonian H^{BFF} . . . 698
Perturbation Expansion of H^{BFF} . . . 699
The Effective Hamiltonian H_{eff}^{11} . . . 701
The NSR Tensor . . . 702
The Perturbation-Dependent Basis Set . . . 706
Rotational London Orbitals . . . 706
Relativistic Rotational London Orbitals . . . 707
Relativistic Mapping Between NMR and NSR Tensors . . . 712
Illustrative Examples . . . 715
NSR Constants of Hydrogen Halide . . . 715
Absolute Nuclear Shielding Constants . . . 717
Summary . . . 720
References . . . 720

Y. Xiao (✉) • W. Liu
Beijing National Laboratory for Molecular Sciences, Institute of Theoretical and Computational Chemistry, State Key Laboratory of Rare Earth Materials Chemistry and Applications, College of Chemistry and Molecular Engineering, and Center for Computational Science and Engineering, Peking University, Beijing, People's Republic of China
e-mail: xiaoyl@pku.edu.cn; liuwjbdf@gmail.com

K. Ruud
Department of Chemistry, Centre for Theoretical and Computational Chemistry, University of Tromsø—The Arctic University of Norway, Tromsø, Norway
e-mail: kenneth.ruud@uit.no

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_16

Abstract

The relativistic theory for the nuclear spin-rotation (NSR) tensor is formulated based on the body-fixed molecular Hamiltonian that treats quantum electrons fully relativistically and quantum nuclei quasi-relativistically. The resulting expression for the NSR tensor is then compared with that for the nuclear magnetic shielding tensor, so as to establish a relativistic mapping between them. This relativistic mapping is very robust and permits an easy and direct translation of experimental NSR tensors into semi-experimental absolute nuclear shielding tensors which are otherwise difficult to obtain experimentally. In contrast, the well-known nonrelativistic mapping (Ramsey–Flygare relation) between the nuclear shielding and NSR tensors breaks down even for relatively light elements. Some classic systems are taken as examples to elucidate the concepts.

Keywords

Relativistic molecular Hamiltonian • Nuclear spin-rotation • Nuclear magnetic resonance • Relativistic mapping • Rotational London orbitals

Introduction

The so-called nuclear spin-rotation (NSR) tensor, M_{uv}^{K}, is a parameter for defining the effective Hamiltonian

$$H_{\text{eff}}^{\text{NSR}} = \sum_{K} I_{Ku} M_{uv}^{K} J_{v}, \quad (u, v = x, y, z), \tag{1}$$

where $\vec{I}_K$ is the spin operator for nucleus K, while $\vec{J}$ is the molecular angular momentum operator but with the nuclear spin excluded. Physically, the effective Hamiltonian Eq. (1) describes the interactions between the nuclear magnetic dipole moment of the target nucleus and the effective magnetic field of the rotating molecule, which includes the magnetic field induced by the rotation of the nuclear framework and that by the polarization of the electron cloud. Such interactions are most prominent in high-resolution microwave studies of rotational spectra, where these interactions appear as hyperfine structures in the rotational spectra [1–3]. These high-resolution microwave spectra represent a valuable source of highly accurate NSR tensors in the gas phase. For this reason, NSR tensors have been important in benchmarking quantum chemical methods. Using highly accurate coupled-cluster calculations with rovibrational corrections, excellent agreement can be obtained between theory and experiment for the NSR tensors [4, 5]. Ab initio calculations are nowadays also being an important component in the analysis of rotational spectra [4]. By contrast, density functional theory (DFT) with approximate functionals performs in general less well for the calculation of NSR tensors [6], the accuracy being, however, comparable to that observed for the nuclear magnetic shielding tensors [6].

NSR tensors can also be determined directly from nuclear magnetic resonance (NMR) spectroscopy experiments, as they are responsible for important relaxation mechanisms for the nuclear spins [7, 8]. However, the so-determined NSR tensors are much less accurate than those obtained from rotational spectroscopy.

In the nonrelativistic (NR) framework, the operator describing the polarization of the electron cloud due to the molecular rotation reads

$$d_v^{10}(\mathrm{NR,NSR}) = -(I^0)_{vv}^{-1}\Big[(\vec{r} - \vec{R}_C) \times \vec{p}\Big]_v, \tag{2}$$

where $\vec{R}_C$ refers to the nuclear center of mass (NCM), I^0 the inertia tensor at the equilibrium structure, and $\vec{p}$ the electronic momentum. It is interesting to see that this operator is proportional to the operator describing the polarization of the electron cloud due to a homogeneous external magnetic field

$$d_v^{10}(\mathrm{NR,NMR}) = \frac{1}{2}\Big[(\vec{r} - \vec{R}_g) \times \vec{p}\Big]_v, \tag{3}$$

provided that the global gauge origin $\vec{R}_g$ is placed at $\vec{R}_C$. Further, in view of the proportionality between the nuclear magnetic dipole moment $\vec{\mu}_K$ and the nuclear spin $\vec{I}_K$, i.e.,

$$\vec{\mu}_K = g_K \mu_\mathrm{n} \vec{I}_K, \tag{4}$$

with g_K and μ_n being the nuclear g-factor and magneton, respectively, it can readily be deduced [9] that the electronic paramagnetic contributions to the NMR shielding $\sigma_{vu,\mathrm{eq}}^{K,\mathrm{p}}(\mathrm{NR})$ and the NSR coupling $M_{uv,\mathrm{eq}}^{K,\mathrm{p}}(\mathrm{NR})$ are related in a simple way, viz.,

$$\sigma_{vu,\mathrm{eq}}^{K,\mathrm{p}}(\mathrm{NR}) = -\frac{I_{vv}^0}{2g_K\mu_\mathrm{n}} M_{uv,\mathrm{eq}}^{K,\mathrm{p}}(\mathrm{NR}). \tag{5}$$

By replacing the theoretical paramagnetic term $M_{uv,\mathrm{eq}}^{K,\mathrm{p}}(\mathrm{NR})$ with the experimental counterpart, $M_{uv,\mathrm{eq}}^{K,\mathrm{p}}(\mathrm{exp}) = M_{uv,\mathrm{eq}}^{K}(\mathrm{exp}) - M_{uv,\mathrm{eq}}^{K,\mathrm{d}}(\mathrm{NR})$, one would obtain the following equality:

$$\sigma_{vu,\mathrm{eq}}^{K,\mathrm{p}}(\mathrm{exp,NR}) = -\frac{I_{vv}^0}{2g_K\mu_\mathrm{n}}\Big[M_{uv,\mathrm{eq}}^{K}(\mathrm{exp}) - M_{uv,\mathrm{eq}}^{K,\mathrm{d}}(\mathrm{NR})\Big]. \tag{6}$$

Further adding in the diamagnetic term $\sigma_{vu,\mathrm{eq}}^{K,\mathrm{d}}(\mathrm{NR})$ calculated with the gauge origin $\vec{R}_g$ placed at $\vec{R}_C$, i.e.,

$$\sigma_{vu,\mathrm{eq}}^{K,\mathrm{d}}(\mathrm{NR}) = \langle\Psi^{00}| \sum_i \frac{\vec{r}_{iC} \cdot \vec{r}_{iK}\delta_{uv} - r_{iC,u} r_{iK,v}}{r_{iK}^3} |\Psi^{00}\rangle, \quad \vec{r}_{iA} = \vec{r}_i - \vec{R}_A, \tag{7}$$

gives rise to the well-known Ramsey–Flygare relation between the "experimental" (more precisely, semi-experimental) magnetic shielding $\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp,NR})$ and the experimental NSR coupling $M^K_{uv,\mathrm{eq}}(\mathrm{exp})$,

$$\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp,NR}) = -\frac{I^0_{vv}}{2g_K\mu_\mathrm{n}}\Big[M^K_{uv,\mathrm{eq}}(\mathrm{exp}) - M^{K,\mathrm{d}}_{uv,\mathrm{eq}}(\mathrm{NR})\Big] + \sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}}(\mathrm{NR}). \tag{8}$$

Due to the importance of absolute nuclear magnetic shielding tensors as benchmarks for theoretical calculations [10–12] as well as in the determination of nuclear magnetic moments [13, 14], the above relation Eq. (8) has been extensively used to determine "experimental" absolute shielding tensors; see Refs. [11, 12, 15] for some recent examples. It can be shown [1–3] that, by shifting the gauge origin from the NCM $\vec{R}_C$ to the position $\vec{R}_K$ of the active nucleus K and further introducing the free-atom (FA) approximation, Eq. (8) can be simplified to

$$\sigma^K_{\mathrm{eq}}(\mathrm{exp,NR}) \approx \sigma^{K,\mathrm{FA}}(\mathrm{exp}) - \sum_{v=x,y,z}\frac{I^0_{vv}}{6g_K\mu_\mathrm{n}}M^K_{vv,\mathrm{eq}}(\mathrm{exp}). \tag{9}$$

Then, only experimentally available data is needed to derive the isotropic shielding $\sigma^K_{\mathrm{eq}}(\mathrm{exp,NR})$. Although it is an approximation to Eq. (8), the relation Eq. (9) has often been used to determine the "experimental" absolute magnetic shielding constants [1–3], which are otherwise difficult to obtain experimentally.

Being nonrelativistic in nature, the validity of relation Eq. (5) and hence the relations Eq. (8)/(9) in the relativistic domain was long questioned [16, 17]. However, the situation became clear only after the relativistic theory of NSR tensors was formulated [18–21]. Specifically, the relativistic counterparts of Eqs. (2) and (3) read [20, 21]

$$d^{10}_v(\mathrm{NSR}) = -(I^0)^{-1}_{vv}\left[\Big((\vec{r}-\vec{R}_C)\times\vec{p}\Big)_v + \frac{1}{2}\Sigma_v\right], \tag{10}$$

$$d^{10}_v(\mathrm{NMR}) = \frac{1}{2}\Big[(\vec{r}-\vec{R}_g)\times c\vec{\alpha}\Big]_v. \tag{11}$$

Since $d^{10}_v(\mathrm{NSR})$ Eq. (10) and $d^{10}_v(\mathrm{NMR})$ Eq. (11) are diagonal and off-diagonal operators, respectively, it is clear that the electronic paramagnetic terms of the NSR coupling and NMR shielding tensors are very different in the relativistic regime. Nonetheless, a "relativistic mapping" [20–23] between the two quantities is still possible by analyzing the internal relationship between the relativistic theories of NMR shielding and NSR coupling tensors. The "relativistic mapping" is very robust in the sense that it is very insensitive to the quality of both one- and many-particle basis sets [23], such that a highly accurate $\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp})$ can readily be obtained by combining the experimental $M^K_{uv,\mathrm{eq}}(\mathrm{exp})$ and the theoretical quantities $M^{K,\mathrm{d}}_{uv,\mathrm{eq}}$, $\sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}}$, and $\Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}}$ (*vide post*) calculated simply at a low level of theory.

To make the subsequent presentation compact, the conventions for the notation are first summarized in section "Conventions." The relativistic theory of the NSR tensor is reviewed in section "Relativistic Theory of the NSR Tensor." At variance

Table 1 Summary of the acronyms

Acronym	Full name
NSR	Nuclear spin-rotation
NMR	Nuclear magnetic resonance
LAB	Laboratory frame of reference
BFF	Body-fixed frame of reference
NCM	Nuclear center of mass
DFT	Density functional theory
NR	Nonrelativistic
KS	Kohn–Sham
DKS	Dirac–Kohn–Sham
DHF	Dirac–Hartree–Fock
SSO	Spin-same-orbit
SOO	Spin-other-orbit
KB	Kinetic balance
RKB	Restricted kinetic balance
KU	Kinetic unbalance
EFUT	External field-dependent unitary transformation
nrl	Nonrelativistic limit
FA	Free-atom
HALA	Heavy-atom effect on the light atoms
AO	Atomic orbital
LO	London orbital
RLO	Rotational London orbital
ppm	Part per million

with the semi-classical Born–Oppenheimer decoupling approach [18, 19], the present rigorous formulation is based on the relativistic molecular Hamiltonian in the body-fixed frame (BFF) of reference [20, 21], which treats both electrons and nuclei as quantum particles. In section "The Perturbation-Dependent Basis Set," a proper perturbation-dependent basis set, the relativistic analog [24] of the rotational London orbitals (RLO) [25], is introduced to speed up the basis set convergence. The "relativistic mapping" [20–23] between the NSR coupling and NMR shielding tensors is then discussed in section "Relativistic Mapping Between NMR and NSR Tensors." Some illustrative examples are provided in section "Illustrative Examples" for both the NSR coupling and NMR shielding constants. The entry ends with a brief summary in section "Summary." Summary of the acronyms can be found in Table 1.

Conventions

All equations will be written in SI-based atomic units. The charge and mass of nucleus K are denoted as Z_K and m_K, respectively. The nuclear ($R_{Ku}, u = x, y, z$) and electronic (r_{iu}) coordinates refer to the BFF of reference, with the origin placed at the NCM along the principal axes of inertia. The relative coordinate $\vec{r}_{ab}$ is defined as $\vec{r}_a - \vec{r}_b$. The moment of inertia tensor I in the BFF is defined as

$$I_{uv} = \varepsilon_{uu'w}\varepsilon_{vv'w}\sum_K m_K R_{Ku'} R_{Kv'}, \tag{12}$$

where ε_{uvw} is the Levi–Civita second-rank antisymmetric tensor. The inertia tensor I in Eq. (12) at the reference (equilibrium) geometry will be denoted as I^0. The so-called effective moment of inertia tensor I' defined as

$$I' = I'' \left(I^0\right)^{-1} I'', \tag{13}$$

$$I''_{uv} = I^0_{uv} + \varepsilon_{uu'w}\varepsilon_{vv'w}\sum_K m_K^{1/2} R^0_{Ku'} L_{(Kv'),k} Q_k \tag{14}$$

will also be invoked. Here, Q_k are the normal coordinates and $L_{(Kv'),k}$ the transformation matrix from the Cartesian displacements to the normal coordinates; see Ref. [20]. The Einstein summation convention over repeated indices is always employed.

Relativistic Theory of the NSR Tensor

The proper formulation of the NSR tensor M^K_{uv} Eq. (1) involves four steps. The very first step is to transform the relativistic molecular Hamiltonian H^{LAB} in the laboratory (LAB) frame to H^{BFF} in the BFF. This is outlined in section "The Molecular Hamiltonian H^{BFF}." The second step (see section "Perturbation Expansion of H^{BFF}") amounts to decomposing the resulting H^{BFF} Hamiltonian into a perturbation series

$$H^{\mathrm{BFF}} = H^{00} + V^{10} + V^{01} + V^{11} + \cdots. \tag{15}$$

The third step is to construct a mixed second-order effective Hamiltonian H^{11}_{eff} (see section "The Effective Hamiltonian H^{11}_{eff}"), which can finally be specified to M^K_{uv} in Eq. (1) based on the perturbation expression Eq. (15) (see section "The NSR Tensor"). Some adaptations need to be made for linear molecules [21], but these will not be discussed any further here.

The Molecular Hamiltonian H^{BFF}

Since the derivation of the BFF molecular Hamiltonian H^{BFF}, which treats quantum electrons fully relativistically and quantum nuclei quasi-relativistically, has recently been discussed in depth for both nonlinear [20] and linear [21] molecules, only a brief sketch is given here. The point of departure is the molecular Hamiltonian H^{LAB} written in the LAB frame. It consists of six terms [20]:

$$H^{\mathrm{LAB}} = T_{\mathrm{n0}}^{\mathrm{LAB}} + T_{\mathrm{n2}}^{\mathrm{LAB}} + T_{\mathrm{e}}^{\mathrm{LAB}} + H_{\mathrm{ee}}^{\mathrm{LAB}} + H_{\mathrm{ne}}^{\mathrm{LAB}} + H_{\mathrm{nn}}^{\mathrm{LAB}}, \tag{16}$$

where the first two terms represent the nonrelativistic nuclear kinetic energy and its first-order relativistic correction, whereas the following four terms are the electronic kinetic energy, electron–electron, nucleus–electron, and nucleus–nucleus interactions, respectively. For explicit expressions, see Ref. [20]. The Hamiltonian H^{LAB} Eq. (16) can be transformed to H^{BFF} through the following unitary transformation:

$$H^{\mathrm{BFF}} = \mu^{-1/4} U_S H^{\mathrm{LAB}} U_S^{-1} \mu^{1/4}, \tag{17}$$

where μ is the determinant of the inverse of the effective inertial tensor I' Eq. (13). The main purpose of the μ-transformation is to simplify the vibrational kinetic energy operator as well as the integrals over the vibrational normal coordinates. The U_S operator in Eq. (17) is defined as

$$U_S = e^{i\gamma S_z^{\mathrm{mol}}} e^{i\beta S_y^{\mathrm{mol}}} e^{i\alpha S_z^{\mathrm{mol}}}, \quad \vec{S}^{\mathrm{mol}} = \sum_i \vec{S}_i + \sum_K \vec{I}_K, \tag{18}$$

where $\vec{S}^{\mathrm{mol}}$ is the total spin operator of the molecule, with the electronic spin $\vec{S}_i$ being

$$\vec{S}_i = \frac{1}{2}\vec{\Sigma}_i, \quad \vec{\Sigma}_i = \begin{pmatrix} \vec{\sigma}_i & 0 \\ 0 & \vec{\sigma}_i \end{pmatrix}, \tag{19}$$

and α, β, and γ the Euler angles. Under such a U_S-transformation, the spin space of both electrons and nuclei is rotated from the LAB frame to the BFF. It is the BFF Hamiltonian H^{BFF} Eq. (17) that is used as the reference for the perturbation expansion Eq. (15).

Perturbation Expansion of H^{BFF}

The Hamiltonian H^{BFF} Eq. (17) can be expanded in powers of the perturbation parameters $\vec{I}_K$ and $\vec{J}$, which are denoted by superscripts '01' and '10', respectively. A natural choice for the zeroth-order term H^{00} is

$$H^{00} = H^{\mathrm{BFF}}|_{\vec{I}_K = \vec{J} = 0} + H^{\mathrm{rot}}, \tag{20}$$

$$H^{\mathrm{rot}} = \frac{1}{2} \sum_{v=x,y,z} J_v B_I J_v, \quad B_I = (I'^{-1})_{vv}. \tag{21}$$

The first term of H^{00} Eq. (20) is obtained from H^{BFF} Eq. (17) by setting $\vec{J} = 0$ and $\vec{I}_K = 0$ for all the nuclei, whereas the second term H^{rot} refers to a rigid rotor.

The latter implies that the effective Hamiltonian is to be constructed for a rovibronic state. One can also decide to remove the rotational term from H^{00}, so as to limit the model space of the effective Hamiltonian only to a vibronic state. For the present choice Eq. (20), it is more correct to state that it is the coupling between $\vec{J}$ and other degrees of freedom than $\vec{J}$ itself that is treated as the '10' perturbation.

The V^{01} operator in Eq. (15) contains those terms in H^{BFF} Eq. (17) [20] that are linear with respect to the nuclear spin $\vec{I}_K$ but independent of the molecular angular momentum $\vec{J}$. Retaining only the term that is of first order in the electron–nucleus mass ratio leads to

$$V^{01} = \sum_K I_{Ku} D_u^{01(K)}, \tag{22}$$

$$D_u^{01(K)} = \sum_i \frac{g_K \mu_\mathrm{n}}{c r_{iK}^3} \left(\vec{r}_{iK} \times \vec{\alpha}_i\right)_u. \tag{23}$$

The V^{10} operator in Eq. (15) contains those terms in H^{BFF} Eq. (17) [20] that are linear with respect to the molecular angular momentum $\vec{J}$ but independent of the nuclear spin $\vec{I}_K$. Again, retaining only the terms that are of first order in the electron–nucleus mass ratio leads to

$$V^{10} = D_v^{10} J_v, \tag{24}$$

$$D_v^{10} = -\sum_i \left(l_{iv} + \frac{1}{2}\Sigma_{iv} + j_v^{\mathrm{vib}}\right) B_I + D_{\mathrm{ne,G(O)},v}^{10}, \tag{25}$$

$$l_{iv} = \left[\left(\vec{r}_i - \vec{R}_C\right) \times \vec{p}_i\right]_v, \tag{26}$$

$$j_v^{\mathrm{vib}} = -i\,\eta_{kl}^v Q_k \frac{\partial}{\partial Q_l}, \quad \eta_{kl}^v = \epsilon_{vuw} \sum_K L_{(Ku),k} L_{(Kw),l}, \tag{27}$$

$$D_{\mathrm{ne,G(O)},v}^{10} = \sum_{iK} \frac{Z_K}{m_K c} \frac{\alpha_{iu}}{r_{iK}} A_{uv}^{(K)}, \tag{28}$$

$$A_{uv}^{(K)} = -m_K \epsilon_{uv'w} R_{Kv'}^0 (I''^{-1})_{wv}. \tag{29}$$

The first term of D_v^{10} Eq. (25) describes the coupling between the electronic orbital motion and the molecular rotation and is the leading term. The second term of D_v^{10} describes the coupling between the electronic spin and the molecular rotation and is of $\mathcal{O}(c^{-2})$ in a relative sense. The third term accounts for the Coriolis coupling, which vanishes identically for diatomic molecules [20] and is also very small for polyatomic molecules, especially when the electronic wave function does not vary much near the equilibrium geometry. The $D_{\mathrm{ne,G(O)}}^{10}$ term Eq. (28) arises from the Gaunt interaction between the electrons and nuclei. It is formally of the same order as the spin term in D_v^{10} Eq. (25) [20,21] but has no discernible contribution [26] due

to the mismatch in parity [23]. Therefore, D_v^{10} Eq. (25) can safely be approximated as

$$D_v^{10} \approx \sum_i d_v^{10}, \tag{30}$$

$$d_v^{10} = -B_I \left(l_{iv} + \frac{1}{2} \Sigma_{iv} \right), \tag{31}$$

which reduces to Eq. (10) at the equilibrium geometry.

The mixed second-order operator V^{11} in Eq. (15) reads

$$V^{11} = \sum_K I_{Ku} D_{uv}^{11,K} J_v, \tag{32}$$

$$D_{uv}^{11,K} = D_{uv}^{11,\mathrm{n},K} + D_{uv}^{11,\mathrm{e},K}, \tag{33}$$

where the purely nuclear term $D_{uv}^{11,\mathrm{n},K}$ stems from the nucleus–nucleus Coulomb spin-same-orbit (SSO) and Breit spin-other-orbit (SOO) interactions, viz.,

$$D_{uv}^{11,\mathrm{n},K} = D_{\mathrm{nn,C(SSO)},uv}^{11(K)} + D_{\mathrm{nn,B(SOO)},uv}^{11(K)}, \tag{34}$$

$$D_{\mathrm{nn,C(SSO)},uv}^{11(K)} = -\sum_{L(L\neq K)} \frac{g_K \mu_\mathrm{n} \gamma_K Z_L}{m_K c^2} \epsilon_{uwv'} \frac{R_{KL,w}}{R_{KL}^3} A_{v'v}^{(K)}, \tag{35}$$

$$D_{\mathrm{nn,B(SOO)},uv}^{11(K)} = \sum_{L(L\neq K)} \frac{g_K \mu_\mathrm{n} Z_L}{m_L c^2} \epsilon_{uwv'} \frac{R_{KL,w}}{R_{KL}^3} A_{v'v}^{(L)}. \tag{36}$$

Here, γ_K is the Thomas precession. The electronic term $D_{uv}^{11,\mathrm{e},K}$ in Eq. (33) arises from the electron–nucleus Coulomb SSO interaction,

$$D_{uv}^{11,\mathrm{e},K} = D_{\mathrm{ne,C(SSO)},uv}^{11(K)} = -\sum_i \frac{g_K \mu_\mathrm{n} \gamma_K}{m_K c^2} \epsilon_{uwv'} \frac{r_{iK,w}}{r_{iK}^3} A_{v'v}^{(K)}. \tag{37}$$

At the (calculated) equilibrium geometry, this contribution will cancel the $D_{\mathrm{nn,C(SSO)}}^{11(K)}$ Eq. (35) term [18, 20], thereby leaving $D_{\mathrm{nn,B(SOO)}}^{11(K)}$ Eq. (36) the only diamagnetic term.

The Effective Hamiltonian H_{eff}^{11}

Given the perturbation form Eq. (15) of the H^{BFF} Hamiltonian, it is assumed that the zeroth-order problem can be solved,

$$H^{00}|p^{00}\rangle = E_p^{00}|p^{00}\rangle. \tag{38}$$

Only a specific n_d-fold degenerate subspace V_{mod}, i.e.,

$$E_p^{00} = \text{const.}, \quad p = 1, \cdots, n_d, \tag{39}$$

is of interest here. The degeneracy of the manifold V_{mod} will split in the presence of the perturbations V^{10}, V^{01}, and V^{11}. The energies can be obtained exactly by solving the reduced eigenvalue problem

$$H_{\text{eff}}|q\rangle = E_q|q\rangle, \tag{40}$$

where the effective Hamiltonian H_{eff} is defined in V_{mod}, whose eigenvectors $|q\rangle$ are just linear combinations of $\{|p^{00}\rangle\}$. For the present purpose, it is the mixed second-order term of H_{eff} that is wanted, viz.,

$$H_{\text{eff}}^{\text{NSR}} = H_{\text{eff}}^{11} = H_{\text{eff}}^{11,\text{d}} + H_{\text{eff}}^{11,\text{p}}, \tag{41}$$

$$H_{\text{eff}}^{11,\text{d}} = P^{00}V^{11}P^{00}, \tag{42}$$

$$H_{\text{eff}}^{11,\text{p}} = P^{00}V^{10}Q^{00}\frac{1}{E^{00} - H^{00}}Q^{00}V^{01}P^{00} + h.c., \tag{43}$$

where the projection operators are defined as

$$P^{00} = \sum_{p \in V_{\text{mod}}} |p^{00}\rangle\langle p^{00}|, \tag{44}$$

$$Q^{00} = 1 - P^{00} = \sum_{q \notin V_{\text{mod}}} |q^{00}\rangle\langle q^{00}|. \tag{45}$$

The NSR Tensor

The Model Space

To apply the previous effective Hamiltonian H_{eff}^{11} Eq. (41) to the NSR tensor, the model space V_{mod} still remains to be specified. In the Born–Oppenheimer approximation, the vectors in the model space are chosen as

$$|p^{00}\rangle = |i_{\text{vib}}^{00} \otimes i_{\text{ele}}^{00} \otimes J_z^{\text{LAB}} \otimes \Phi_m\rangle, \tag{46}$$

where $|i_{\text{vib}}^{00} \otimes i_{\text{ele}}^{00}\rangle$ stands for the nondegenerate, zeroth-order vibronic state, $\{|J_z^{\text{LAB}}\rangle\}$ denotes the $2J+1$ degenerate rotational states (with $z = -J, -J+1, \cdots, J-1, J$ being the z component of the rotational quantum number in the LAB frame), whereas $\{|\Phi_m\rangle\}$ is a complete and orthonormal basis describing different nuclear spin states, the dimension of which is $n_{\text{spin}} = \prod_K (2I_K + 1)$. Therefore, the dimension of the model space V_{mod} is $n_d = (2J+1)n_{\text{spin}}$.

The electronic state $|i_{\rm ele}^{00}\rangle$ with energy $E_{\rm e}(Q_k)$ can be obtained by solving the electronic Dirac–Coulomb or Dirac–Coulomb–Breit equation within the clamped nucleus approximation.

The rotational states, with rotational quantum number J, can be calculated for any given Q_k (because Q_k determines I' and hence $H^{\rm rot}$ Eq. (21)). The corresponding eigenvalue problem is

$$H^{\rm rot}|J^2 J_z^{\rm LAB}\eta\rangle = E^{\rm rot}_{J^2,\eta}(Q_k)|J^2 J_z^{\rm LAB}\eta\rangle, \tag{47}$$

where $\eta = 1, 2, \cdots, 2J+1$ denotes the energy levels, each of which is $(2J+1)$-fold degenerate, corresponding to the range of $J_z^{\rm LAB}$ from $-J$ to J. Since J^2, $J_z^{\rm LAB}$ and J_z mutually commute, their $(2J+1)^2$ common eigenvectors $\{|J^2 J_z^{\rm LAB} J_z\rangle\}$ (which are solutions of a spherical top rotor) can be taken as the complete and orthonormal basis to solve Eq. (47) algebraically. Since both J^2 and $J_z^{\rm LAB}$ commute with $H^{\rm rot}$ whereas J_z generally does not, those states $\{|J^2 J_z^{\rm LAB} J_z\rangle\}$ with the same J^2 and $J_z^{\rm LAB}$ but different J_z will get mixed, with the mixing coefficients dependent on Q_k. The rotational energies $E^{\rm rot}_{J^2,\eta}(Q_k)$ therefore depend on Q_k. The Q_k-dependent solutions $|J^2 J_z^{\rm LAB}\eta\rangle$ are denoted simply as $|J_z^{\rm LAB}\rangle$ in Eq. (46).

The vibrational state $|i_{\rm vib}^{00}\rangle$ can be obtained by solving the nuclear Schrödinger-like equation

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial(Q_k)^2} + E_{\rm e}(Q_k) + V_J\right]|i_{\rm vib}^{00}\rangle = E_{\rm ve}^{00}|i_{\rm vib}^{00}\rangle, \tag{48}$$

where $V_J = E^{\rm rot}_{J^2,\eta}(Q_k) - E^{\rm rot}_{J^2,\eta}(Q_k^0)$ stands for the J-dependent rotational energy correction to the potential energy surface. This equation can be solved using a variety of methods; see, e.g., Ref. [27] and references therein.

The Diamagnetic Term

To calculate the diamagnetic term Eq. (42), the projection operator P^{00} Eq. (44) is first written explicitly as

$$P^{00} = \sum_{J_z^{\rm LAB}=-J}^{J}\sum_{m=1}^{n_{\rm spin}} |i_{\rm ele}^{00} i_{\rm vib}^{00} J^2 J_z^{\rm LAB}\eta\Phi_m\rangle\langle i_{\rm ele}^{00} i_{\rm vib}^{00} J^2 J_z^{\rm LAB}\eta\Phi_m|, \tag{49}$$

$$= |i_{\rm ele}^{00} i_{\rm vib}^{00}\rangle\langle i_{\rm ele}^{00} i_{\rm vib}^{00}| P^{\rm rot}(Q_k), \tag{50}$$

$$P^{\rm rot}(Q_k) = \sum_{J_z^{\rm LAB}=-J}^{J} |J^2 J_z^{\rm LAB}\eta\rangle\langle J^2 J_z^{\rm LAB}\eta|. \tag{51}$$

Use of the fact that $\{|\Phi_m\rangle\}$ constitutes a complete basis for the nuclear spin states has been made when going from Eqs. (49) to (50). It is now easily seen that the rotational projection operator $P^{\rm rot}(Q_k)$ depends on Q_k. As a consequence, the

effective Hamiltonian cannot be written as in Eq. (1), in which the $\vec{J}$ operator does not depend on Q_k. This can be avoided if a vibronic instead of a rovibronic state is chosen as the model function, i.e., neglecting H^{rot} in H^{00} Eq. (20). This conundrum can anyway be resolved for semirigid rotors by making the approximation

$$P^{\mathrm{rot}}(Q_k) \approx P^{\mathrm{rot}}(Q_k^0), \tag{52}$$

where Q_k^0 stands for the reference (equilibrium) geometry. Since $D_{uv}^{11,K}$ Eq. (33) is independent of the Euler angles, $H_{\mathrm{eff}}^{11,\mathrm{d}}$ can be calculated as

$$\begin{aligned} H_{\mathrm{eff}}^{11,\mathrm{d}} = {} & |i_{\mathrm{ele}}^{00} i_{\mathrm{vib}}^{00}\rangle\langle i_{\mathrm{ele}}^{00} i_{\mathrm{vib}}^{00}| D_{uv}^{11,K} |i_{\mathrm{ele}}^{00} i_{\mathrm{vib}}^{00}\rangle\langle i_{\mathrm{ele}}^{00} i_{\mathrm{vib}}^{00}| \\ & \times I_{Ku} p^{\mathrm{rot}}(Q_k^0) J_v p^{\mathrm{rot}}(Q_k^0). \end{aligned} \tag{53}$$

In a simplified notation, the $H_{\mathrm{eff}}^{11,\mathrm{d}}$ Hamiltonian can be rewritten as

$$H_{\mathrm{eff}}^{11,\mathrm{d}} = \langle i_{\mathrm{ele}}^{00} i_{\mathrm{vib}}^{00}| D_{uv}^{11,K} |i_{\mathrm{ele}}^{00} i_{\mathrm{vib}}^{00}\rangle I_{Ku} J_v, \tag{54}$$

provided that the domain of $H_{\mathrm{eff}}^{11,d}$ Eq. (54) is confined only to the model space V_{mod}.

The Paramagnetic Term

As for the paramagnetic term Eq. (43), an important simplification arises from the fact that $D_u^{01(K)}$ is a time-odd operator, such that the intermediate states in Q^{00} Eq. (45) cannot take i_{ele}^{00} as its electronic-state component. The effective Q^{00} operator is therefore given by

$$\begin{aligned} Q^{00} = {} & \sum_{a\neq i} \sum_{p} \sum_{J=0}^{\infty} \sum_{\eta} \sum_{J_z^{\mathrm{LAB}}=-J}^{J} \sum_{m=1}^{n_{\mathrm{spin}}} \\ & |a_{\mathrm{ele}}^{00} p_{\mathrm{vib}}^{00} J^2 \eta J_z^{\mathrm{LAB}} \Phi_m\rangle\langle a_{\mathrm{ele}}^{00} p_{\mathrm{vib}}^{00} J^2 \eta J_z^{\mathrm{LAB}} \Phi_m| \end{aligned} \tag{55}$$

$$= \sum_{a\neq i} |a_{\mathrm{ele}}^{00}\rangle\langle a_{\mathrm{ele}}^{00}|. \tag{56}$$

By approximating the energy gap of two rovibronic states by the energy gap of their electronic states at the equilibrium geometry, the resolvent in Eq. (43) can be written as

$$\begin{aligned} & Q^{00} \frac{1}{E^{00} - H^{00}} Q^{00} \\ & = \sum_{a\neq i} |a_{\mathrm{ele}}^{00}\rangle\langle a_{\mathrm{ele}}^{00}| \frac{1}{E_{\mathrm{e}}(i_{\mathrm{ele}}^{00}, Q_k^0) - E_{\mathrm{e}}(a_{\mathrm{ele}}^{00}, Q_k^0)} |a_{\mathrm{ele}}^{00}\rangle\langle a_{\mathrm{ele}}^{00}|. \end{aligned} \tag{57}$$

Inserting Eqs. (50) and (57) into Eq. (43) leads to

$$H_{\text{eff}}^{11,\text{p}} = P^{\text{rot}}(Q_k^0)|i_{\text{ele}}^{00} i_{\text{vib}}^{00}\rangle \Big[\sum_{a\neq i} \frac{\langle i_{\text{ele}}^{00} i_{\text{vib}}^{00}| D_v^{10} J_v |a_{\text{ele}}^{00}\rangle\langle a_{\text{ele}}^{00}| D_u^{01(K)} I_{Ku} |i_{\text{ele}}^{00} i_{\text{vib}}^{00}\rangle}{E_{\text{e}}(i_{\text{ele}}^{00}, Q_k^0) - E_{\text{e}}(a_{\text{ele}}^{00}, Q_k^0)} + h.c. \Big] \langle i_{\text{ele}}^{00} i_{\text{vib}}^{00}| P^{\text{rot}}(Q_k^0), \tag{58}$$

which can also be rewritten as

$$H_{\text{eff}}^{11,\text{p}} = \sum_{a\neq i} \frac{\langle i_{\text{ele}}^{00} i_{\text{vib}}^{00}| D_v^{10} |a_{\text{ele}}^{00}\rangle\langle a_{\text{ele}}^{00}| D_u^{01(K)} |i_{\text{ele}}^{00} i_{\text{vib}}^{00}\rangle}{E_{\text{e}}(i_{\text{ele}}^{00}, Q_k^0) - E_{\text{e}}(a_{\text{ele}}^{00}, Q_k^0)} I_{Ku} J_v + h.c. \tag{59}$$

by confining the domain of $H_{\text{eff}}^{11,\text{p}}$ Eq. (59) in the model space V_{mod}. For practical use, the Hamiltonian $H_{\text{eff}}^{11,\text{p}}$ Eq. (59) can be further approximated as

$$H_{\text{eff}}^{11,\text{p}} = \langle i_{\text{vib}}^{00}| \sum_{a\neq i} \frac{\langle i_{\text{ele}}^{00}| D_v^{10} |a_{\text{ele}}^{00}\rangle\langle a_{\text{ele}}^{00}| D_u^{01(K)} |i_{\text{ele}}^{00}\rangle}{E_e(i_{\text{ele}}^{00}, Q_k) - E_e(a_{\text{ele}}^{00}, Q_k)} |i_{\text{vib}}^{00}\rangle I_{Ku} J_v + h.c, \tag{60}$$

where the integral is nothing but the vibrational average of a second-order electronic property calculated at an arbitrary geometry Q_K.

The NSR Tensor

The desired effective Hamiltonian H_{eff}^{11} Eq. (41) is the sum of $H_{\text{eff}}^{11,\text{d}}$ Eq. (54) and $H_{\text{eff}}^{11,\text{p}}$ Eq. (60). By comparing H_{eff}^{11} Eq. (41) with Eq. (1), the NSR tensor M_{uv}^K can be calculated as

$$M_{uv}^K = \langle i_{\text{vib}}^{00}(Q_k)| M_{uv}^K(Q_k) |i_{\text{vib}}^{00}(Q_k)\rangle. \tag{61}$$

In view of Eqs. (23), (30), and (33), $M_{uv}^K(Q_k)$ can be decomposed as

$$M_{uv}^K(Q_k) = M_{uv}^{K,\text{d}}(Q_k) + M_{uv}^{K,\text{p}}(Q_k), \tag{62}$$

$$M_{uv}^{K,\text{d}}(Q_k) = M_{uv}^{K,\text{dn}}(Q_k) + M_{uv}^{K,\text{de}}(Q_k), \tag{63}$$

$$M_{uv}^{K,\text{dn}}(Q_k) = D_{uv}^{11,\text{n},K}, \quad M_{uv}^{K,\text{de}}(Q_k) = \langle i_{\text{ele}}^{00}| D_{uv}^{11,\text{e},K} |i_{\text{ele}}^{00}\rangle, \tag{64}$$

$$M_{uv}^{K,\text{p}}(Q_k) = \frac{\langle i_{\text{ele}}^{00}| D_u^{01(K)} |a_{\text{ele}}^{00}\rangle\langle a_{\text{ele}}^{00}| D_v^{10} |i_{\text{ele}}^{00}\rangle}{E_e(i_{\text{ele}}^{00}, Q_k) - E_e(a_{\text{ele}}^{00}, Q_k)} + c.c., \tag{65}$$

where $M_{uv}^{K,\text{dn}}(Q_k)$ is the nuclear term (which is simply a constant for a given geometry), while $M_{uv}^{K,\text{de}}(Q_k)$ and $M_{uv}^{K,\text{p}}(Q_k)$ are the electronic diamagnetic and paramagnetic terms, respectively.

The Perturbation-Dependent Basis Set

The electronic paramagnetic term $M_{uv}^{K,\mathrm{p}}(Q_k)$ Eq. (65) converges very slowly with respect to the one-electron basis sets. This can be improved greatly by introducing proper relativistic rotational London orbitals (RLO) [24] as a perturbation-dependent basis set. Before doing so in section "Relativistic Rotational London Orbitals," the introduction of RLO in the nonrelativistic framework for the NSR tensor is first recapitulated briefly in section "Rotational London Orbitals."

Rotational London Orbitals

The introduction of the RLO [25],

$$\chi_\mu = e^{i\left(\vec{R}_{\mu C}\times\vec{r}\right)\cdot B_I \vec{J}}\chi_\mu^{00}, \tag{66}$$

was inspired by the improved basis set convergence of London orbitals (LO) [8,28–30] for the calculation of nuclear shielding tensors and the similarity between the operators in Eqs. (2) and (3). It is the latter that dictates that the complex phase factor multiplying the conventional atomic orbital (AO) χ_μ^{00} is simply proportional to that appearing in the LO

$$\chi_\mu = e^{-\frac{i}{2}\left(\vec{R}_{\mu g}\times\vec{r}\right)\cdot\vec{B}_{\mathrm{ext}}}\chi_\mu^{00}. \tag{67}$$

For nuclear shielding tensors, the effect of the phase factor in χ_μ Eq. (67) is to move the single global gauge origin $\vec{R}_g$ to more suited *local* gauge origins for each AO. These local gauge origins are the atomic centers to which the AOs are attached. The reason for this being an (close to) optimal choice can be realized from the fact that, in the case of a one-electron, one-center system in the presence of an external magnetic field, the LO Eq. (67) is correct to first order in the external magnetic field, whereas the conventional AO is only correct to zeroth order [31]. In the presence of an external magnetic field, the LO Eq. (67) are thus better basis functions than the conventional AO. Although there is no arbitrariness in the choice of origin in the case of NSR tensors, being well defined to be the NCM of the molecule, the improved basis set convergence is an argument for introducing the RLO Eq. (66) in calculations of NSR tensors. Another important reason lies in the fact that the mapping relation Eq. (8), together with $\sigma_{vu,\mathrm{eq}}^{K,\mathrm{d}}(\mathrm{NR})$ defined in Eq. (7), also holds when the RLO Eq. (66) are used for the NSR tensor and the LO Eq. (67) for the nuclear shielding tensor. It will be shown in section "Relativistic Rotational London Orbitals" that these two advantages can also be realized in the relativistic domain by introducing properly the relativistic RLO [24].

Relativistic Rotational London Orbitals

The underlying reason for the particular forms of the LO Eq. (67) and RLO Eq. (66) lies in that both $d_v^{10}(\mathrm{NR},\mathrm{NMR})$ Eq. (3) and $d_v^{10}(\mathrm{NR},\mathrm{NSR})$ Eq. (2) can be decomposed into a (reduced) paramagnetic term and a diamagnetic term. To see this, the $d_v^{10}(\mathrm{NR},\mathrm{NSR})$ Eq. (2) and $d_v^{10}(\mathrm{NR},\mathrm{NMR})$ Eq. (3) operators are first written in a unified form

$$d_v^{10}(\mathrm{NR},\mathrm{X}) = f\big[(\vec{r}-\vec{R}_O)\times\vec{p}\big]_v, \tag{68}$$

where $\mathrm{X}=\mathrm{NSR}$ with $f=-B_I$ and $\vec{R}_O=\vec{R}_\mathrm{C}$ or $\mathrm{X}=\mathrm{NMR}$ with $f=1/2$ and $\vec{R}_O=\vec{R}_\mathrm{g}$. Eq. (68) can obviously be rewritten as

$$d_v^{10}(\mathrm{NR},\mathrm{X}) = \tilde{d}_v^{10}(\mathrm{NR},\mathrm{X}) + \bar{d}_v^{10}(\mathrm{NR},\mathrm{X}), \tag{69}$$

where

$$\tilde{d}_v^{10}(\mathrm{NR},\mathrm{X}) = f\big[(\vec{r}-\vec{R}_\mu)\times\vec{p}\big]_v, \tag{70}$$

$$\bar{d}_v^{10}(\mathrm{NR},\mathrm{X}) = f\big[\vec{R}_{\mu O}\times\vec{p}\big]_v, \quad \vec{R}_{\mu O}=\vec{R}_\mu-\vec{R}_O. \tag{71}$$

Further in view of the operator identity

$$\left[\frac{1}{2}p^2,\vec{r}\right] = -i\,\vec{p}, \tag{72}$$

$\bar{d}_v^{10}(\mathrm{NR},\mathrm{X})$ Eq. (71) can be replaced by

$$\bar{d}_v^{10}(\mathrm{NR},\mathrm{X}) = if\big[h^{00}(\mathrm{NR}),(\vec{R}_{\mu O}\times\vec{r})_v\big]. \tag{73}$$

In the revised form Eq. (69) of $d_v^{10}(\mathrm{NR},\mathrm{X})$, $\tilde{d}_v^{10}(\mathrm{NR},\mathrm{X})$ Eq. (70) still contributes as a paramagnetic term but with the (gauge) origin placed at the center $\vec{R}_\mu$ of the AO. In contrast, $\bar{d}_v^{10}(\mathrm{NR},\mathrm{X})$ Eq. (73) will contribute as a diamagnetic term, because the commutator therein will cancel the energy denominator in the corresponding response function $\psi_{i,b}^{10,v}$, so as to make the latter a fixed instead of a response function, i.e., $\psi_{i,b}^{10,v}=\chi_\mu^{10,v}C_{\mu i}^{00}$ with $\chi_\mu^{10,v}=-if\,(\vec{R}_{\mu O}\times\vec{r})_v\chi_\mu^{00}$ [24].

The above procedure can also be applied to the relativistic $d_v^{10}(\mathrm{NMR})$ Eq. (11) operator by rewriting it as

$$d_v^{10}(\mathrm{NMR}) = \tilde{d}_v^{10}(\mathrm{NMR}) + \bar{d}_v^{10}(\mathrm{NMR}), \tag{74}$$

$$\tilde{d}_v^{10}(\mathrm{NMR}) = \frac{1}{2}\big[(\vec{r}-\vec{R}_\mu)\times c\vec{\alpha}\big]_v, \tag{75}$$

$$\bar{d}_v^{10}(\mathrm{NMR}) = \frac{1}{2}(\vec{R}_{\mu g}\times c\vec{\alpha})_v, \quad \vec{R}_{\mu g}=\vec{R}_\mu-\vec{R}_g. \tag{76}$$

In place of the identity Eq. (72), the following identity

$$[c\vec{\alpha}\cdot\vec{p},\vec{r}] = -ic\vec{\alpha} \tag{77}$$

allows one to replace $\bar{d}_v^{10}$(NMR) Eq. (76) by

$$\bar{d}_v^{10}(\text{NMR}) = \frac{i}{2}\big[h^{00},(\vec{R}_{\mu g}\times\vec{r})_v\big], \tag{78}$$

where h^{00} is the one-electron Dirac operator. The similarity between $\bar{d}_v^{10}$(NR, NMR) Eq. (73) and $\bar{d}_v^{10}$(NMR) Eq. (78) implies that the nonrelativistic LO [8, 28–30] for nuclear shielding tensors can directly be extended to the relativistic domain [32, 33].

However, the above procedure cannot be applied to d_v^{10}(NSR) Eq. (31) because no operator O can satisfy the desired identity

$$[h^{00},O] = -i(\vec{R}_{\mu C}\times\vec{p})_v, \quad \vec{R}_{\mu C} = \vec{R}_\mu - \vec{R}_C. \tag{79}$$

This is of course not surprising, as the dependence of the NSR tensor on the NCM $\vec{R}_C$ is a physical consequence. Therefore, the introduction of "relativistic RLO" $\{\chi_\mu(\text{NSR})\}$ can only be done in a special way.

Following the principles of "orbital decomposition" [34], what is really wanted is the following form:

$$\psi_i^{10,v}(\text{NSR}) = \psi_{i,b}^{10,v}(\text{NSR}) + \psi_{i,r}^{10,v}(\text{NSR}) \tag{80}$$

for the first-order $\psi_i^{10,v}$(NSR) of the bi-spinor ψ_i(NSR), which is featured by a known term $\psi_{i,b}^{10,v}$(NSR) and a residual $\psi_{i,r}^{10,v}$(NSR). That the nonrelativistic limit (nrl) of the large component $\psi_{i,b}^{L,10,v}$(NSR) of $\psi_{i,b}^{10,v}$(NSR) should take the same form as the first order of the nonrelativistic RLO Eq. (66) gives rise to the following condition:

$$\psi_{i,b}^{L,10,v}(\text{NSR}) = Z_{g,v}^{10}(\text{NSR})\left(\psi_i^{L,00} + \mathcal{O}(c^{-2})\right), \tag{81}$$

$$\vec{Z}_g^{10}(\text{NSR}) = iB_I\vec{R}_{\mu C}\times\vec{r}. \tag{82}$$

To see what constraint can be imposed on the small component of $\psi_{i,b}^{10,v}$(NSR) suffices it to consider the one-electron Dirac equation

$$\begin{pmatrix} V_{\text{eff}} & c\vec{\sigma}\cdot(\vec{p}+\vec{A}_N) \\ c\vec{\sigma}\cdot(\vec{p}+\vec{A}_N) & V_{\text{eff}}-2c^2 \end{pmatrix}\begin{pmatrix}\psi_i^L(\text{NSR})\\ \psi_i^S(\text{NSR})\end{pmatrix} = \epsilon_i\begin{pmatrix}\psi_i^L(\text{NSR})\\ \psi_i^S(\text{NSR})\end{pmatrix}, \tag{83}$$

where

$$V_{\text{eff}} = V_N - \left(\vec{l}_{\text{C}} + \frac{1}{2}\vec{\sigma}\right) \cdot B_I \vec{J}. \tag{84}$$

The second line of Eq. (83) gives rise to the following relation between the small and large components of a positive-energy spinor ψ_i:

$$(2c^2 + \epsilon_i - V_{\text{eff}})\psi_i^S(\text{NSR}) = c\vec{\sigma} \cdot (\vec{p} + \vec{A}_N)\psi_i^L(\text{NSR}), \tag{85}$$

which can be expanded first in c^{-1}

$$\psi_i^S(\text{NSR}) = \frac{\vec{\sigma} \cdot (\vec{p} + \vec{A}_N)}{2c} \left(\psi_i^L(\text{NSR}) + \mathcal{O}(c^{-2})\right), \tag{86}$$

and then in $\vec{J}_v$

$$\psi_i^{S,00} = \frac{\vec{\sigma} \cdot \vec{p}}{2c} \left(\psi_i^{L,00} + \mathcal{O}(c^{-2})\right), \tag{87}$$

$$\psi_i^{S,10,v}(\text{NSR}) = \frac{\vec{\sigma} \cdot \vec{p}}{2c} \left(\psi_i^{L,10,v}(\text{NSR}) + \mathcal{O}(c^{-2})\right). \tag{88}$$

Suppose now that the residual $\psi_{i,r}^{10,v}(\text{NSR})$ of $\psi_i^{10,v}(\text{NSR})$ Eq. (80) can effectively be expanded in the basis of zeroth-order orbitals $\{\psi_p^{00}\}$, viz.,

$$\psi_{i,r}^{10,v}(\text{NSR}) = \begin{pmatrix} \psi_{i,r}^{L,10,v} \\ \psi_{i,r}^{S,10,v} \end{pmatrix} = \begin{pmatrix} \psi_p^{L,00} \\ \psi_p^{S,00} \end{pmatrix} U_{pi}^{10}, \tag{89}$$

where the summation over p includes both positive (p^+) and negative (p^-) energy states. However, it can be required that the contribution of the negative energy states vanishes in the nrl, which can only be satisfied if

$$\psi_{i,r}^{10,v}(\text{NSR}) = \begin{pmatrix} \psi_{i,r}^{L,10,v} \\ \psi_{i,r}^{S,10,v} \end{pmatrix} = \begin{pmatrix} \psi_{p^+}^{L,00} \\ \psi_{p^+}^{S,00} \end{pmatrix} U_{p^+i}^{10} + O(c^{-2}). \tag{90}$$

Applying the condition Eq. (87) to $\psi_{p^+}^{S,00}$ leads to that the small component of $\psi_{i,r}^{10,v}(\text{NSR})$ Eq. (90) should have the following structure:

$$\psi_{i,r}^{S,10,v}(\text{NSR}) = \frac{\vec{\sigma} \cdot \vec{p}}{2c} \left(\psi_{p^+}^{L,00} + \mathcal{O}(c^{-2})\right) U_{p^+i}^{10} \tag{91}$$

$$= \frac{\vec{\sigma} \cdot \vec{p}}{2c} \left(\psi_{i,r}^{L,10,v}(\text{NSR}) + \mathcal{O}(c^{-2})\right) \tag{92}$$

in order to guarantee the correct nrl. The difference between $\psi_i^{S,10,v}(\mathrm{NSR})$ Eq. (88) and $\psi_{i,r}^{S,10,v}(\mathrm{NSR})$ Eq. (92) then gives rise to the desired condition for determining the small component of $\psi_{i,b}^{10,v}$:

$$\psi_{i,b}^{S,10,v}(\mathrm{NSR}) = \frac{\vec{\sigma}\cdot\vec{p}}{2c}\left(\psi_{i,b}^{L,10,v}(\mathrm{NSR}) + \mathcal{O}(c^{-2})\right) \tag{93}$$

$$= \frac{\vec{\sigma}\cdot\vec{p}}{2c}\left(Z_{g,v}^{10}(\mathrm{NSR})\psi_i^{L,00} + \mathcal{O}(c^{-2})\right) \tag{94}$$

$$= \left(Z_{g,v}^{10}(\mathrm{NSR})\psi_i^{S,00} + \mathcal{O}(c^{-2})\right) + \tilde{Z}_{m,v}^{10}(\mathrm{NSR})\left(\psi_i^{L,00} + \mathcal{O}(c^{-2})\right), \tag{95}$$

where

$$\tilde{Z}_{m,v}^{10}(\mathrm{NSR}) = \left[\frac{\vec{\sigma}\cdot\vec{p}}{2c}, Z_{g,v}^{10}(\mathrm{NSR})\right] = \frac{1}{2c}B_I(\vec{R}_{\mu C}\times\sigma)_v. \tag{96}$$

Equations (81) and (95) are the necessary and sufficient conditions of the RLO for four-component relativistic NSR calculations with the correct nrl. Given the weakness and diagonal form of the perturbation J_v, the $\mathcal{O}(c^{-2})$ terms therein can be neglected, thereby leading to

$$\psi_{i,b}^{10,v}(\mathrm{NSR}) = \begin{pmatrix}\psi_{i,b}^{L,10,v}(\mathrm{NSR})\\ \psi_{i,b}^{S,10,v}(\mathrm{NSR})\end{pmatrix} = \begin{pmatrix}Z_{g,v}^{10}(\mathrm{NSR}) & 0\\ \tilde{Z}_{m,v}^{10}(\mathrm{NSR}) & Z_{g,v}^{10}(\mathrm{NSR})\end{pmatrix}\begin{pmatrix}\psi_i^{L,00}\\ \psi_i^{S,00}\end{pmatrix}. \tag{97}$$

The first-order relativistic RLO can be extracted as

$$\chi_\mu^{10,v}(\mathrm{NSR}) = \begin{pmatrix}\chi_\mu^{L,10,v}(\mathrm{NSR})\\ \chi_\mu^{S,10,v}(\mathrm{NSR})\end{pmatrix} = \begin{pmatrix}Z_{g,v}^{10}(\mathrm{NSR}) & 0\\ \tilde{Z}_{m,v}^{10}(\mathrm{NSR}) & Z_{g,v}^{10}(\mathrm{NSR})\end{pmatrix}\chi_\mu^{00}, \tag{98}$$

because the zeroth-order functions $\{\chi_\mu^{00}\}$ span the same space as $\{\psi_p^{00}\}$. Given the non-determination of the $\mathcal{O}(c^{-2})$ terms, it is perfectly legitimate to antisymmetrize the operator in Eq. (98), viz.,

$$\chi_\mu^{10,v}(\mathrm{NSR}) = \begin{pmatrix}\chi_\mu^{L,10,v}(\mathrm{NSR})\\ \chi_\mu^{S,10,v}(\mathrm{NSR})\end{pmatrix} = \begin{pmatrix}Z_{g,v}^{10}(\mathrm{NSR}) & -\tilde{Z}_{m,v}^{10}(\mathrm{NSR})\\ \tilde{Z}_{m,v}^{10}(\mathrm{NSR}) & Z_{g,v}^{10}(\mathrm{NSR})\end{pmatrix}\chi_\mu^{00} \triangleq Z_v^{10}(\mathrm{NSR})\chi_\mu^{00}, \tag{99}$$

where

$$\vec{Z}^{10}(\mathrm{NSR}) = \vec{Z}_g^{10}(\mathrm{NSR}) + \vec{Z}_m^{10}(\mathrm{NSR}) = -\vec{Z}^{10\dagger}(\mathrm{NSR}), \tag{100}$$

$$\vec{Z}_m^{10}(\mathrm{NSR}) = -\frac{B_I}{2c}\beta \vec{R}_{\mu C} \times \vec{\alpha}. \tag{101}$$

Apart from elegance in the operator, a clear advantage of Eq. (99) over Eq. (98) lies in that the one-center, first-order overlap matrix elements $S_{\mu v}^{10,v} = \langle \chi_\mu^{00} | Z_v^{10}(\mathrm{NSR}) + h.c. | \chi_v^{00} \rangle$ vanish for the former.

Had the $\tilde{\vec{Z}}_m^{10}(\mathrm{NSR})/\vec{Z}_m^{10}(\mathrm{NSR})$ term been neglected, Eqs. (98) and (99) would be reduced to

$$\chi_\mu^{10,v}(\mathrm{NSR}) = Z_{g,v}^{10}(\mathrm{NSR})\chi_\mu^{00}, \tag{102}$$

which amounts to a naive extension of the nonrelativistic RLO [25] to the four-component case. In this case, the nrl cannot be guaranteed, in contrast with the Ansätze Eqs. (98) and (99). It is the $\tilde{\vec{Z}}_m^{10}(\mathrm{NSR})/\vec{Z}_m^{10}(\mathrm{NSR})$ term that restores the kinetic balance missed by $\vec{Z}_g^{10}(\mathrm{NSR})$. For this reason, the Ansätze in Eqs. (98)/(99) and Eq. (102) can be denoted as "kinetically balanced RLO" (KB-RLO) and "kinetically unbalanced RLO" (KU-RLO), respectively.

The above Ansätze fits all kinds of kinetic balances [35] for the zeroth-order functions $\{\chi_\mu^{00}\}$. If the restricted kinetic balance (RKB) [36] is considered, the following Ansatz can also be used:

$$\chi_\mu^{L,10,v}(\mathrm{NSR}) = Z_{g,v}^{10}(\mathrm{NSR})g_\mu,\ \chi_\mu^{S,10,v}(\mathrm{NSR}) = \frac{\vec{\sigma}\cdot\vec{p}}{2c}Z_{g,v}^{10}(\mathrm{NSR})g_\mu. \tag{103}$$

That is, the nonrelativistic RLO Eq. (66) are used for the large-component basis, from which the small-component basis is generated according to the RKB prescription. This Ansatz also belongs to the KB-RLO category.

Some remarks are in order.

1. Unlike the zeroth-order (in J_v) relation Eq. (87) where the $\mathcal{O}(c^{-2})$ terms cannot be neglected for a finite speed of light, neglecting the $\mathcal{O}(c^{-2})$ terms in Eqs. (81) and (95), thereby leading to Eq. (97) for $\psi_{i,b}^{10,v}(\mathrm{NSR})$, is *not* an approximation. Even setting $\psi_{i,b}^{10,v}(\mathrm{NSR})$ to zero is not an approximation. A different choice of $\psi_{i,b}^{10,v}(\mathrm{NSR})$ means simply a different decomposition of $\psi_i^{10,v}(\mathrm{NSR})$ and hence a different definition of the diamagnetic and paramagnetic terms. In this regard, all the Ansätze in Eqs. (98), (99), (102), and (103) are equivalent. However, the Ansätze Eqs. (98), (99), and (103) ensure that the contribution of the negative-energy states to the paramagnetic term vanishes in the nrl and are hence much preferred.

2. In terms of the field-dependent basis, only the electronic paramagnetic term $M_{uv}^{K,\mathrm{p}}(Q_k)$ Eq. (65) of $M_{uv}^{K}(Q_k)$ Eq. (62) needs to be revised; see Ref. [24] for explicit expressions.

As a final note, it is interesting to see that the Ansatz Eq. (99) for the NSR tensor is closely parallel to the EFUT (external field-dependent unitary transformation) formulation [37, 38] of the nuclear shielding tensor, where the first-order magnetically balanced LO $\chi_{\mu}^{10,v}(\mathrm{NMR})$ reads [32]

$$\chi_{\mu}^{10,v}(\mathrm{NMR}) = Z_{v}^{10}(\mathrm{NMR})\chi_{\mu}^{00} = \left(Z_{g,v}^{10}(\mathrm{NMR}) + Z_{m,v}^{10}(\mathrm{NMR})\right)\chi_{\mu}^{00}, \tag{104}$$

$$Z_{g,v}^{10}(\mathrm{NMR}) = -\frac{i}{2}\left(\vec{R}_{\mu g}\times\vec{r}\right)_{v}, \quad \vec{R}_{\mu g} = \vec{R}_{\mu} - \vec{R}_{g}, \tag{105}$$

$$Z_{m,v}^{10}(\mathrm{NMR}) = -\frac{1}{4c}\beta\left(\vec{r}_{\mu}\times\vec{\alpha}\right)_{v}, \quad \vec{r}_{\mu} = r - \vec{R}_{\mu}. \tag{106}$$

Equation (104) obviously matches the form of Eq. (99).

Relativistic Mapping Between NMR and NSR Tensors

The interest in the NSR tensor resides not only in the understanding of hyperfine structures of molecular rotational spectra but also in the determination of absolute nuclear magnetic shielding tensors [3, 11, 12, 15, 22]. It is well known that, although the relative shielding between two nuclei, the chemical shift, can be measured accurately in NMR experiments, the absolute nuclear magnetic shielding cannot [8]. In most cases, the reported "experimental" values of isotropic nuclear magnetic shielding constants were mapped via Eq. (9) from the experimental NSR coupling constants. As emphasized before, the relation Eq. (8) and hence Eq. (9) are nonrelativistic in nature. In reality, the relation Eq. (8) is actually "inversely relativistic" [23], i.e., nonrelativistic for heavy atoms but relativistic for (very) light atoms. This peculiarity arises from the fact that the so-mapped $\sigma_{vu,\mathrm{eq}}^{K}(\mathrm{exp,NR})$ for light atoms inherits some relativistic effects from the experimental $M_{uv,\mathrm{eq}}^{K}(\mathrm{exp})$, whereas relativistic corrections to the $M_{uv,\mathrm{eq}}^{K}$ of heavy atoms are generally very small (*vide post*).

In the relativistic framework, the NMR and NSR tensors are distinct properties. Yet, a formal relation between them can still be established by starting with the apparent identity

$$\sigma_{vu,\mathrm{eq}}^{K,\mathrm{p}} = -\frac{I_{vv}^{0}}{2g_K\mu_{\mathrm{n}}}M_{uv,\mathrm{eq}}^{K,\mathrm{p}} + \Delta_{vu,\mathrm{eq}}^{K,\mathrm{p}}, \tag{107}$$

$$\Delta_{vu,\mathrm{eq}}^{K,\mathrm{p}} = \sigma_{vu,\mathrm{eq}}^{K,\mathrm{p}} + \frac{I_{vv}^{0}}{2g_K\mu_{\mathrm{n}}}M_{uv,\mathrm{eq}}^{K,\mathrm{p}}, \tag{108}$$

where the gauge origin of the external vector potential is supposed to be at the NCM. Just like the step going from Eqs. (5) to (6), the theoretical paramagnetic term $M^{K,\mathrm{p}}_{uv,\mathrm{eq}}$ in the first term of Eq. (107) can be replaced with the experimental counterpart, $M^{K,\mathrm{p}}_{uv,\mathrm{eq}}(\mathrm{exp}) = M^{K}_{uv,\mathrm{eq}}(\mathrm{exp}) - M^{K,\mathrm{d}}_{uv,\mathrm{eq}}$, so as to obtain

$$\sigma^{K,\mathrm{p}}_{vu,\mathrm{eq}}(\mathrm{exp}) = -\frac{I^0_{vv}}{2g_K\mu_\mathrm{n}}\left[M^{K}_{uv,\mathrm{eq}}(\mathrm{exp}) - M^{K,\mathrm{d}}_{uv,\mathrm{eq}}\right] + \Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}}. \tag{109}$$

The desired "relativistic mapping" between NMR and NSR tensors then reads

$$\begin{aligned}\sigma^{K}_{vu,\mathrm{eq}}(\mathrm{exp}) &= \sigma^{K,\mathrm{p}}_{vu,\mathrm{eq}}(\mathrm{exp}) + \sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}} \\ &= \sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}} - \frac{I^0_{vv}}{2g_K\mu_\mathrm{n}}\left[M^{K}_{uv,\mathrm{eq}}(\mathrm{exp}) - M^{K,\mathrm{d}}_{uv,\mathrm{eq}}\right] + \Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}},\end{aligned} \tag{110}$$

where $\sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}}$ is again the calculated diamagnetic term of the shielding. Any four-component relativistic formulation (for a recent review, see Ref. [33]) of NMR shielding can be employed here, provided that it has an explicit diamagnetic term. However, only the EFUT approach [37] allows for a direct evaluation [20, 21] of $\Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}}$ defined in Eq. (108). Specifically, the EFUT paramagnetic term $\sigma^{K,\mathrm{p}}_{vu}(Q_k)$ reads

$$\sigma^{K,\mathrm{p}}_{vu}(Q_k) = \frac{\langle i^{00}_{\mathrm{ele}}|D^{01(K)}_u(\mathrm{NMR})|a^{00}_{\mathrm{ele}}\rangle\langle a^{00}_{\mathrm{ele}}|D^{10}_v(\mathrm{NMR})|i^{00}_{\mathrm{ele}}\rangle}{E_e(i^{00}_{\mathrm{ele}}, Q_k) - E_e(a^{00}_{\mathrm{ele}}, Q_k)} + c.c., \tag{111}$$

where

$$D^{01(K)}_u(\mathrm{NMR}) = \sum_i \frac{(\vec{r}_{iK}\times c\vec{\alpha}_i)_u}{c^2 r^3_{iK}}, \tag{112}$$

$$D^{10}_v(\mathrm{NMR}) = \sum_i \frac{\beta}{2}\left(\left[(\vec{r}_i - \vec{R}_g)\times\vec{p}_i\right]_v + \Sigma_{iv}\right). \tag{113}$$

By comparing $D^{01(K)}_u(\mathrm{NMR})$ with $D^{01(K)}_u$ Eq. (23) and $D^{10}_v(\mathrm{NMR})$ with D^{10}_v Eq. (30), the difference $\Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}}$ Eq. (108) between $\sigma^{K,\mathrm{p}}_{vu}(Q_k)$ Eq. (111) and $-\frac{I^0_{vv}}{2g_K\mu_\mathrm{n}}M^{K,\mathrm{p}}_{uv}(Q_k)$ Eq. (65) can be obtained as

$$\Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}} = \frac{\langle i^{00}_{\mathrm{ele}}|D^{01(K)}_u|a^{00}_{\mathrm{ele}}\rangle\langle a^{00}_{\mathrm{ele}}|\Delta D^{10}_v|i^{00}_{\mathrm{ele}}\rangle}{E_e(i^{00}_{\mathrm{ele}}, Q^0_k) - E_e(a^{00}_{\mathrm{ele}}, Q^0_k)} + \mathrm{c.c.}, \tag{114}$$

$$\Delta D^{10}_v = D^{10}_v(\mathrm{NMR})|_{\vec{R}_g=\vec{R}_C} - \left(-\frac{1}{2}I^{00}D^{10}_v\right) \tag{115}$$

$$= \frac{1}{2}\sum_i\left[(\beta_i - 1)l_{i,v} + \left(\beta_i - \frac{1}{2}\right)\Sigma_{i,v}\right]. \tag{116}$$

To complete Eq. (110), the EFUT $\sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}}$ (with the gauge origin at $\vec{R}_C$) should be specified [37]

$$\sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}} = \langle\Psi^{00}|\beta \sum_i \frac{\vec{r}_{iC}\cdot\vec{r}_{iK}\delta_{uv} - r_{iC,u}r_{iK,v}}{r^3_{iK}}|\Psi^{00}\rangle. \tag{117}$$

Of course, Eq. (110) is only meaningful if $\Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}}$ Eq. (114) is insensitive to both one- and many-particle basis sets. This is indeed the case [23], simply because ΔD^{10}_v Eq. (116) is a differential one-electron operator. In other words, while both $\sigma^{K,\mathrm{p}}_{vu,\mathrm{eq}}$ and $M^{K,\mathrm{p}}_{uv,\mathrm{eq}}$ are very sensitive to basis sets, their difference $\Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}}$ is not (due to systematic cancellation [23]). Close inspections reveal [23] that $\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp})$ Eq. (110) has three relativistic corrections to $\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp,NR})$ Eq. (8):

- $\Delta_A = \Delta^{K,\mathrm{p}}_{vu,\mathrm{eq}}$, relativistic effect in the paramagnetic mapping
- $\Delta_B = \sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}} - \sigma^{K,\mathrm{d}}_{vu,\mathrm{eq}}(\mathrm{NR})$, relativistic effect in the NMR diamagnetism
- $\Delta_C = \frac{I^0_{vv}}{2g_K\mu_\mathrm{n}}\left[M^{K,\mathrm{d}}_{uv,\mathrm{eq}} - M^{K,\mathrm{d}}_{uv,\mathrm{eq}}(\mathrm{NR})\right]$, relativistic effect in the NSR diamagnetism

$\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp})$ Eq. (110) and $\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp,NR})$ Eq. (8) are hence related by

$$\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp}) = \sigma^K_{vu,\mathrm{eq}}(\mathrm{exp,NR}) + \Delta_A + \Delta_B + \Delta_C, \tag{118}$$

which shows that if the "nonrelativisitic mapping" Eq. (8) or Eq. (9) is to be used for deriving the nuclear shielding tensor $\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp,NR})$ from the experimental NSR tensor $M^{K,\mathrm{d}}_{uv,\mathrm{eq}}(\mathrm{exp})$, the difference between the relativistic effects in the calculated nuclear shielding and NSR tensors must further be added in.

For notional reasons, the combination of Eq. (110) with Eq. (114) and that of Eq. (110) with Eq. (108) have been dubbed [23] as "direct relativistic mapping" and "indirect relativistic mapping," respectively. It is the latter that was used in Ref. [22]. If the absolute shielding $\sigma^K_{vu,\mathrm{eq}}(\mathrm{exp})$ is the sole target, the former is more advantageous, for then only one set of coupled-perturbed equations needs to be solved for the responses of the electronic wave function to the differential perturbation ΔD^{10}_v Eq. (116), which converges very fast. If both the shielding and NSR tensors are wanted, the "indirect relativistic mapping" is instead recommended even for EFUT, for which Eqs. (108) and (114) yield identical results as long as the coupled-perturbed equations for the shielding and NSR tensors are both sufficiently convergent.

Without going into details, Eq. (110) along with Eqs. (114) and (117) holds [24] also when the relativistic RLO Eq. (99) are used for the NSR tensor and the relativistic LO Eq. (104) for the nuclear shielding tensor. This is another indicator that the "direct relativistic mapping" is indeed very robust.

Illustrative Examples

NSR Constants of Hydrogen Halide

The hydrogen halide series HX (X = F, Cl, Br, I) has been taken [19, 23, 24, 26, 39] as a prototypical test set for the relativistic theory of NSR tensors. Consider first the basis set convergence. For this purpose, the performance of the RLO discussed in section "The Perturbation-Dependent Basis Set" is compared in Fig. 1 with that of the conventional basis sets for the NSR constants of HBr and HI. It is seen that, for the hydrogen NSR constants, the RLO do not show any significant improvement over the conventional basis sets, neither at the KS nor at the DKS level. However, the situation changes for the halogen NSR constants. Both nonrelativistic and relativistic RLO improve the conventional calculations significantly, at very much the same rate. Consequently, when combined with the RLO, triple-zeta basis sets are already sufficient for accurate predictions (within 1 % of the KS/DKS limit) of NSR constants. It is also clear that the EFUT variant Eq. (99) of the KB-RLO performs more uniformly than the KU-RLO Eq. (102). The underlying reason can be traced

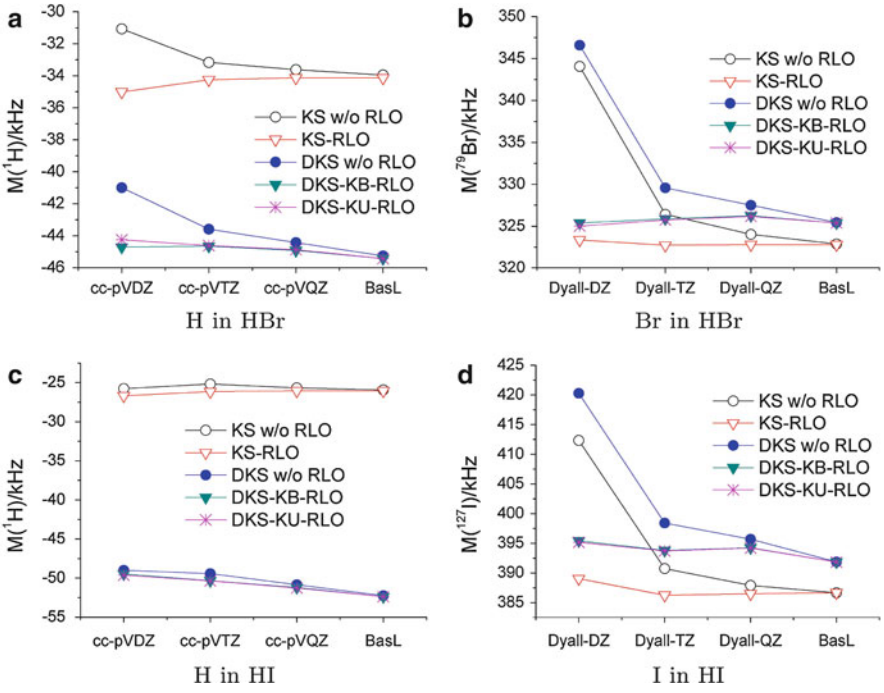


Fig. 1 Basis set convergence in the Kohn–Sham (KS) and Dirac–Kohn–Sham (DKS) calculations of the NSR constants of HBr and HI. w/o-RLO, without rotational London orbitals; KB-RLO, kinetically balanced RLO Eq. (99); KU-RLO, kinetically unbalanced RLO Eq. (102) (Reprinted with permission from J. Chem. Phys. 141, 164110 (2014). ©2014, American Institute of Physics)

to that, in the nrl, the contribution of negative-energy states vanishes in the former formulation but still survives in the latter formulation [24]. Therefore, the former is clearly favored both conceptually and numerically.

As for relativistic effects, it can be seen from Fig. 1 that they are negligibly small for the halogen NSR constants but significant for the hydrogen NSR constants. More specifically, relativity increases the hydrogen NSR constants by 1.1, 5.5, 31.5, and 96.7 % going from F to I [23]. This peculiar "inverse relativity" can be well understood by taking a close look at the $D_u^{01(K)}$ Eq. (23) and D_v^{10} Eq. (30) operators. Due to the r^{-2} behavior, the $D_u^{01(K)}$ operator is effective only in the vicinity of nucleus K, such that the matrix elements $D_{ia}^{01(K)}$ are only large for innermost atom-like orbitals ψ_i^{00}. On the other hand, the D_v^{10} operator can be rewritten as

$$\begin{aligned} D_v^{10} &= -\sum_i \left[(\vec{r}_i - \vec{R}_K) \times \vec{p}_i + \frac{1}{2}\Sigma_i \right]_v B_I \\ &\quad - \sum_i \left[(\vec{R}_K - \vec{R}_{\mathrm{NCM}}) \times \vec{p}_i \right]_v B_I \\ &= -\sum_i j_{iv}^{(K)} B_I - \sum_i \left[(\vec{R}_K - \vec{R}_{\mathrm{NCM}}) \times \vec{p}_i \right]_v B_I . \end{aligned} \tag{119}$$

Since atomic core orbitals are approximate eigenfunctions of the angular momentum $\vec{j}_i^{(K)}$, it is clear that the first term of Eq. (119) does not contribute much. The second term also contributes very little because it has an opposite parity to $D^{01(K)}$ Eq. (23), just like the previously discussed electron–nucleus Gaunt term Eq. (28). In short, only atomic core orbitals that are strongly distorted by neighboring heavy atoms can have significant relativistic contributions to NSR coupling constants. This is the case for H in HX but not for the halogen atoms. It is also instructive to make a cross comparison of the relativistic effects on the NSR and shielding constants. That the relativistic effects on the hydrogen NSR constants are enhanced by the heavier halogen is in line with the importance of relativistic corrections to hydrogen shielding constants due to the so-called heavy-atom effect on the light atoms (HALA) [40]. As the spin-orbit operator is an important contributor to both the shielding and NSR constants and is responsible for the HALA effect [18, 43], the strong relativistic effects on the hydrogen NSR constants should be expected. However, the situation is quite different for heavy atoms. While relativistic effects are in general significant for the shielding constants of heavy atoms, spin-orbit interactions no longer dominate. Instead, it is the atom-centered interaction operators that dominate the relativistic effects of the heavy atoms themselves [41,42]. On the contrary, many of these contributions do not enter in the expression for the relativistic corrections to the NSR constants [18,43]. Therefore, the relativistic effects on the halogen NSR constants appear numerically small. However, it deserves to be pointed out that relativistic effects on the NSR constants

of heavy atoms have similar scaling powers as those of the shielding constants [39]. It is just that the prefactors of the former are much smaller.

Finally, the available Dirac–Hartree–Fock (DHF) and Dirac–Kohn–Sham (DKS) results are collected in Table 2 to compare with experimental data [44–47]. Since the zero-point vibrational corrections are quite significant, e.g., as large as 10 % in the case of the fluorine NSR constant [23], they are subtracted from the experimental values. It can be seen that in most cases, there exists a rather substantial electron correlation effect as described by DFT. However, neither the generalized gradient approximation nor the hybrid exchange-correlation functionals perform better than DHF, especially for the halogen NSR constants. Instead, it is the simplest local density approximation (LDA) that performs best for these systems. It is clear that more refined exchange-correlation functionals need to be developed.

Absolute Nuclear Shielding Constants

The "direct relativistic mapping" discussed in section "Relativistic Mapping Between NMR and NSR Tensors" permits an easy and direct translation of experimentally measured NSR constants to absolute nuclear shielding constants. As shown in Fig. 2, the so-derived semi-experimental nuclear shielding constants of HX are indeed very close to the best theoretical values (designated as 4C-CASSCF in the figure; for more details, see Ref. [23]). In contrast, the "NR-Expt" results obtained by the "nonrelativistic mapping" Eq. (8) are far off, amounting to 4.5 ppm already for F in HF and even to 1,276 ppm for I in HI. They are actually very close to the nonrelativistic values (NR-CASSCF in the figure) for the halogen atoms. Yet, for the hydrogen atom, these "NR-Expt" results are close to the relativistic instead of the nonrelativistic values. As explained in section "NSR Constants of Hydrogen Halide," this "inverse relativity" behavior of the otherwise "nonrelativistic mapping" Eq. (8) arises from the inheritance of relativistic effects on the hydrogen NSR constants, which are, however, very small for the halogen NSR constants.

Similar findings were also observed for the absolute shielding constants of ^{119}Sn in SnH_4, $Sn(CH_3)_4$, and $SnCl_4$ [22] as well as in SnX (X = O, S, Se, Te) [43]. In these cases, the "nonrelativistic mapping" Eq. (8) is in error by about 1,000 ppm. Since the relativistic effects on the NSR constants of ^{119}Sn are very small, this large error is due to the substantial relativistic effects on the shielding of ^{119}Sn, which are atomic in nature and hence insensitive to the chemical environments. This argument can further be confirmed by a perturbational analysis of the relativistic corrections to the nuclear shielding and NSR constants [18, 43]. Specifically, there are five contributions that only appear for the nuclear shielding constants, four of which tend to sample the vicinity of the nucleus by sampling the electron density either at the nucleus ($\delta(r_K)$) or close to the nucleus ($r_K^{-n}, n \geq 2$). The only exception is the contribution $\sigma_{K,\epsilon\tau}^{p/KE-OZ}$ in the notation of Ref. [42]. However, this contribution involves the orbital angular momentum around the nucleus of interest and will

Table 2 Theoretical and experimental NSR constants (in kHz) of HX (X = F, Cl, Br, I)

Molecule	Nucleus	DHF[a]	DKS				Exp.	$\delta^{d}_{v,J}$	Exp.[e]
			LDA[b]	PP86[b]	BP86[c]	B3LYP[c]			
HF	^{19}F	336.52	297.01	333.17	360.23	356.99	307.65 ± 0.02[f]	29.08	278.57 ± 0.02
	^{1}H	−62.71	−74.75	−77.39	−71.87	−69.74	−71.10 ± 0.02[f]	1.06	−72.16±0.02
HCl	^{35}Cl	56.93	53.93	58.46	60.74	61.72	54.00 ± 0.15[g]	2.53	51.47 ± 0.15
	^{1}H	−39.78	−43.01	−44.25	−41.76	−41.52	−42.32 ± 0.70[g]	0.35	−42.67 ± 0.70
HBr	^{79}Br	290.84	301.64	325.44	335.10	336.46	290.83 ± 0.08[h]	11.88	278.95 ± 0.08
	^{1}H	−44.45	−43.14	−45.26	−44.48	−44.98	−41.27 ± 0.31[h]	−0.27	−41.00 ± 0.31
HI	^{127}I	331.99	365.16	391.86	403.06	403.90	351.1 ± 0.3[i]	10.66	340.44 ± 0.3
	^{1}H	−61.30	−49.22	−52.25	−54.43	−56.65	−49.22 ± 0.22[i]	−1.03	−48.19 ± 0.22
SD[j]	X	30.00	19.18	44.26	58.80	58.29			
	^{1}H	8.39	1.77	4.01	3.60	4.86			

[a] Ref. [39]
[b] Ref. [23]
[c] Ref. [39]
[d] Rovibrational corrections [23]
[e] Rovibronically corrected experimental values
[f] Ref. [44]
[g] Ref. [45]
[h] Ref. [46]
[i] Ref. [47]
[j] Standard deviations from Exp.[e]

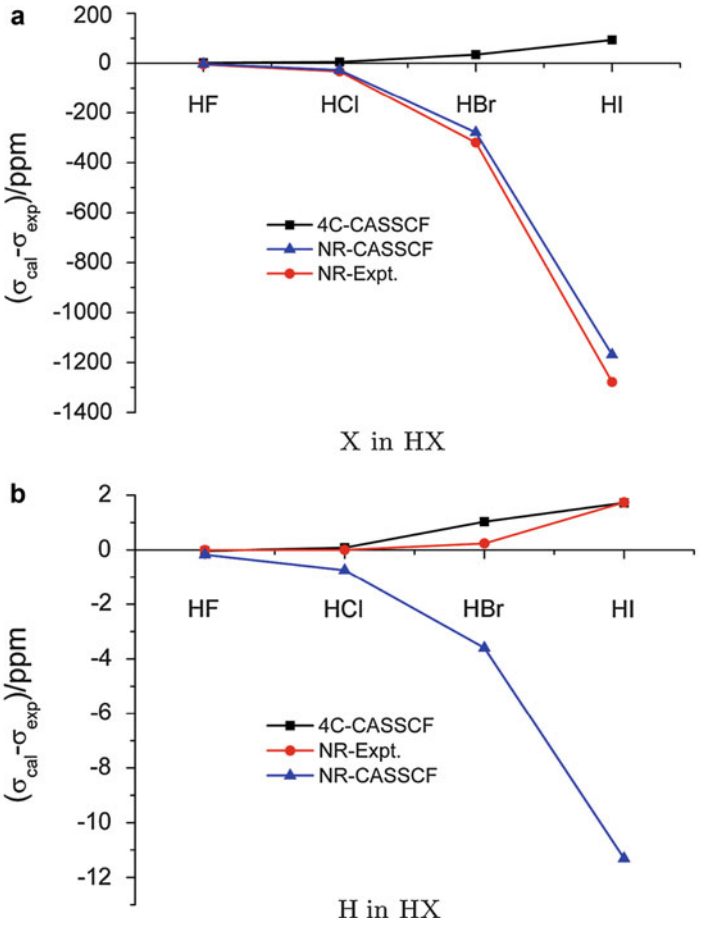


Fig. 2 Deviations of theoretical from experimental nuclear shielding constants of HX (X = F, Cl, Br, I) (The data was taken from Tables 4 and 7 of Ref. [23])

vanish for molecules with spherical symmetry or have only one nonvanishing component for linear molecules. Thus, for highly symmetric molecules, the relativistic corrections signaling the breakdown of the "nonrelativistic mapping" Eq. (8) have a strongly atomic character. Indeed, the perturbational analysis [48] of the relativistic corrections to the shielding of ^{119}Sn in SnH_4 reveals that these five contributions amount to 978 ppm, very close to the 1,000 ppm [22].

The above results reveal that if the "nonrelativisitic mapping" Eqs. (8) or (9) is used for deriving the nuclear shielding constants from the experimental NSR constants, the difference between the relativistic effects on the calculated nuclear shielding and NSR constants must further be added in [22] as shown clearly in Eq. (118) [23]. Along this line, accurate absolute shielding constants have been established for various systems such as XF_6 (X = B, Al, Ga, In, Tl) [49], XF_6 (X = S,

Se, Te, Mo, W) [50], and XCO (X = Ni, Pd, Pt) [51]. It is also possible to determine in this way more accurate nuclear magnetic dipole moments [52].

Given the robustness of the "relativistic mapping" Eqs. (110) or (118), the "nonrelativistic mapping" Eqs. (8) or (9) remains only useful for generating absolute shielding constants to calibrate nonrelativistic ab initio calculations. Still, however, the relativistic effects must first be subtracted from the experimental NSR constants. For instance, for the shielding of ^{33}S in $H_2^{33}S$, a value of 716.4(4.6) ppm could be obtained from the "nonrelativistic mapping" Eq. (8), which is to be compared with the nonrelativistic coupled-cluster value of 719.0 ppm [11]. After subtracting the relativistic effects from the experimental NSR constant of ^{33}S, the former was increased to 718.3(4.6) ppm [12], which is in even better agreement with the latter. In contrast, the true shielding of ^{33}S in $H_2^{33}S$ amounts to 742.9(4.6) ppm [12] according to the "relativistic mapping" Eqs. (110)/(118), a very clear example of the breakdown of the "nonrelativistic mapping" Eqs. (8)/(9) even for atoms as light as sulfur.

Summary

The relativistic theory [18–21] of the nuclear spin-rotation tensor has critically been reviewed, focusing especially on the relativistic mapping [20–23] between the nuclear magnetic shielding and NSR tensors. Thanks to the robustness [23] of this relativistic mapping, the nonrelativistic counterpart [1–3, 9] that has been in use for many decades should now be regarded as outdated.

Since the theory itself was formulated only rather recently, practical applications have so far been performed only at the DHF and DKS levels. The available results suggest that the standard exchange-correlation functionals cannot match the very high experimental accuracy that can be obtained for the NSR tensors. As such, the NSR tensors represent an important test ground for future development of more refined exchange-correlation functionals. Nevertheless, accurate absolute shielding tensors can still be obtained through the combined use of experimental NSR tensors and relativistic corrections calculated with DFT. More accurate approaches such as four-component relativistic coupled-cluster theory are highly desired. Before such approaches are made available, perturbation theory [43] appears very promising.

Acknowledgements The research of this work was supported by grants from the NSFC (Project Nos. 21033001, 21273011, 21473002, 21173006, and 21290192). KR acknowledges support from the Research Council of Norway through a Centre of Excellence Grant (Grant No. 179568) and helpful discussions with S. Komorovský.

References

1. Flygare WH (1964) Spin-rotation interaction and magnetic shielding in molecules. J Chem Phys 41:793
2. Flygare WH, Goodisman J (1968) Calculation of diamagnetic shielding in molecules. J Chem Phys 49:3122

3. Flygare WH (1974) Magnetic interactions in molecules and an analysis of molecular electronic charge distribution from magnetic parameters. Chem Rev 74:653
4. Puzzarini C, Stanton JF, Gauss J (2010) Quantum-chemical calculation of spectroscopic parameters for rotational spectroscopy. Int Rev Phys Chem 29:273
5. Helgaker T, Coriani S, Jørgensen P, Kristensen K, Olsen J, Ruud K (2012) Recent advances in wave function-based methods of molecular-property calculations. Chem Rev 112:543
6. Teale AM, Lutnæs OB, Helgaker T, Tozer DJ, Gauss J (2013) Benchmarking density-functional theory calculations of NMR shielding constants and spin-rotation constants using accurate coupled-cluster calculations. J Chem Phys 138:024111
7. Abragam A (1961) The principles of nuclear magnetic resonance. Oxford University Press, Oxford
8. Helgaker T, Jaszuński M, Ruud K (1999) Ab initio methods for the calculation of NMR shielding and indirect spin-spin coupling constants. Chem Rev 99:293
9. Ramsey NF (1950) Magnetic shielding of nuclei in molecules. Phys Rev 78:699
10. Sundholm D, Gauss J, Schäfer A (1996) Rovibrationally averaged nuclear magnetic shielding tensors calculated at the coupled-cluster level. J Chem Phys 105:11051
11. Helgaker T, Gauss J, Cazzoli G, Puzzarini C (2013) ^{33}S hyperfine interactions in H_2S and SO_2 and revision of the sulfur nuclear magnetic shielding scale. J Chem Phys 139:244308
12. Komorovsky S, Repisky M, Malkin E, Ruud K, Gauss J (2015) Communication: the absolute shielding scales of oxygen and sulfur revisited. J Chem Phys 142:091102
13. Antušek A, Jackowski K, Jaszuński M, Makulski W, Wilczek M (2005) Nuclear magnetic dipole moments from NMR spectra. Chem Phys Lett 411:111
14. Jaszuński M, Antušek A, Garbacz P, Jackowski K, Makulski W, Wilczek M (2012) The determination of accurate nuclear magnetic dipole moments and direct measurement of NMR shielding constants. Prog Nucl Magn Reson 67:49
15. Puzzarini C, Cazzoli G, Harding ME, Vázquez J, Gauss J (2015) The hyperfine structure in the rotational spectra of $D_2^{17}O$ and $HD^{17}O$: confirmation of the absolute nuclear magnetic shielding scale for oxygen. J Chem Phys 142:124308
16. Visscher L, Enevoldsen T, Saue T, Jensen HJA, Oddershede J (1999) Full four-component relativistic calculations of NMR shielding and indirect spin-spin coupling tensors in hydrogen halides. J Comput Chem 20:1262
17. Saue T (2005) Spin-interactions and the non-relativistic limit of electrodynamics. Adv Quantum Chem 48:383
18. Aucar IA, Gómez SS, Ruiz de Azúa MC, Giribet CG (2012) Theoretical study of the nuclear spin-molecular rotation coupling for relativistic electrons and non-relativistic nuclei. J Chem Phys 136:204119
19. Aucar IA, Gómez SS, Melo JI, Giribet CC, Ruiz de Azúa MC (2013) Theoretical study of the nuclear spin-molecular rotation coupling for relativistic electrons and non-relativistic nuclei. II. Quantitative results in HX (X=H,F,Cl,Br,I) compounds. J Chem Phys 138:134107
20. Xiao Y, Liu W (2013) Body-fixed relativistic molecular Hamiltonian and its application to nuclear spin-rotation tensor. J Chem Phys 138:134104
21. Xiao Y, Liu W (2013) Body-fixed relativistic molecular Hamiltonian and its application to nuclear spin-rotation tensor: linear molecules. J Chem Phys 139:034113
22. Malkin E, Komorovsky S, Repisky M, Demissie TB, Ruud K (2013) The absolute shielding constants of heavy nuclei: resolving the enigma of the ^{119}Sn absolute shielding. J Phys Chem Lett 4:459
23. Xiao Y, Zhang Y, Liu W (2014) New experimental NMR shielding scales mapped relativistically from NSR: theory and application. J Chem Theory Comput 10:600
24. Xiao Y, Zhang Y, Liu W (2014) Relativistic theory of nuclear spin-rotation tensor with kinetically balanced rotational London orbitals. J Chem Phys 141:164110
25. Gauss J, Ruud K, Helgaker T (1996) Perturbation-dependent atomic orbitals for the calculation of spin-rotation constants and rotational g tensors. J Chem Phys 105:2804
26. Aucar IA, Gómez SS, Giribet CG, Ruiz de Azúa MC (2013) Breit interaction effects in relativistic theory of the nuclear spin-rotation tensor. J Chem Phys 139:094112

27. Åstrand PO, Ruud K, Taylor PR (2000) Calculation of the vibrational wave function of polyatomic molecules. J Chem Phys 112:2655
28. London F (1937) The quantic theory of inter-atomic currents in aromatic combinations. J Phys Radium 8:397
29. Ditchfield R (1972) Molecular orbital theory of magnetic shielding and magnetic susceptibility. J Chem Phys 56:5688
30. Wolinski K, Hinton JF, Pulay P (1990) Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. J Am Chem Soc 112:8251
31. Bak KL, Jørgensen P, Helgaker T, Ruud K, Jensen HJA (1993) Gauge-origin independent multiconfigurational self-consistent-field theory for vibrational circular dichroism. J Chem Phys 98:8873
32. Cheng L, Xiao Y, Liu W (2009) Four-component relativistic theory for nuclear magnetic shielding: magnetically balanced gauge-including atomic orbitals. J Chem Phys 131:244113
33. Xiao Y, Sun Q, Liu W (2012) Fully relativistic theories and methods for NMR parameters. Theor Chem Acc 131:1080
34. Xiao Y, Peng D, Liu W (2007) Four-component relativistic theory for nuclear magnetic shielding constants: the orbital decomposition approach. J Chem Phys 126:081101
35. Sun Q, Liu W, Kutzelnigg W (2011) Comparison of restricted, unrestricted, inverse, and dual kinetic balances for four-component relativistic calculations. Theor Chem Acc 129:423
36. Stanton RE, Havriliak S (1984) Kinetic balance: a partial solution to the problem of variational safety in Dirac calculations. J Chem Phys 81:1910
37. Xiao Y, Liu W, Cheng L, Peng D (2007) Four-component relativistic theory for nuclear magnetic shielding constants: critical assessments of different approaches. J Chem Phys 126:214101
38. Cheng L, Xiao Y, Liu W (2009) Four-component relativistic theory for NMR parameters: unified formulation and numerical assessment of different approaches. J Chem Phys 130:144102
39. Komorovsky S, Repisky M, Malkin E, Demissie TB, Ruud K (2015) Four-component relativistic density-functional theory calculations of nuclear spin-rotation constants: relativistic effects in p-block hydrides. J Chem Theory Comput Accpte 11:3729
40. Pyykkö P, Görling A, Rösch N (1987) A transparent interpretation of the relativistic contribution to the NMR Heavy-atom chemical-shift. Mol Phys 61:195
41. Manninen P, Lantto P, Vaara J, Ruud K (2003) Perturbational ab initio calculations of relativistic contributions to nuclear magnetic resonance shielding tensors. J Chem Phys 119:2623
42. Manninen P, Ruud K, Lantto P, Vaara J (2005) Leading-order relativistic effects on nuclear magnetic resonance shielding tensors. J Chem Phys 122:114107; Erratum *ibid*. 124:149901 (2006)
43. Komorovsky S, Repisky M, Demissie TB, Malkin E, Ruud K, Nuclear spin-rotation and shielding constants in the SnX (S = O, S, Se, Te) series. Unpublished
44. Bass SM, Deleon RL, Muenter JS (1987) Stark, Zeeman, and hyperfine properties of v=0, v=1, and the equilibrium configuration of hydrogen fluoride. J Chem Phys 86:4305
45. Cazzoli G, Puzzarini C (2004) Hyperfine structure of the J=1 ← 0 transition of $H^{35}Cl$ and $H^{37}Cl$: improved ground state parameters. J Mol Spectrosc 226:161
46. Van Dijk FA, Dymanus A (1969) Hyperfine structure of the rotational spectrum of HBr and in the submillimeter wave region. Chem Phys Lett 4:170
47. Van Dijk FA, Dymanus A (1968) Hyperfine structure of the rotational spectrum of HI in the submillimeter region. Chem Phys Lett 2:235
48. Jaszuński M, Ruud K (2006) Nuclear magnetic resonance shielding constants in XH4 group XIV hydrides. Mol Phys 104:2139

49. Jaszuński M, Demissie TB, Ruud K (2014) Spin-rotation and NMR shielding constants in XF molecules (X = B, Al, Ga, In, and Tl). J Phys Chem A 118:9588
50. Ruud K, Demissie TB, Jasuński M (2014) Ab initio and relativistic DFT study of spin-rotation and NMR shielding constants in XF_6 molecules, X = S, Se, Te, Mo, and W. J Chem Phys 140:194308
51. Demissie TB, Jaszuński M, Malkin E, Komorovsky S, Ruud K (2015) NMR shielding and spin-rotation constants in XCO (X = Ni, Pd, Pt) molecules. Mol Phys 113:1576
52. Jaszuński M, Repisky M, Demissie TB, Komorovsky S, Malkin E, Ruud K, Garbacz P, Jackowski K, Makulski W (2013) Spin-rotation and NMR shielding constants in HCl. J Chem Phys 139:234302

Relativistic Methods for Calculating Electron Paramagnetic Resonance (EPR) Parameters

23

Hélène Bolvin and Jochen Autschbach

Contents

Preliminaries: Units, Notations, Acronyms . . . 726
Introduction and Background: EPR Parameters and Spin Hamiltonians . . . 726
Computational Methods for EPR Parameter Calculations . . . 731
Representation of the Pseudo-spin Hamiltonian in an Ab Initio Framework . . . 731
Wavefunction-Based Methods for EPR Calculations . . . 736
Hartree-Fock and Kohn-Sham Methods for EPR Calculations . . . 736
Operators for the Zeeman and HFC Interactions . . . 739
Higher-Order EPR Parameters and Mapping of Ab Initio to Pseudo-spin Functions . . . 743
Signs of EPR *g*-Factors and Hyperfine Couplings . . . 749
Selected Case Studies . . . 751
Summary . . . 761
References . . . 761

Abstract

Basic concepts for calculating electronic paramagnetic resonance are discussed, with a focus on methods that are suitable for molecules containing heavy elements. Inclusion of relativistic effects is essential in such calculations. Selected examples are presented to illustrate practical applications of these theoretical methods.

H. Bolvin
Laboratoire de Physique et de Chimie Quantiques, Université Toulouse 3, Toulouse, France
e-mail: bolvin@irsamc.ups-tlse.fr

J. Autschbach (✉)
Department of Chemistry, University at Buffalo, State University of New York, Buffalo, NY, USA
e-mail: jochena@buffalo.edu

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_12

Keywords

Electron paramagnetism • Electron paramagnetic resonance • Spin Hamiltonian • Spin–orbit coupling • Hyperfine coupling • Electron g-factor • Zeeman interaction • Electron spin • Nuclear spin • Finite nucleus effects

Preliminaries: Units, Notations, Acronyms

The reader is assumed to be familiar with basic concepts of quantum mechanics – including relativistic methods covered in other chapters – and basic concepts of computational chemistry. SI units are employed. Nuclear motion is not considered; the focus is on electronic structure and the resulting magnetic properties. The symbols $\cdot$ and $\times$ indicate inner and outer products, respectively, for vectors and matrices or tensors. Bold-italic notation such as $\boldsymbol{r}$, $\hat{\boldsymbol{S}}$, and $\boldsymbol{\mu}$ is used for vectors and vector operators, while upright-bold such as $\mathbf{a}$, $\mathbf{G}$, and $\boldsymbol{\upmu}$ is used for matrices and rank-2 tensors.

The following acronyms are used occasionally in the text:

AO	atomic orbital (basis function or actual AO)
CAS	complete active space
DFT	Density Functional Theory (usually KS, "pure" and generalized KS variants)
EM	electromagnetic
GIAO	gauge-including atomic orbital
HF	Hartree-Fock
HFC	hyperfine coupling
KS	Kohn-Sham
MO	molecular orbital
NR	nonrelativistic (calculation excluding any relativistic effects)
PV	principal value (of a tensor)
PAS	principal axis system (of a tensor)
QM	quantum mechanical (e.g., in reference to Dirac, Schrödinger Eqs.)
SO	spin-orbit (usually means calculation also includes SR effects)
SOS	sum over states
SR	scalar relativistic (relativistic calculation without SO effects)
WFT	wavefunction theory
ZFS	zero-field splitting

Introduction and Background: EPR Parameters and Spin Hamiltonians

Many of the chemical species encountered in the laboratory and in everyday life have nondegenerate closed-shell ground states. But there are also many exceptions, such as open-shell metal complexes, stable radicals, and most atoms. In the absence

of external electromagnetic (EM) fields, such species may afford a degenerate electronic ground state and degenerate excited states. Species with closed-shell ground states may also afford degenerate excited states such as excited spin triplets. The term *degeneracy* means that an electronic state λ with energy E_λ may have d_λ state components $|\lambda, a\rangle$, with $a = 1 \ldots d_\lambda$, such that each $|\lambda, a\rangle$ and any linear combination thereof is a solution to the field-free quantum mechanical (QM) equation describing the system (e.g., Schrödinger equation, Dirac equation, approximate two-component relativistic QM methods, as discussed elsewhere in this handbook) with the same energy E_λ. The index λ may simply be a number counting the energy levels of the system, or it may be a spectroscopic symbol characterizing a state of interest, or a symmetry label, or a combination thereof. The discussion excludes cases of accidental degeneracy.

Electron paramagnetic resonance (EPR) [1–4] is a primary tool for studying degenerate and nearly degenerate electronic states experimentally. An external magnetic field $\boldsymbol{B}$ splits the degeneracy (*Zeeman effect*) to yield a new set of states. EM radiation of a suitable frequency may then induce transitions among them and allow to measure the energy splittings spectroscopically. The parameters extracted from the spectra (vide infra) contain a wealth of information about the electronic structure and molecular structure.

To illustrate the effect utilized in EPR spectroscopy, consider a single unpaired electron and – first – neglect spin-orbit (SO) coupling. This situation represents a spin doublet (a twofold degenerate state) with spin quantum number $S = 1/2$. If there is no external magnetic field, the two possible orientations of the spin projection onto a quantization axis ($M_S = \pm 1/2$) have the same energy. Associated with the spin angular momentum vector $\boldsymbol{S}$ is a magnetic dipole moment $\boldsymbol{m}_e = -g_e \beta_e \boldsymbol{S}$ of the electron, with $\beta_e = e\hbar/(2m_e)$ being the Bohr magneton and g_e the free electron g-value or g-factor with a current experimental value [5] of 2.002 319 304 361 53 (53). The Dirac equation predicts $g_e = 2$ exactly. Therefore, $g_e = 2$ is used occasionally in the following. The small differences in the values of g_e are due to quantum electrodynamic corrections.

For $S = 1/2$, the magnetic moment associated with the spin has two projections onto a quantization axis. If a static external magnetic field is applied, the direction of $\boldsymbol{B}$ defines the quantization axis. With the field present, the quantized projections of $\boldsymbol{m}_e$ do not have the same energy. In classical physics, the energy of a magnetic dipole $\boldsymbol{m}$ in a magnetic field $\boldsymbol{B}$ is

$$E = -\boldsymbol{m} \cdot \boldsymbol{B} \tag{1}$$

The lowest energy is for $\boldsymbol{m}$ and $\boldsymbol{B}$ being antiparallel, and the highest energy is for $\boldsymbol{m}$ and $\boldsymbol{B}$ being parallel. This is the physical mechanism that keeps a compass needle pointing toward the magnetic north pole. Quantum mechanically, a phenomenological Zeeman Hamiltonian that describes such an effect for a quantized spin magnetic moment is

$$\hat{\mathcal{H}}^{\mathrm{Z}} = g_e \beta_e \hat{\boldsymbol{S}} \cdot \boldsymbol{B} \tag{2}$$

with $\hat{\boldsymbol{S}}$ being the spin vector operator. The negative sign in (1) is canceled by the negative sign relating $\boldsymbol{m}_e$ to the spin. One may choose a coordinate system such that $\boldsymbol{B}$ is oriented along the z axis, with amplitude B_0. The Hamiltonian (2) then reads

$$\hat{\mathcal{H}}^Z = g_e \beta_e B_0 \hat{S}_z \tag{3}$$

The eigenvalues are those of $\hat{S}_z$ times $g_e \beta_e B_0$, i.e., $\pm(1/2) g_e \beta_e B_0$. The magnetic field lifts the degeneracy of the spin doublet. A transition from the lower energy level ($\boldsymbol{B}$ and $\boldsymbol{S}$ "antiparallel") to the higher one ($\boldsymbol{B}$ and $\boldsymbol{S}$ "parallel") requires an energy of

$$\Delta E = h\nu = g_e \beta_e B_0 \tag{4}$$

In the equation above, ν is the frequency of EM radiation used to induce the transition. The frequency is approximately 28 GHz/T. In a typical EPR spectrometer, B_0 is 0.34 T (3,400 Gauss), which translates to a free electron spin-1/2 resonance frequency of about 9.5 GHz. This is radiation in the X-band microwave region of the EM spectrum. For general spin values S, the degeneracy of the projection is $2S + 1$. A magnetic field splits these into $2S + 1$ individual states. The selection rule is that transitions with $\Delta M_S = 1$ are allowed.

One may repeat the calculation for the spin $\boldsymbol{I}_K$ of some nucleus no. K, by substituting the electron spin magnetic dipole moment by the nuclear spin magnetic dipole moment $\boldsymbol{m}_K = g_K \beta_N \boldsymbol{I}_K$. Here, g_K is the g-factor for a given nuclear isotope, and $\beta_N = e\hbar/(2M_P)$ is the nuclear magneton, with M_P being the proton mass. The latter is approximately 2,000 times greater than the electron mass, and therefore, the associated transition energies for nuclear spins are in the radio-frequency range (tens of MHz) at magnetic field strengths used in EPR spectrometers. Transitions between nuclear spin projections are observed directly in nuclear magnetic resonance (NMR). At EPR spectrometer field strengths, the energy splitting for nuclear transitions is very small, and one may assume equal Boltzmann populations of the nuclear spin projection states in the absence of other magnetic fields. However, there is a magnetic coupling of electronic and nuclear magnetic moments called *hyperfine coupling* (HFC). Hyperfine coupling gives rise to hyperfine structure of EPR spectra when nuclei with nonzero spin are present.

There is another type of fine structure that can be observed in EPR experiments for systems with $S \geq 1$. For a given magnetic field strength B_0, the $2S + 1$ spin projections with different M_S would be expected to be equally spaced energetically, meaning that any of the possible $2S$ allowed transitions with $\Delta M_S = 1$ give resonances at the same frequency. However, the spectra may exhibit $2S$ distinct features even at lower field strengths [1], which indicates unequal energetic separations of the M_S components in the presence of the field. This fine structure can be traced back to a removal of the degeneracy of the spin multiplet already at zero field. The effect is therefore called *zero field splitting* (ZFS). A spin $S \geq 1$

implies that there are two or more unpaired electrons. The physical origin of ZFS is a magnetic interaction between pairs of electrons, either directly (a dipolar spin–spin interaction arising from relativistic corrections to the electron–electron interaction) or mediated via SO coupling (which has one- and two-electron contributions). The ZFS can therefore be associated with relativistic effects. For systems with heavy elements, as well as for many lighter systems, the SO contribution to ZFS is the dominant one.

The observed electronic magnetic moment resulting from an electronically degenerate state may differ from what is expected based solely on the spin quantum number. For instance, there may be an orbital degeneracy present in an electronic state, meaning that the observed magnetic moment is not only due to an electron spin but also due to an orbital angular momentum. The interaction with the external field can be expressed via the total angular momentum $\boldsymbol{J}$, which is obtained from the vector addition of the quantized spin and orbital angular momenta, $\boldsymbol{S}$ and $\boldsymbol{L}$. As an example, the $S = 1/2, L = 1$ (^2P) ground state of the fluorine atom has a total angular momentum quantum number $J = 3/2$, with total spin and orbital angular momentum parallel, giving a fourfold degenerate state (M_J ranging from $+3/2$ to $-3/2$) whose components split in the presence of a magnetic field. Transitions in the EPR experiment may be observed for $\Delta M_J = \pm 1$. Instead of Eq. (4), the transition frequencies are determined by

$$\Delta E = h\nu = g_J \beta_e B_0 \tag{5}$$

with $g_J = 4/3$ for the fluorine $^2P_{3/2}$ state, instead of 2. Here, g_J is the Landé g-factor; the experimentally observed g-factor obtained from matching the measured resonance frequency with Eq. (5) for known field strength B_0 is very close to this number. The large difference of g_J from the free-electron $g_e = 2$ arises because the state reflects not only a spin doublet but also an orbital triplet.

In the absence of orbital degeneracy, an orbital magnetic moment and deviations of observed g-factors from the free electron value may arise because of SO coupling. For organic doublet radicals with only light elements, $S = 1/2, M_S = \pm 1/2$ are basically good quantum numbers. However, even in this situation, the observed g-factors may differ from 2 (typically ranging from 1.9 to 2.1) because of SO coupling. Because of the small deviations, it is sometimes preferred to report g-shifts $\Delta g = g - g_e$ (often in units of parts per thousand (ppt)), in analogy to NMR chemical shifts. Since SO coupling is a relativistic effect, the presence of g-shifts for orbitally nondegenerate states directly indicates relativistic effects. If SO coupling is strong, S and M_S may not be good quantum numbers at all. In this case, g-shifts can become very large even in the absence of orbital degeneracy (meaning an absence of orbital angular momentum in the corresponding scalar relativistic (SR) state). A case in point is the doublet ground state of NpF_6 for which the observed [6] g-factor is -0.6.

To summarize: A degenerate paramagnetic electronic state gives rise to a more or less complicated pattern of EPR resonances. Given the potential influence of SO coupling, orbital angular momenta, and ZFS, the spectrum is usually interpreted and

quantified by invoking the concept of a *pseudo-spin* $\mathcal{S}$ rather than the actual electron spin S. The value of $\mathcal{S}$ defines the degeneracy, $2\mathcal{S}+1$, of the state that is split by the magnetic field into components with different $\mathcal{M}_\mathcal{S}$. The various effects discussed above can be included in a phenomenological pseudo-spin Hamiltonian, which, in lowest order, reads

$$\hat{\mathcal{H}}_S = \beta_e \boldsymbol{B}\cdot\mathbf{g}\cdot\hat{\boldsymbol{\mathcal{S}}} + \boldsymbol{I}_K\cdot\mathbf{a}_K\cdot\hat{\boldsymbol{\mathcal{S}}} + \hat{\boldsymbol{\mathcal{S}}}\cdot\mathbf{d}\cdot\hat{\boldsymbol{\mathcal{S}}} \tag{6}$$

The parameters in the Hamiltonian are determined by requiring that the transitions between its eigenstates reproduce the observed spectrum. The spin Hamiltonian is designed such that its elements within the set of fictitious spin eigenstates are the same as the matrix elements of the true Hamiltonian within the set of true eigenstates. It supposes a correspondence between pseudo-spin and true eigenstates, up to a phase factor common to all the eigenvectors. While this assignment can be rather arbitrary, the basic requirement is that the spin Hamiltonian in the fictitious space transforms in coherence as the real Hamiltonian does in the real space, either by time reversal or by the spatial symmetries of the molecule [7].

On the right-hand side of Eq. (6) are, from left to right, the pseudo-spin operators for the Zeeman interaction, the HFC interaction, and the ZFS. Only the pseudo-spin related to the electronic state is treated quantum mechanically. The nuclear spin and external magnetic field are parameters. In Eq. (6), $\mathbf{g}$, $\mathbf{a}_K$, and $\mathbf{d}$ are 3×3 matrices parametrizing the various interactions. The fact that they are written in matrix form reflects the possibility that the observed interactions may be *anisotropic*. For example, observed g-factors for a molecule with axial symmetry may be very different if the magnetic field is oriented along the axial direction or perpendicular to it. Higher-order terms requiring additional sets of parameters in Eq. (6) may be required to reflect the full complexity of an EPR spectrum, as discussed in section "Higher-Order EPR Parameters and Mapping of Ab Initio to Pseudo-spin Functions."

The matrices $\mathbf{g}$ and $\mathbf{a}_K$ are often referred to as the g-"tensor" and the HFC "tensor." It was pointed out in the book by Abragam and Bleaney [4] that they are in fact not proper rank-2 tensors. Following a suggestion by Atherton [1], one may refer to them as the Zeeman coupling matrix ($\mathbf{g}$) and the HFC matrix ($\mathbf{a}_K$) instead. The g-factor observed for a magnetic field in the direction of a unit vector $\boldsymbol{u}$ in the laboratory coordinate system is given by

$$g_{\boldsymbol{u}} = \pm(\boldsymbol{u}\cdot\mathbf{g}\mathbf{g}^T\cdot\boldsymbol{u})^{1/2} \tag{7}$$

(superscript T indicating a matrix transpose). One may define a hyperfine coupling associated with a particular quantization direction chosen for the pseudo-spin in a similar way. The corresponding objects $\mathbf{G}=\mathbf{g}\mathbf{g}^T$ and $\mathbf{A}_K=\mathbf{a}_K\mathbf{a}_K^T$ are rank-2 tensors whose eigenvalues and eigenvectors define the squares of the principal values (PVs) of $\mathbf{g}$ and $\mathbf{a}_K$ and a principal axis system (PAS) for each of the interactions. For example, the principal g-factors correspond to the g-factors that are observed when the direction of the magnetic field $\boldsymbol{B}$ coincides with one of

the principal axes. Section "Signs of EPR g-Factors and Hyperfine Couplings" addresses the question of the signs of the PVs of the Zeeman and HFC interaction in more detail.

Section "Computational Methods for EPR Parameter Calculations" sketches different computational relativistic methods by which to obtain the EPR parameters in the spin Hamiltonian of Eq. (6) from first principles. As already mentioned, SO coupling plays an important role. So do SR effects. Selected illustrative examples are presented in section "Selected Case Studies."

Computational Methods for EPR Parameter Calculations

Representation of the Pseudo-spin Hamiltonian in an Ab Initio Framework

Equation (6) and generalizations thereof present some conceptual challenges when addressing the problem by relativistic or nonrelativistic (NR) molecular quantum mechanics. The reason is that the pseudo-spin operator $\hat{\boldsymbol{\mathcal{S}}}$ may have little in common with the electron spin operator $\hat{\boldsymbol{S}}$ if SO coupling or ZFS is strong, or if the orbital angular momenta are not quenched. What can be done instead follows roughly the following sequence [3], if a calculation starts out with a pure spin multiplet (i.e., from a NR or SR reference without orbital degeneracy) and if effects from SO coupling can be dealt with as a perturbation:

- Define a Hamiltonian $\hat{\mathcal{H}}_0$ for the system in the absence of external EM fields and find its eigenstates with energies E_λ, spin degeneracies d_λ, and corresponding orthonormal QM eigenfunctions $|\lambda, a\rangle$.
- Consider a perturbation: a homogeneous external field $\boldsymbol{B}$ for the Zeeman interaction, the hyperfine magnetic field from a nuclear spin magnetic moment $\boldsymbol{m}_K$, or spin-dependent perturbations. Define a corresponding perturbation Hamiltonian $\hat{\mathcal{H}}'$. The effects from SO coupling may also be absorbed into $\hat{\mathcal{H}}'$. The perturbations are assumed to be weak enough such that one can identify the eigenstates of $\hat{\mathcal{H}}_0 + \hat{\mathcal{H}}'$ corresponding to a multiplet λ of $\hat{\mathcal{H}}_0$ of interest. Diagonalization of the matrix $\mathbf{H}_0 + \mathbf{H}'$ representing the Hamiltonian in the complete set (or a large subset) of eigenstates of $\hat{\mathcal{H}}_0$, or finding a selected number of eigenfunctions and eigenvalues with techniques such as the Davidson or Lanczos algorithms, would give the eigenfunctions and energies for the perturbed system. Typically, there would be a mixing among eigenvectors belonging to a multiplet λ as well as some admixture from components of other multiplets.
- Instead, select a multiplet λ with degeneracy d_λ of interest corresponding to $\hat{\mathcal{H}}_0$, and seek a $d_\lambda \times d_\lambda$ matrix representation $\mathbf{H}_{\text{eff}}$ of an *effective Hamiltonian* $\hat{\mathcal{H}}_{\text{eff}}$ with the following properties: (i) the eigenvalues of $\mathbf{H}_{\text{eff}}$ are the same as those of $\mathbf{H}_0 + \mathbf{H}'$ for the perturbed multiplet λ. (ii) The eigenvectors of $\mathbf{H}_{\text{eff}}$ describe how

the components of the unperturbed multiplet mix under the perturbation. There are various ways by which $\mathbf{H}_{\text{eff}}$ can be calculated. A well-known approach is by perturbation theory as an approximation to second order, which gives for a matrix element related to the multiplet λ:

$$\langle \lambda a|\hat{\mathcal{H}}_{\text{eff}}|\lambda a'\rangle = \delta_{aa'}E_\lambda + \langle \lambda a|\hat{\mathcal{H}}^{(a)}|\lambda a'\rangle + \sum_{\mu\neq\lambda}\sum_{b=1}^{d_\mu}\frac{\langle \lambda a|\hat{\mathcal{H}}^{(b)}|\mu b\rangle\langle \mu b|\hat{\mathcal{H}}^{(c)}|\lambda a'\rangle}{E_\lambda - E_\mu} \tag{8}$$

Here, $\hat{\mathcal{H}}^{(a)}$, $\hat{\mathcal{H}}^{(b)}$, and $\hat{\mathcal{H}}^{(c)}$ are parts of $\hat{\mathcal{H}}'$ such that the overall matrix element (minus the $\delta_{aa'}E_\lambda$ part) affords terms that are linear in $\boldsymbol{B}$ or $\boldsymbol{I}_K$, or bi-linear in electron spin operators.

- Apart from a constant shift on the diagonal, the matrix representation of the pseudo-spin Hamiltonian, Eq. (6), written in terms of $|\mathscr{M}_\mathcal{S}\rangle$ pseudo-spin projects, is then supposed to have the same elements as those of $\mathbf{H}_{\text{eff}}$. For the Zeeman interaction, the contribution to $\mathbf{H}_{\text{eff}}$ should be linear in $\boldsymbol{B}$. Then, $\hat{\mathcal{H}}^{(a)}$ and $\hat{\mathcal{H}}^{(b)}$ are the Zeeman operator (in a suitable relativistic form), and $\hat{\mathcal{H}}^{(c)}$ additionally considers SO effects. If $\hat{\mathcal{H}}^{(a)}$ and $\hat{\mathcal{H}}^{(b)}$ are instead QM operators linear in $\boldsymbol{I}_K$ describing the nuclear hyperfine field, a mapping onto the HFC part of the pseudo-spin Hamiltonian can be made. Finally, if $\hat{\mathcal{H}}^{(a)}$ is the dipolar spin–spin interaction operator, and $\hat{\mathcal{H}}^{(b)}$ and $\hat{\mathcal{H}}^{(c)}$ represent SO interactions, one obtains an effective Hamiltonian quadratic in the electron spin which represents ZFS.

In the previous approach, the matching between real and fictitious states is made according to $|\lambda, M\rangle \equiv |\mathcal{S}, \mathscr{M}_\mathcal{S}\rangle$ since the model space is a spin multiplet; it supposes a similarity between the real spin and the pseudo-spin. This is only valid in the weak SO limit, but the procedure permits the calculation of the spin Hamiltonian parameters for all values of the pseudo-spin $\mathcal{S}$ [8]. Equation (8) misses quadratic contributions which may be important [9].

The value of the pseudo-spin $\mathcal{S}$ defines the size of the model space. If the state of interest, usually the ground state, is d-fold degenerate or nearly degenerate, $\mathcal{S}$ is chosen such that $d = 2\mathcal{S}+1$. In the case of weak SO coupling and a spatially nondegenerate state ${}^{2S+1}\Gamma$ where Γ is a nondegenerate irreducible representation (irrep) of the system's point group, the SO coupling with the excited states splits the spin degeneracy. This ZFS splitting is on the order of a few cm^{-1} and gives the fine structure to the EPR spectra. In some cases, the ZFS splitting may be on the order of several ten cm^{-1}, and high-field high-frequency EPR (HF-HF EPR) is necessary to detect the transitions between these components. In this case, one usually takes $\mathcal{S} = S$. When the ground state is spatially degenerate or if there are very low-lying excited states, there can be large orbital contributions to the magnetic moment, and the choice of the spin Hamiltonian becomes more complicated and must be treated case by case.

For systems where SO coupling is strong, a close correspondence of a degenerate electronic state of interest with an *electron spin* multiplet may simply not exist. A corresponding QM method used to determine the electronic states may already include SO coupling in some form, possibly along with dipolar a spin–spin interaction term. In this case, the ZFS effects are included in the electronic spectrum. It is shown next how the pseudo-spin Hamiltonian parameters for the Zeeman and HFC interactions can be extracted from such QM calculations. For further discussion, see section "Higher-Order EPR Parameters and Mapping of Ab Initio to Pseudo-spin Functions."

When the SO coupling is large, the degeneracy of the states is related to symmetry. For odd-electron systems, the degeneracy is even due to Kramers' theorem. Fourfold degenerate irreps only appear in the cubic and icosahedral groups, along with sixfold in the latter. Therefore, except for highly symmetric molecules, the ground state of Kramers systems is modeled using $\mathcal{S} = 1/2$. For even-electron systems, if there is a high-order rotation axis, states may be doubly degenerate and $\mathcal{S} = 1/2$ represents a non-Kramers doublet. Only cubic and icosahedral groups may have higher degeneracies. Therefore, the states of even-electron systems with a heavy element are usually nondegenerate or, in the presence of symmetry, can be considered as non-Kramers doublets. In lanthanide and actinide complexes, the term ${}^{2S+1}L_J$ of the free ion is split due to the environment of the ligands. This splitting is usually on the order of some tens of cm^{-1} for lanthanides since the $4f$ orbitals are mostly inner shell and interact weakly with the environment. The splitting of the free ion term of an actinide is larger since the $5f$ orbitals interact more with the ligands, even forming covalent bonds for the early actinides; it can be on the order of several hundred cm^{-1}. Therefore, in the case of heavy elements, states are at the most doubly degenerate or nearly degenerate, and there are usually no EPR transitions with excited states, except for cubic systems.

An ab initio calculation provides the $2\mathcal{S}+1$ quasidegenerate wave functions $|\lambda, a\rangle_{a=1,2\mathcal{S}+1}$ in the absence of external magnetic field, defining the model space, and the corresponding energies E_a. Let

$$\hat{\mathcal{H}}^Z = -(-\beta_e \hat{\boldsymbol{\mu}}) \cdot \boldsymbol{B} \tag{9}$$

be the electronic Zeeman operator, with $\hat{\boldsymbol{\mu}}$ being a dimensionless time-odd QM operator that gives the electron magnetic moment operator upon multiplication with $-\beta_e$. The Zeeman interaction is characterized by the three matrices of the magnetization operator $(\boldsymbol{\mu}^k)_{a,b} = \langle \lambda, a|\hat{\mu}_k|\lambda, b\rangle$ with $a, b \in [1, 2\mathcal{S}+1]$, and $k = x, y, z$ being defined in the physical space. Hyperfine matrices can be defined analogously. These matrices are further discussed in sections "Higher-Order EPR Parameters and Mapping of Ab Initio to Pseudo-spin Functions" and "Signs of EPR g-Factors and Hyperfine Couplings."

The approach is outlined in this section for the Zeeman interaction and a doublet of Kramers states ($\mathcal{S} = 1/2$). In this case, there is no ZFS and the spin Hamiltonian reduces to the Zeeman term. In the basis of the pseudo-spin projection

eigenfunctions $|\mathscr{M}_S\rangle$, the operator $\hat{\mathcal{S}}_k$ ($k = x, y, z$) is represented by a matrix $\mathbf{S}_k = (1/2)\boldsymbol{\sigma}_k$, with $\boldsymbol{\sigma}_k$ being one of the 2×2 Pauli spin matrices. The magnetic field vector is expressed in terms of its components as $\boldsymbol{B} = (B_x, B_y, B_z)$. The matrix representation of the spin Hamiltonian for the Zeeman interaction then reads

$$\mathbf{H} = \sum_k h_k \mathbf{S}_k \quad \text{with} \quad h_k = \beta_e \sum_l B_l g_{lk} \tag{10}$$

The eigenvalues are easily obtained as $\pm$ the square roots of the eigenvalues of $\mathbf{H}^2$, which is diagonal because of $\mathbf{S}_k\mathbf{S}_l + \mathbf{S}_l\mathbf{S}_k = (\delta_{kl}/2)\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$:

$$\mathbf{H}^2 = \frac{1}{4}\begin{pmatrix} \sum_k h_k^2 & 0 \\ 0 & \sum_k h_k^2 \end{pmatrix} \tag{11}$$

Therefore, the energy difference ΔE for the two spin projections is $2[(1/4)\sum_k h_k^2]^{1/2} = [\sum_k h_k^2]^{1/2}$, i.e.,

$$\Delta E = \beta_e \Big[\sum_{k,l} B_k B_l \sum_m g_{km} g_{lm} \Big]^{1/2} = \beta_e \Big[\sum_{k,l} B_k B_l G_{kl} \Big]^{1/2} \tag{12}$$

In the previous equation, G_{kl} is an element of the tensor $\mathbf{G}$ introduced below Eq. (7).

Next, consider a quantum mechanical framework with a doublet state with two wavefunction components, ψ_1, ψ_2, assumed to be orthonormal for convenience. Further, for the time being, it is assumed that the doublet components ψ_1 and ψ_2 have the time-reversal properties of a Kramers pair. In the basis $\{\psi_1, \psi_2\}$, the QM Zeeman operator can also be expressed with the help of the spin-1/2 matrices, as

$$\mathbf{H}' = \sum_k h'_k \mathbf{S}_k \quad \text{with} \quad \begin{aligned} h'_x &= 2\beta_e \mathrm{Re}\langle\psi_2|\hat{\boldsymbol{\mu}}|\psi_1\rangle \cdot \boldsymbol{B} \\ h'_y &= 2\beta_e \mathrm{Im}\langle\psi_2|\hat{\boldsymbol{\mu}}|\psi_1\rangle \cdot \boldsymbol{B} \\ h'_z &= 2\beta_e \langle\psi_1|\hat{\boldsymbol{\mu}}|\psi_1\rangle \cdot \boldsymbol{B} \end{aligned} \tag{13}$$

Note that $\langle\psi_1|\hat{\boldsymbol{\mu}}|\psi_1\rangle \cdot \boldsymbol{B} = -\langle\psi_2|\hat{\boldsymbol{\mu}}|\psi_2\rangle \cdot \boldsymbol{B}$ because of the time-reversal symmetry. As with the pseudo-spin Hamiltonian, one can calculate twice the square root of the eigenvalues of $\mathbf{H}'^2$ to obtain the energy splitting in the presence of a magnetic field. The result can be rearranged as follows:

$$\Delta E = \Big[\sum_k {h'_k}^2 \Big]^{1/2} = \beta_e \Big[2 \sum_{k,l} B_k B_l \sum_{a=1}^{2} \sum_{b=1}^{2} \langle\psi_a|\hat{\mu}_k|\psi_b\rangle\langle\psi_b|\hat{\mu}_l|\psi_a\rangle \Big]^{1/2} \tag{14}$$

A factor 1/2 enters inside the square root when the double sum over a, b is introduced, to avoid double counting of contributions. By comparison of Eqs. (14) and (12), one finds for the elements of the tensor $\mathbf{G}$:

$$G_{kl} = 2\sum_{a=1}^{2}\sum_{b=1}^{2}\langle\psi_a|\hat{\mu}_k|\psi_b\rangle\langle\psi_b|\hat{\mu}_l|\psi_a\rangle \tag{15}$$

At this point, the assumption that ψ_1 and ψ_2 transform as a Kramers pair can be dropped, because any linear combination obtained from $\{\psi_1, \psi_2\}$ by unitary transformation gives the same tensor $\mathbf{G}$ from Eq. (15). Therefore, a computation of $\mathbf{G}$ can utilize a pair of doublet wavefunction components without imposing time-reversal symmetry explicitly. The reader is reminded that the definition of the magnetic moment operator components in Eq. (15) excludes pre-factors of β_e. As written, Eq. (15) assumes a complete one-particle basis set to represent ψ_1 and ψ_2, such that there is no dependence of the results on the gauge origin chosen for the external magnetic field. In calculations with finite basis sets, an origin dependence can be avoided by adopting a distributed gauge origin such as gauge-including atomic orbitals (GIAOs). When distributed origin methods are not available, calculations of magnetic properties of complexes with one paramagnetic metal center often place the metal center at the gauge origin.

The eigenvectors of $\mathbf{G}$ represent the molecule-fixed PAS of the Zeeman interaction, sometimes referred to as the "magnetic axes" of the system under consideration [10]. The square roots of the eigenvalues are absolute values of the principal g-factors. The signs of the g-factors are not obtained directly. For further discussion, see section "Signs of EPR g-Factors and Hyperfine Couplings."

The tensor $\mathbf{A}$ plays an analogous role for HFC as $\mathbf{G}$ plays for the Zeeman interaction. Therefore, after a QM operator $\hat{\boldsymbol{F}}_K$ has been defined for the hyperfine interaction as follows:

$$\hat{\mathcal{H}}^{\mathrm{HFC}} = \hat{\boldsymbol{F}}_K \cdot \boldsymbol{m}_K = g_K \beta_N \hat{\boldsymbol{F}}_K \cdot \boldsymbol{I}_K \tag{16}$$

the HFC tensor for a Kramers doublet can be calculated via

$$A_{kl} = 2(g_K\beta_N)^2\sum_{a=1}^{2}\sum_{b=1}^{2}\langle\psi_a|\hat{F}_{Kk}|\psi_b\rangle\langle\psi_b|\hat{F}_{Kl}|\psi_a\rangle \tag{17}$$

For hyperfine coupling, a natural choice for the gauge origin of the hyperfine field is the nucleus for which the HFC tensor is calculated.

A different route has been proposed [11] for calculating $\mathbf{G}$ for arbitrary values of $\mathcal{S}$, which is briefly discussed in section "Higher-Order EPR Parameters and Mapping of Ab Initio to Pseudo-spin Functions." The expression given in Eq. (45) is the same as (15) for $\mathcal{S} = 1/2$. As for the $\mathcal{S} = 1/2$ case, the expression for $\mathbf{G}$ of Eq. (45) should be adaptable for calculations of HFC.

Wavefunction-Based Methods for EPR Calculations

Popular starting points for wavefunction-based computations of EPR parameters are complete active-space self-consistent field (CASSCF) calculations and related restricted and generalized active-space approaches [12], often followed by perturbation theory (PT)-based treatments of dynamic correlation on top of CASSCF. For the latter, second-order perturbation theory (CASPT2) [13] and *n*-electron valence state perturbation theory (NEVPT) [14] are in relatively widespread use. Limitations arise from an insufficient description of *spin polarization* with the size of active spaces commonly achievable in these types of calculations, which is detrimental for HFC calculations. *g*-factors, ZFS parameters, and magnetic susceptibilities, on the other hand, can be obtained with good accuracy. Recently developed combinations of active-space methods with density matrix renormalization group (DMRG) techniques allow for larger active spaces, which is beneficial for treating electron correlation as well as spin polarization. "Proof of concept" calculations of HFC appear promising [15]. Linear response methods have also been developed for multi-configurational SCF wavefunctions in order to generate spin polarization suitable for HFC calculations without the need of very large active spaces [16]. In principle, multi-reference coupled-cluster (MRCC) methods should be suitable for EPR parameter calculations. To the authors' knowledge, relativistic MRCC calculations have not been used to predict EPR parameters at the time of writing this chapter.

Relativistic effects have been/can be included in wavefunction-based EPR calculations in a variety of ways, for instance: (i) by using all-electron SR Hamiltonians or SR effective core potentials (ECPs) to generate wavefunctions for a range of electronic states, followed by treatment of SO coupling via state-interaction (SI), [17] (ii) by including SR and SO effects either via an all-electron Hamiltonian [18] or with ECPs from the outset, and (iii) by calculating SR components of a spin multiplet and treatment of SO coupling as a perturbation in the EPR step [3]. In case (ii), SO effects are treated fully variationally, whereas in case (i), an SO Hamiltonian matrix is calculated in a limited basis of active-space wavefunctions and subsequently diagonalized. Approach (iii) is applicable in the weak SO coupling limit. Note that without application of specialized techniques, the use of a relativistic ECP for a given atom prevents calculations of the HFC for the same atom because the inner core nodal structure of the valence orbitals is needed. An order-by-order treatment of SO coupling via perturbation theory is also viable, for instance, based on four-component relativistic perturbation theory [19] after separation of SR and SO components of the QM operators.

Hartree-Fock and Kohn-Sham Methods for EPR Calculations

The approaches to obtaining EPR parameters outlined above assume that the wavefunction components of a degenerate state of interest are available explicitly

from a calculation. With single-reference methods such as Hartree-Fock (HF) theory and Kohn-Sham (KS) Density Functional Theory (DFT) and generalized KS methods, the usual approach in the absence of strong SO coupling is to start from a spin-unrestricted SR calculation. For brevity, HF theory is considered as a special case of a generalized KS hybrid functional from here on. The use of a spin-unrestricted single-determinant reference typically leads to spin contamination: while the spin-unrestricted KS reference can be designed as an eigenfunction of $\hat{S}_z$, it is not necessarily an eigenfunction of $\hat{S}^2$. Perdew et al. have pointed out that some degree of spin contamination is good because the KS reference is not the true wavefunction [20]. Spin polarization is generated straightforwardly in spin-unrestricted calculations but can be severely over- or underestimated. Single-reference KS methods with *approximate* functionals are often not suited to represent degenerate states. The calculation then results in breaking of spin or spatial symmetry of the KS wavefunction or both. Projection techniques can be used to restore lost symmetries.

In the absence of orbital degeneracy, the components of a pseudo-spin doublet can often be treated reasonably well by standard spin-unrestricted KS methods for the purpose of calculating EPR spin Hamiltonian parameters. In an SR or NR framework, the g-factors then simply become equal to g_e, while the isotropic average of the HFC matrix, the isotropic HFC constant, is calculated from averaging $g_K \beta_N \langle \psi | \hat{F}_{Kk} | \psi \rangle$ over k, with $|\psi\rangle$ being the $\mathscr{M}_{\mathcal{S}} = +1/2$ component of the doublet. There is also extensive literature on utilizing the same expectation value approach within single-reference correlated wavefunction methods. Extensions to treat cases with $\mathcal{S} > (1/2)$, and ways for additional inclusion of SO coupling via first-order perturbation theory, have been devised. The reader can find details in Refs. [21–27] and citations to original literature provided therein.

Some of the KS methods that are currently in use for relativistic EPR parameter calculations with SO coupling being included variationally [28–30] employ two different approaches. The first utilizes a variant of Eq. (13), but within a single-electron framework where the many-electron wavefunctions are replaced by one-electron orbitals. In the second approach, three separate SCF cycles are typically performed, with different quantization axes of spin, magnetic moment, or total angular momentum, and the quantization axis is identified with the directional index "k" of Eqs. (10) and (13).

The first approach [28, 30] as it was devised and implemented in a two-component relativistic form is quasi-spin restricted and limited to Kramers doublets. An SCF calculation is performed with the unpaired electron distributed over two degenerate frontier orbitals, with occupations of 1/2 each. In the absence of SO coupling, these would be an α and β spin pair of orbitals with identical spatial components. The method has some resemblance to restricted open-shell HF (ROHF) but is not the same. After the SCF step, one of these orbitals, say φ, is chosen to represent the component ϕ_1 of the Kramers pair of orbitals. Its conjugate ϕ_2 is then constructed from ϕ_1. Written explicitly in terms of real (R) and imaginary (I) parts of the two spin components of the SCF orbital φ, the Kramers pair is

$$\phi_1 = \begin{pmatrix} \varphi_\alpha^R \\ \varphi_\beta^R \end{pmatrix} + i \begin{pmatrix} \varphi_\alpha^I \\ \varphi_\beta^I \end{pmatrix} \tag{18a}$$

$$\phi_2 = \begin{pmatrix} -\varphi_\beta^R \\ \varphi_\alpha^R \end{pmatrix} + i \begin{pmatrix} \varphi_\beta^I \\ -\varphi_\alpha^I \end{pmatrix} \tag{18b}$$

For calculations of g-factors, the matrix elements of the Zeeman operator are then calculated as in Eq. (13), but with the orbital pair $\{\phi_1, \phi_2\}$, and then processed similar to Eqs. (14)–(15) to yield the g-factor. Alternatively, the h'_k of Eq. (13) are directly identified with the h_k terms of Eq. (10), and the g-matrix elements can be extracted from the calculation results without the detour via $\mathbf{G}$. HFC matrix elements can be calculated in an analogous way. However, due to the lack of spin polarization, the performance of the quasi-restricted approach is unsatisfactory for the latter. The performance for g-factors has frequently been satisfactory.

Regarding the "three SCF cycles" techniques, van Wüllen and coworkers have provided a justification for their use in KS calculations [31]. The approach of Ref. [30] is illustrated for HFC: the expectation value of the HFC part of the EPR spin Hamiltonian taken with a Kohn-Sham determinant calculated with a spin-quantization axis $\boldsymbol{u}$

$$E(\boldsymbol{u}) = \sum_i n_i \langle \varphi_i^{\boldsymbol{u}} | \boldsymbol{I}_K \cdot \mathbf{a} \cdot \hat{\boldsymbol{\mathcal{S}}} | \varphi_i^{\boldsymbol{u}} \rangle = \sum_{k,l} a_{kl} I_{Kk} \sum_i n_i \langle \varphi_i^{\boldsymbol{u}} | \hat{\mathcal{S}}_l | \varphi_i^{\boldsymbol{u}} \rangle \tag{19}$$

with $k, l \in \{x, y, z\}$. The $\varphi_i^{\boldsymbol{u}}$ are assumed to be orbitals obtained from a "generalized-collinear" KS calculation with selected spin-quantization axis $\boldsymbol{u}$, and the n_i are the occupation numbers. Assume next that $\boldsymbol{u}$ is along the Cartesian direction k, that the orbitals are $\hat{S}_k$ eigenfunctions, that the electron spin S is the same as the pseudo-spin $\mathcal{S}$, and that the KS determinant is a solution corresponding to $\langle S_k \rangle = M_S = S = \mathcal{S}$. One then finds $\sum_i n_i \langle \varphi_i^k | \hat{S}_l | \varphi_i^k \rangle = \mathcal{S} \delta_{kl}$, such that

$$E(k) = \mathcal{S} \sum_l a_{kl} I_{Kl} \tag{20}$$

Instead, calculate an analogous expectation value, but this time with the QM hyperfine operator $g_K \beta_N \hat{\boldsymbol{F}} \cdot \boldsymbol{I}_K$ and with the actual relativistic generalized-collinear two-component KS orbitals:

$$E(k) = g_K \beta_N \sum_l \sum_i n_i \langle \varphi_i^k | \hat{F}_{Kl} | \varphi_i^k \rangle I_{Kl} \tag{21}$$

One can now map the result (20) for the pseudo-spin onto the result (21) calculated by KS, which gives

$$a_{kl} = \frac{g_K \beta_N}{\mathcal{S}} \sum_i n_i \langle \varphi_i^k | \hat{F}_{Kl} | \varphi_i^k \rangle \tag{22}$$

An analogous approach is possible for calculations of g-factors, which gives with the QM Zeeman operator $\beta_e \hat{\boldsymbol{\mu}} \cdot \boldsymbol{B}$

$$g_{kl} = \frac{1}{\mathcal{S}} \sum_i n_i \langle \varphi_i^k | \hat{\mu}_l | \varphi_i^k \rangle \tag{23}$$

Both for the Zeeman and HFC matrices, one can form the rank-2 tensors $\mathbf{A}_K$ and $\mathbf{G}$ afterward and diagonalize them in order to obtain the PAS. As with variational wavefunction methods, GIAO basis sets are sometimes employed in order to generate Zeeman coupling matrices that are strictly origin invariant.

Within the generalized-collinear KS framework, it is also possible to obtain elements of the ZFS tensor $\mathbf{d}$, from the magnetic anisotropy of the KS energy with respect to the spin-quantization axis $\boldsymbol{u}$. With the pseudo-spin Hamiltonian and a $\mathscr{M}_\mathcal{S} = \mathcal{S}$ pseudo-spin eigenfunction, one obtains

$$E^{\mathrm{ZFS}}(\boldsymbol{u}) = \mathcal{S}(\mathcal{S} - \frac{1}{2})\, \boldsymbol{u} \cdot \mathbf{d} \cdot \boldsymbol{u} \tag{24}$$

As for the other parts of the EPR spin Hamiltonian, the result of a QM calculation of $E(\boldsymbol{u})$ for different directions of $\boldsymbol{u}$ can then be mapped onto Eq. (24). The approach was first introduced in Refs. [31,32], where the reader can find comments regarding some subtleties leading to the $\mathcal{S}(\mathcal{S} - 1/2)$ factor instead of $\mathcal{S}^2$. For weak SO coupling, $E(\boldsymbol{u})$ can also be calculated by perturbation theory. In this case, the "sum over states" (SOS) – as in Eq. (8) – can be interpreted as the result of a double perturbation of the energy by SO coupling and the dipolar ZFS interaction, and a KS coupled-perturbed analog can be devised instead. For details, see Ref. [31]. In cases where SO coupling dominates the ZFS, and for spin triplets, there is another KS route: starting with a closed-shell reference state, one calculates energy differences between the reference and a triplet state of interest by time-dependent linear response KS ("time-dependent DFT") within a framework that includes SO coupling variationally or as a perturbation.

Operators for the Zeeman and HFC Interactions

In principle, the Zeeman and hyperfine operators that are used in QM calculations of EPR parameters should match the Hamiltonian used for calculating the wavefunctions or KS orbitals in order to avoid picture-change errors. For further details, the reader is referred to the chapters in this handbook that are concerned with calculations of NMR parameters within various relativistic frameworks, because derivatives of the Zeeman and hyperfine operators with respect to the external field components and the nuclear spin magnetic moment components, respectively, are needed for those calculations. In order to render this chapter somewhat self-contained, for illustration, the Zeeman and hyperfine one-electron operators are provided here for the NR case, for the two-component zeroth-order regular approximation (ZORA),

and for the four-component case in its standard notation where diamagnetism is not explicit. For brevity, field-dependent two-electron operators are not listed.

Assuming point nuclei for the hyperfine terms, the gauge origin for the external field coinciding with the coordinate origin, and Coulomb gauge for the nuclear and external vector potential, the nonrelativistic Zeeman (Z) and HFC one-electron operators read

$$\text{NR}: \quad \hat{h}^Z = \frac{\beta_e}{\hbar}[(\boldsymbol{r} \times \hat{\boldsymbol{p}}) + \hbar\boldsymbol{\sigma}] \cdot \boldsymbol{B}$$

$$= \beta_e[\hat{\boldsymbol{L}} + 2\hat{\boldsymbol{S}}] \cdot \boldsymbol{B} \tag{25a}$$

$$\hat{h}_K^{\text{HFC}} = \frac{2\beta_e}{\hbar}\frac{\mu_0}{4\pi}[\frac{\boldsymbol{r}_K}{r_K^3} \times \hat{\boldsymbol{p}}] \cdot \boldsymbol{m}_K \tag{25b}$$

$$+ \beta_e \frac{\mu_0}{4\pi}[\boldsymbol{\sigma} \cdot \{\boldsymbol{m}_K(\nabla \cdot \frac{\boldsymbol{r}_K}{r_K^3}) - (\boldsymbol{m}_K \cdot \nabla)\frac{\boldsymbol{r}_K}{r_K^3}\}] \tag{25c}$$

Curly brackets, $\{\cdots\}$, in the operator expressions indicate that derivatives are only taken inside the operator, not of functions to its right-hand side. As elsewhere in this chapter, $\boldsymbol{m}_K = g_K\beta_N\boldsymbol{I}_K$. Further, $\boldsymbol{r}_K$ is the electron-nucleus distance vector and r_K its length. The Zeeman operator is a sum of contributions from orbital and spin angular momentum. Likewise, in the hyperfine operator, there is the "paramagnetic nuclear spin–electron orbital" (PSO) term in Eq. (25b) which is independent of the electron spin, and there is the electron spin dependent sum of the Fermi contact (FC) and spin dipole (SD) operators in Eq. (25c). The usual expressions for the FC and SD operators are obtained by taking the derivatives of $\boldsymbol{r}_K/r_K^3$, which gives

$$\text{NR}: \quad \hat{h}_K^{\text{FC}} = \beta_e \frac{\mu_0}{4\pi}\frac{8\pi}{3}\delta(\boldsymbol{r}_K)\,\boldsymbol{\sigma} \cdot \boldsymbol{m}_K \tag{26a}$$

$$\hat{h}_K^{\text{SD}} = \beta_e \frac{\mu_0}{4\pi}\frac{3(\boldsymbol{\sigma} \cdot \boldsymbol{r}_K)(\boldsymbol{m}_K \cdot \boldsymbol{r}_K)}{r_K^5} \tag{26b}$$

The "contact" part of the name of the FC operator refers to the presence of the Dirac δ-distribution.

Due to the fact that code for calculating matrix elements of these operators with Gaussian-type atomic orbital (AO) basis functions is rather widely available, nonrelativistic operators are sometimes used in relativistic calculations of EPR parameters. For the Zeeman operator, the relativistic corrections from the operator are likely small because it samples the valence and outer regions of light and heavy atoms. The hyperfine operators are to be used in relativistic calculations only with caution, because of the singular behavior evident from Eqs. (26a)–(26b). Due to their local nature, it is possible to use them for light nuclei in a system that also contains heavy elements, because then no significant relativistic effects are generated around the nucleus for which the HFC is calculated. It may also be possible to generate estimates of a heavy-element HFC if the relevant orbitals have

high angular momentum, and the HFC is dominated by the PSO mechanism. It is certainly *not* physically meaningful to use the nonrelativistic hyperfine operators in other relativistic scenarios such as HFC tensors of heavy alkali metal atoms or for small radicals containing mercury (see section "Selected Case Studies").

When adopting the two-component ZORA framework, the operators relevant for the Zeeman and HFC interactions read

$$\text{ZORA}: \quad \hat{h}^Z = \frac{\beta_e}{2\hbar}[\mathcal{K}(\boldsymbol{r}\times\hat{\boldsymbol{p}}) + (\boldsymbol{r}\times\hat{\boldsymbol{p}})\mathcal{K}]\cdot\boldsymbol{B} \tag{27a}$$

$$+ \frac{\beta_e}{2}\boldsymbol{\sigma}\cdot\left\{\boldsymbol{B}(\nabla\cdot\mathcal{K}\boldsymbol{r}) - (\boldsymbol{B}\cdot\nabla)\mathcal{K}\boldsymbol{r}\right\} \tag{27b}$$

$$\hat{h}_K^{\text{HFC}} = \frac{\beta_e}{\hbar}\frac{\mu_0}{4\pi}[\mathcal{K}\frac{\boldsymbol{r}_K}{r_K^3}\times\hat{\boldsymbol{p}} + \frac{\boldsymbol{r}_K}{r_K^3}\times\hat{\boldsymbol{p}}\mathcal{K}]\cdot\boldsymbol{m}_K \tag{27c}$$

$$+ \frac{\beta_e\mu_0}{4\pi}[\boldsymbol{\sigma}\cdot\{\boldsymbol{m}_K(\nabla\cdot\frac{\mathcal{K}\boldsymbol{r}_K}{r_K^3}) - (\boldsymbol{m}_K\cdot\nabla)\frac{\mathcal{K}\boldsymbol{r}_K}{r_K^3}\}] \tag{27d}$$

The function $\mathcal{K} = 2m_ec^2/(2m_ec^2 - V)$ is a "relativistic kinematic factor" that typically shows up in equations derived within the ZORA framework. Formally, the NR limit is obtained for $\mathcal{K} \to 1$. In this case, (27a) becomes the orbital Zeeman (OZ) operator, (27b) becomes spin Zeeman, (27c) becomes PSO, and (27d) becomes FC + SD. It therefore makes sense to adopt the same terminology with two-component methods such as ZORA, Douglas Kroll Hess (DKH), and other approximate or formally exact two-component methods that afford operators of similar structure. In the vicinity of heavy nuclei, $\mathcal{K}$ is very different from unity which generates the desired relativistic effects. It is noted that for point nuclei with a charge below 118, there is no "contact" term (i.e., a delta distribution) [33], because it is suppressed by $\mathcal{K} \to 0$ for $r_K \to 0$ in the operator.

The one-electron Zeeman and hyperfine operators in the four-component (Dirac) framework involve the 4×4 Dirac $\boldsymbol{\alpha}$ matrices:

$$\text{Dirac}: \quad \hat{h}^Z = \frac{ce}{2}\boldsymbol{\alpha}\times\boldsymbol{r}\cdot\boldsymbol{B} \tag{28}$$

$$\hat{h}_K^{\text{HFC}} = \frac{ce\mu_0}{4\pi}\boldsymbol{\alpha}\times\frac{\boldsymbol{r}_K}{r_K^3}\cdot\boldsymbol{m}_K \tag{29}$$

Unlike the NR and two-component versions, the operators do not explicitly include derivative terms. However, the derivative terms are implicitly contained in the formalism because of the relation between the large (upper) and small (lower) components of the electronic wavefunctions or orbitals.

For HFC that is nominated by *s* orbitals (heavy alkali metals and Hg in particular), finite nucleus effects can be large. There are different ways to treat finite nuclear volume effects [34]. Due to the ubiquity of Gaussian-type basis functions in quantum chemical calculations, the spherical Gaussian nuclear model

is in widespread use. Here, the charge distribution ρ_K of nucleus A is "smeared out" by a Gaussian function as

$$\rho_K(\boldsymbol{R}) = Z_K \left(\frac{\xi_K}{\pi}\right)^{3/2} e^{-\xi_K |\boldsymbol{R}-\boldsymbol{R}_K|^2} \tag{30}$$

The exponent ζ_K is inversely proportional to the mean square radius $\langle R_K^2 \rangle$ of the nucleus:

$$\zeta_K = \frac{3}{2\langle R_K^2 \rangle} \tag{31}$$

The root mean square nuclear radius is, in turn, related to the nuclear mass M_K (in amu) as follows:

$$\langle R_K^2 \rangle^{1/2} = (0.863 M_K^{1/3} + 0.570)\ \text{fm} \tag{32}$$

The electron–nucleus attraction term for nucleus K with charge Z_K in the Hamiltonian for point nuclei

$$V_K^{\text{point}} = -\frac{e^2}{4\pi\varepsilon_0}\frac{Z_K}{r_K} \tag{33a}$$

changes to

$$V_K^{\text{gauss.}} = -\frac{e^2}{4\pi\varepsilon_0}\frac{Z_K}{r_K} P(1/2, \tilde{r}_K^2) \tag{33b}$$

with $\tilde{r}_K^2 = \zeta_K r_K^2$. Further,

$$P(a,x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt \tag{34}$$

is the lower incomplete Gamma function ratio, and $\Gamma(a)$ is the gamma function. Assuming as a first approximation that the magnetization density of the nucleus can also be described by a spherical Gaussian, the vector potential for a point nucleus

$$\boldsymbol{A}_K^{\text{point}} = \frac{\mu_0}{4\pi}\frac{\boldsymbol{m}_K \times \boldsymbol{r}_K}{r_K^3} \tag{35a}$$

changes to

$$\boldsymbol{A}_K^{\text{gauss.}} = \frac{\mu_0}{4\pi}\frac{\boldsymbol{m}_K \times \boldsymbol{r}_K}{r_K^3} P(3/2, \tilde{r}_K^2) \tag{35b}$$

The presence of the incomplete Gamma function terms in the expressions serves to dampen the inverse powers of r_K such that the resulting potential and vector potential remain finite as $r_K \to 0$. In calculations, there are two effects: The first one is via the potential (33b) and affects the electron spin and orbital magnetizations around the nucleus. The second one is the modification of the hyperfine operators by (35b). In combination, they tend to reduce the magnitude of hyperfine coupling constants.

Higher-Order EPR Parameters and Mapping of Ab Initio to Pseudo-spin Functions

This section focuses on ZFS and the Zeeman interaction as examples. The HFC can be treated in an analogous fashion as the Zeeman interaction as far as higher-order pseudo-spin Hamiltonian terms are concerned. For a unified formalism and examples, see Ref. [35].

The larger the pseudo-spin $\mathcal{S}$, the more degrees of freedom there are. Higher-order spin operators are then added to the spin Hamiltonian to describe the supplementary degrees of freedom. Higher orders include terms with polynomials of order l, m, and n in the components of $\hat{\boldsymbol{\mathcal{S}}}$, $\boldsymbol{B}$, and $\boldsymbol{I}$ respectively, symbolically denoted here as a term of order $\mathcal{S}^l B^m I^n$ where l, m, and n are nonnegative integers and $l + m + n$ is even to preserve time even parity of the Hamiltonian. An exception concerns the description of non-Kramers doublets. This point is presented in section "Selected Case Studies." The expansion is limited to $l \leq 2\mathcal{S}$ since all matrix elements of the operators with $l > 2\mathcal{S}$ are zero. The ZFS term corresponds to $m = n = 0$

$$\hat{\mathcal{H}}_S^{\mathrm{ZFS}} = \hat{\mathcal{H}}_{(2)}^{\mathrm{ZFS}} + \hat{\mathcal{H}}_{(4)}^{\mathrm{ZFS}} + \cdots \tag{36}$$

where $\hat{\mathcal{H}}_{(l)}^{\mathrm{ZFS}}$ is a term of order l even in $\mathcal{S}$. The term linear in the magnetic field, with $m = 1$ and $n = 0$, is the Zeeman term

$$\begin{aligned} \hat{\mathcal{H}}_S^Z &= \hat{\mathcal{H}}_{(1)}^Z + \hat{\mathcal{H}}_{(3)}^Z + \cdots \\ &= \beta_e(\hat{\boldsymbol{\mu}}_{(1)} + \hat{\boldsymbol{\mu}}_{(3)} + \cdots) \cdot \boldsymbol{B} \end{aligned} \tag{37}$$

where $\hat{\boldsymbol{\mu}}_{(l)}$ is a term of order l odd in $\mathcal{S}$. The next term with $m = 2$ describes the quadratic contribution in the magnetic field. This term is usually negligible due to the smallness of the magnetic interaction [36].

According to the irreducible tensor operator decomposition, the preceding terms can be written as

$$\hat{\mathcal{H}}_{(l)}^{\mathrm{ZFS}} = \sum_{m=-l}^{l} a_{l,m} \mathcal{T}_{l,m}(\hat{\boldsymbol{\mathcal{S}}}) \tag{38}$$

and

$$\hat{\mu}^u_{(l)} = \sum_{m=-l}^{l} b^u_{l,m} \mathcal{T}_{l,m}(\hat{\boldsymbol{\mathcal{S}}}) \tag{39}$$

where $\hat{\mu}^u_{(l)}$ is the component of $\hat{\boldsymbol{\mu}}_l$ in direction $\boldsymbol{u}$ and $\mathcal{T}_{lm}$ are the tesseral combinations of the spherical-tensor operators $T_{l,m}$

$$\begin{aligned} \mathcal{T}_{l,m}(\hat{\boldsymbol{\mathcal{S}}}) &= \tfrac{1}{\sqrt{2}}[(-1)^m T_{l,m}(\hat{\boldsymbol{\mathcal{S}}}) + T_{l,-m}(\hat{\boldsymbol{\mathcal{S}}})] \\ \mathcal{T}_{l,-m}(\hat{\boldsymbol{\mathcal{S}}}) &= \tfrac{i}{\sqrt{2}}[(-1)^{m+1} T_{l,m}(\hat{\boldsymbol{\mathcal{S}}}) + T_{l,-m}(\hat{\boldsymbol{\mathcal{S}}})] \end{aligned} \tag{40}$$

with $0 \leq m \leq l$. Equation (39) becomes for $l = 1$

$$\hat{\mu}^u_{(1)} = b^u_{1,1}\hat{\mathcal{S}}_x + b^u_{1,-1}\hat{\mathcal{S}}_y + b^u_{1,0}\hat{\mathcal{S}}_z \tag{41}$$

with $u = x, y, z$. This defines nine parameters $b^u_{1,m}$ corresponding to the elements of the $\mathbf{g}$ matrix. Equation (41) corresponds to the first term of Eq. (6) and appears for all values of $\mathcal{S} \geq 1/2$. The third term of Eq. (6) appears for $\mathcal{S} \geq 1$

$$\begin{aligned} \hat{\boldsymbol{\mathcal{H}}}^{\mathrm{ZFS}}_{(2)} = {} & a_{2,2}\tfrac{1}{\sqrt{2}}(\hat{\mathcal{S}}^2_x - \hat{\mathcal{S}}^2_y) + a_{2,-2}\tfrac{1}{\sqrt{2}}(\hat{\mathcal{S}}_x\hat{\mathcal{S}}_y + \hat{\mathcal{S}}_y\hat{\mathcal{S}}_x) + a_{2,1}\tfrac{1}{\sqrt{2}}(\hat{\mathcal{S}}_x\hat{\mathcal{S}}_z + \hat{\mathcal{S}}_z\hat{\mathcal{S}}_x) \\ & + a_{2,-1}\tfrac{1}{\sqrt{2}}(\hat{\mathcal{S}}_y\hat{\mathcal{S}}_z + \hat{\mathcal{S}}_z\hat{\mathcal{S}}_y) + a_{2,0}\tfrac{1}{\sqrt{6}}(2\hat{\mathcal{S}}^2_z - \hat{\mathcal{S}}^2_x - \hat{\mathcal{S}}^2_y) \end{aligned} \tag{42}$$

The five parameters $a_{2,m}$ define the symmetric and traceless $\mathbf{d}$ tensor. The cubic term in $\mathcal{S}$ contributes to the Zeeman interaction for $\mathcal{S} \geq 3/2$:

$$\hat{\mu}^u_{(3)} = \sum_{m=-3}^{3} b^u_{3,m} \mathcal{T}_{3,m}(\hat{\boldsymbol{\mathcal{S}}}) \tag{43}$$

All $\mathcal{T}_{3,m}(\hat{\boldsymbol{\mathcal{S}}})$ can be expressed as a product of three spin components $\hat{\mathcal{S}}_u$ defining a third-rank tensor $\mathbf{g}'$.

$$\hat{\mathcal{H}}^Z_{(3)} = \beta_e \boldsymbol{B} \cdot \mathbf{g}' \cdot \hat{\boldsymbol{\mathcal{S}}}^3 \tag{44}$$

where $\mathbf{g}'$ is a third-rank tensor.

In the case of weak SO coupling, the $|\lambda, a\rangle$ wavefunctions correspond closely to the SR components $|\lambda', a\rangle$. Without SO coupling, the $|\lambda', a\rangle$ functions are the $2S + 1$ spin components of the real spin and degenerate. These $|\lambda', M\rangle$ behave properly under all spin operations, time inversion, and spatial symmetries of the molecule and can be assigned to the pseudo-spin states $|\lambda', M_S\rangle \equiv |\mathcal{S}, \mathcal{M}_\mathcal{S}\rangle$. As it has been shown in section "Representation of the Pseudo-spin Hamiltonian in an Ab Initio Framework," the spin Hamiltonian parameters can be calculated in this case by a perturbative approach using wavefunctions which do not include SO effects.

In the case where SO coupling is included in the QM calculation, the assignment of the $|\lambda, a\rangle$ functions to pseudo-spin functions becomes more difficult. There are currently two types of methods to calculate the spin Hamiltonian parameters from the E_a and the three matrices $\boldsymbol{\mu}^k$ of the electron magnetic moment (see Eq. (9)) represented in the basis of $|\lambda, a\rangle$ calculated by the ab initio methods (i) either by projecting the Zeeman matrices using the irreducible tensor operators algebra [10], since these latter operators form a basis of orthogonal and linearly independent matrices, each matrix has a unique expansion in this basis set, (ii) or by mapping the matrix elements of the real and pseudo-spin matrices one by one once the correspondence between the real and pseudo-spin states is performed.

For the projection technique, one considers first that the term linear in $\mathcal{S}$ is the dominant one in the $\hat{\mathcal{H}}_S^Z$ operator, and the tensor $\mathbf{G} = \mathbf{g}\,\mathbf{g}^T$ can be calculated as [11]

$$\mathbf{G} = \frac{6}{\mathcal{S}(\mathcal{S}+1)(2\mathcal{S}+1)}\mathcal{A} \tag{45}$$

where the tensor $\mathcal{A}$ (not to be confused with $\mathbf{A}$ of Eq. (17)) is defined as

$$\mathcal{A}_{k,l} = \frac{1}{2}\,\mathrm{tr}(\boldsymbol{\mu}^k\,\boldsymbol{\mu}^l) \tag{46}$$

The diagonalization of $\mathbf{G}$ provides the absolute values of the g factors $g_i = \pm\sqrt{G_i}$ with $i = X, Y, Z$. Equation (45) is equivalent to Eq. (15) in the case $\mathcal{S} = 1/2$.

The matching technique consists in rotating the three matrices $\boldsymbol{\mu}^k$ $(k = x, y, z)$

(i) In the Euclidean space of spatial coordinates

$$\boldsymbol{\mu}^{l'} = \sum_k \mathbf{R}_{k,l'}\boldsymbol{\mu}^k \quad \text{with} \quad k = x, y, z \quad \text{and} \quad l' = x', y', z' \tag{47}$$

where $\mathbf{R}$ is a 3×3 rotation matrix of the Cartesian coordinates.

(ii) In the Hilbert space generated by $|\lambda, a\rangle$ $(a = 1, 2\mathcal{S}+1)$.

$$(\boldsymbol{\mu}^k)' = \mathcal{R}^\dagger \cdot \boldsymbol{\mu}^k \cdot \mathcal{R} \quad \text{for all} \quad k = x, y, z \tag{48}$$

where $\mathcal{R}$ is a rotation in the $(1 + 2\mathcal{S})^2$ Hilbert space.

These rotations are performed in order to put the three matrices $\boldsymbol{\mu}^k$ $(k = x, y, z)$ to suit the matrices of the spin Hamiltonian. The rotations in coordinate space may rotate the real space in the principal axis of the $\mathbf{D}$ or $\mathcal{A}$, for example. Then the rotations in the Hilbert space of wavefunctions may diagonalize the $\boldsymbol{\mu}^Z$ matrix, make $\boldsymbol{\mu}^X$ real, and $\boldsymbol{\mu}^Y$ imaginary. No information is lost during these transformations. In the final form, one can fit the spin Hamiltonian parameters on the matrix elements of the $\boldsymbol{\mu}'^k$ $(k = x', y', z')$ matrices. The deviation through the fitting procedure can be evaluated and scores the propensity of the model to reproduce the ab initio data. But this procedure needs some symmetry in order to find the proper rotations.

For the calculation of the ZFS parameters, one needs the assignment of the combination of the $|\lambda, a\rangle$ to the pseudo-spin states $|\mathcal{S}, \mathcal{M}_\mathcal{S}\rangle$. One must find a rotation $\mathcal{R}$ in the model space such the transformed wavefunctions fulfill the time inversion properties $\hat{\Theta}|\lambda, \mathcal{M}_\mathcal{S}\rangle = (\pm 1)^{(\mathcal{S}-\mathcal{M}_\mathcal{S})}|\lambda, -\mathcal{M}_\mathcal{S}\rangle$. The real Hamiltonian, including the interactions attributed to the ZFS, is diagonal in the basis of the $|\lambda, a\rangle$ if SO coupling and the dipolar two-electron spin–spin interaction are included variationally in the ground state. In the new basis set, it becomes

$$\mathbf{H}^{\mathrm{ZFS}} = \mathcal{R}^{\dagger} \cdot \mathbf{E} \cdot \mathcal{R} \tag{49}$$

where $\mathbf{E}$ is the diagonal matrix with $\mathbf{E}_{a,b} = E_a \delta_{a,b}$. Then, these matrix elements match the matrix elements of the pseudo-spin matrix $\mathbf{H}^{\mathrm{ZFS}}_{\mathcal{M}_\mathcal{S}, \mathcal{M}'_\mathcal{S}} = \langle \mathcal{S}, \mathcal{M}_\mathcal{S} | \hat{\mathcal{H}}^{\mathrm{ZFS}}_S | \mathcal{S}, \mathcal{M}_\mathcal{S} \rangle$

If SO coupling is weak, the $|\lambda, a\rangle$ derives mostly of a pure spin state ${}^{2S+1}\Gamma$ with components $|\lambda', M\rangle$ where M is the projection on the quantification axis of the real spin. This assignment is easily performed using the effective Hamiltonian technique [37] briefly outlined in section "Representation of the Pseudo-spin Hamiltonian in an Ab Initio Framework."

$$\mathcal{P}|\lambda, a\rangle = \sum_M C_{i,M} |\lambda', M\rangle \tag{50}$$

where $\mathcal{P}$ is the projector on the pure spin space. The effective Hamiltonian in this target space is

$$\mathbf{H}^{\mathrm{ZFS}}_{\mathrm{eff}} = \mathbf{C}^{-1} \cdot \mathbf{E} \cdot \mathbf{C} \tag{51}$$

$C_{i,M}$ is not an orthogonal matrix since it is a projector. The wavefunctions $|\lambda', M\rangle$ satisfy all the properties of transformations of spin with time inversion and spin operators since they are eigenfunctions for a real spin. Equations (49) and (51) are closely related. In the case of weak SO coupling, $\mathbf{C}$ is close to being orthogonal, and the effective Hamiltonian procedure is a convenient way to obtain the rotation $\mathcal{R}$, eventually using an orthogonalization procedure. This procedure is applicable for all values of $\mathcal{S}$, while for large values of $\mathcal{S}$, the determination of $\mathcal{R}$ becomes more complicated due to the increase of the number of degrees of freedom.

The spin Hamiltonian is designed in order to reduce as much as possible the number of parameters to fit the EPR spectra. In the case of high $\mathcal{S}$, many of the spin Hamiltonian parameters are negligible. In the case of weak SO coupling, the Zeeman interaction is almost Isotropic and the magnetic anisotropy arises from the ZFS term. The determination of the ZFS tensor is then the key step of the fitting.

In the case of $\mathcal{S} = 1$, without any spatial symmetry, the three $|\lambda, a\rangle$ are not degenerate and are not magnetic to first-order $\langle \lambda, a | \hat{\boldsymbol{\mu}}_u | \lambda, a \rangle = 0$. Magnetic properties arise from the off-diagonal matrix elements. The spin Hamiltonian

expressed in the principal axis of the **d** tensor with an isotropic Zeeman interaction writes

$$\begin{aligned}\hat{\mathcal{H}}_S &= \beta_e g \boldsymbol{B} \cdot \hat{\boldsymbol{\mathcal{S}}} + d_X \hat{\mathcal{S}}_X^2 + d_Y \hat{\mathcal{S}}_Y^2 + d_Z \hat{\mathcal{S}}_Z^2 \\ &= \beta_e g \boldsymbol{B} \cdot \hat{\boldsymbol{\mathcal{S}}} + D\left(\hat{\mathcal{S}}_Z^2 - \tfrac{1}{3}\mathcal{S}(\mathcal{S}+1)\right) + E(\hat{\mathcal{S}}_X^2 - \hat{\mathcal{S}}_Y^2)\end{aligned} \tag{52}$$

In the basis set $\left\{|0_X\rangle = \frac{1}{\sqrt{2}}(-|1,1\rangle + |1,-1\rangle), |0_Y\rangle = \frac{i}{\sqrt{2}}(|1,1\rangle + |1,-1\rangle), |0_Z\rangle = |1,0\rangle\right\}$ where $|0_k\rangle$ is the spin state with $\mathscr{M}_\mathcal{S} = 0$ for direction k

$\hat{\mathcal{H}}_S$	$\lvert 0_X\rangle$	$\lvert 0_Y\rangle$	$\lvert 0_Z\rangle$
$\langle 0_X\rvert$	$\frac{1}{3}D - E$	$-i\beta_e g B_Z$	$i\beta_e g B_Y$
$\langle 0_Y\rvert$	$i\beta_e g B_Z$	$\frac{1}{3}D + E$	$-i\beta_e g B_X$
$\langle 0_Z\rvert$	$-i\beta_e g B_Y$	$i\beta_e g B_X$	$-\frac{2}{3}D$

(53)

Figure 1 represents the variation of the energy of the three states as a function of B for the three directions X, Y, and Z for $g = 2$, $D = 10\,\mathrm{cm}^{-1}$, and $E = 1\,\mathrm{cm}^{-1}$. The largest magnetization (slope of the $E = f(B)$ curve) is obtained for $|0_X\rangle$ and

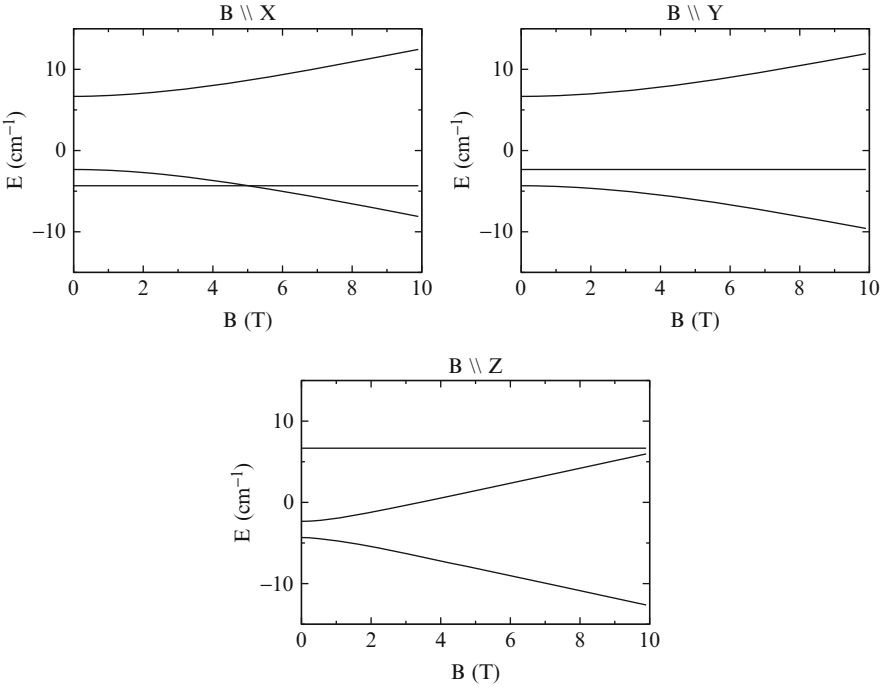


Fig. 1 Energy versus magnetic field for $\mathcal{S} = 1$ as solution of Eq. (53) for $g = 2$, $D = -10\,\mathrm{cm}^{-1}$ and $E = 1\,\mathrm{cm}^{-1}$

$|0_Y\rangle$ when the field is applied along Z since $|0_X\rangle$ and $|0_Y\rangle$ are the closest, and the Zeeman interaction couples them in this direction. When $D < 0$, these two states are lower than $|0_Z\rangle$, and the ground state is magnetic along Z; the smaller the E is, larger the magnetic interaction between these states is. This corresponds to an axial magnetization in direction Z.

The spin Hamiltonian of Eq. (52) is widely used for $\mathcal{S} = 3/2$ systems as well. It is shown in section "Selected Case Studies" that the cubic term in $\mathcal{S}$ is negligible in the case of weak SO coupling. In the basis of the $|\mathcal{S}, \mathcal{M}_\mathcal{S}\rangle$, its matrix is

$\hat{\mathcal{H}}_S$	$\|3/2\rangle$	$\|1/2\rangle$	$\|-1/2\rangle$	$\|-3/2\rangle$
$\langle 3/2\|$	$D + \frac{3}{2}\beta_e g B_Z$	$\frac{\sqrt{3}}{2}\beta_e g\,(B_X - iB_Y)$	$\sqrt{3}E$	0
$\langle 1/2\|$	$\frac{\sqrt{3}}{2}\beta_e g\,(B_X + iB_Y)$	$-D + \frac{1}{2}\beta_e g B_Z$	$\beta_e g\,(B_X - iB_Y)$	$\sqrt{3}E$
$\langle -1/2\|$	$\sqrt{3}E$	$\beta_e g\,(B_X + iB_Y)$	$-D - \frac{1}{2}\beta_e g B_Z$	$\frac{\sqrt{3}}{2}\beta_e g\,(B_X - iB_Y)$
$\langle -3/2\|$	0	$\sqrt{3}E$	$\frac{\sqrt{3}}{2}\beta_e g\,(B_X + iB_Y)$	$D - \frac{3}{2}\beta_e g B_Z$

(54)

The energies from diagonalization of this spin Hamiltonian are plotted as functions of B for $g = 2$, $D = -10\,\mathrm{cm}^{-1}$, and $E = 1\,\mathrm{cm}^{-1}$ in Fig. 2. The two Kramers doublets $|\pm 1/2\rangle$ and $|\pm 3/2\rangle$ are split by an energy $2\sqrt{D^2 + 3E^2}$. With $D < 0$,

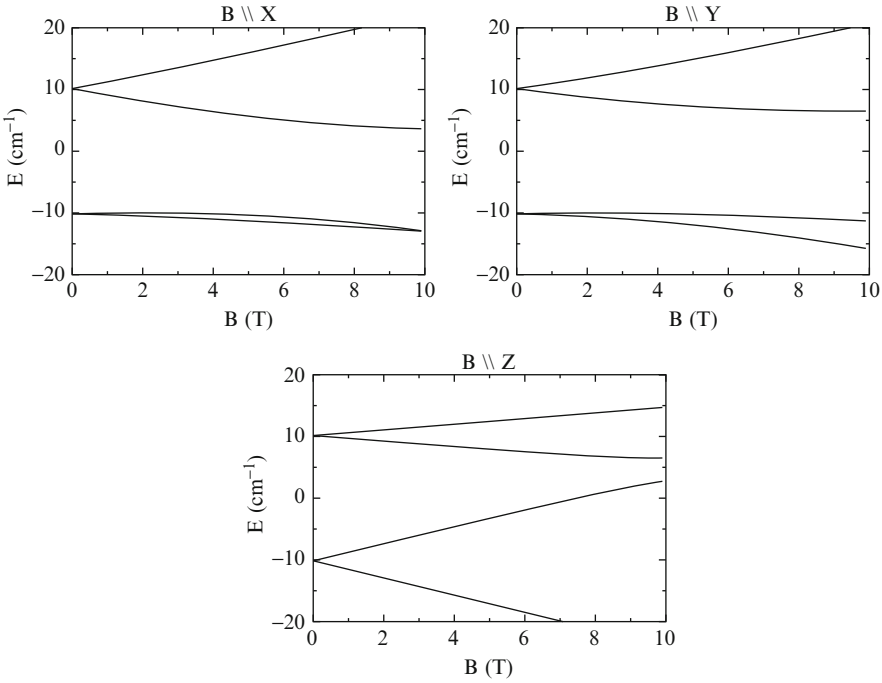


Fig. 2 Energy versus magnetic field for $\mathcal{S} = 3/2$ as solution of Eq. (54) for $g = 2$, $D = -10\,\mathrm{cm}^{-1}$ and $E = 1\,\mathrm{cm}^{-1}$

the $|\pm 3/2\rangle$ doublet is the lowest. When $E = 0$, the $|\pm 3/2\rangle$ doublet is purely axial along Z with a magnetization of $3/2g$ while the $|\pm 1/2\rangle$ doublet has a magnetization $1/2g$ along Z and g along X and Y. The E parameter induces some mixing of the $|\pm 1/2\rangle$ component in the ground state at zero field and couples the two doublets through the Zeeman interaction. The magnetization becomes less axial. When D is large, the second Kramers doublet may not be detected by EPR, even with HF-HF EPR. In this case, the ground Kramers doublet can be modeled with a restricted model space with $\mathcal{S} = 1/2$; the spin Hamiltonian is then pure Zeeman, and the g matrix is purely axial $g_Z = 3g$ and $g_X = g_Y = 0$ for $D < 0$ and $g_Z = g$ and $g_X = g_Y = 2g$ for $D > 0$.

Most cases with large values of $\mathcal{S}$ are in the weak SO coupling limit with $\mathcal{S}$ close to the real spin S of the spin-free state. When there is no very low-lying state, the orbital contribution arises through second-order coupling with the excited states, the Zeeman interaction is almost isotropic, and the anisotropic behavior of the magnetization arises from the **D** tensor. In this case, whatever the value of $\mathcal{S}$ is, the effective Hamiltonian technique permits a simple assignment between the "real" wavefunctions to the pseudo-spin components.

Signs of EPR *g*-Factors and Hyperfine Couplings

In the Zeeman term of Eq. (6), the matrix **g** links the pseudo-spin operator $\hat{\boldsymbol{\mathcal{S}}}$, which acts in the spin space, with the magnetic field, which acts in the physical space. Rotations in each of the two spaces are a priori disconnected, and consequently, the matrix **g** is not a tensor. It is rather arbitrary since any rotation in the spin space affects the **g** matrix but gives the same electronic magnetic moment $\boldsymbol{m}_e$, which is the physical observable coupling to the magnetic field. The g-factors are calculated as the square roots of the principal values of the **G** tensor of Eq. (15), which therefore determines the absolute values of the g-factors but does not provide any information about their sign. Experimentally, the g-factors are deduced from Eq. (7). Conventional EPR does not provide the sign of the g-factors. In the weak SO limit, they are close to $g_e \simeq 2$ and they are positive.

Let X, Y, Z denote the magnetic axes of a system. Pryce has shown [38] that the sign of the product $g_X g_Y g_Z$ determines the direction of the precession of the magnetic moment around the magnetic field. Experimentally, this sign has been measured for octahedral compounds where the three factors are identical. Negative signs were deduced relative to the sign of the hyperfine coupling [39] for $PaCl_6^{2-}$ as well as [40] for UF_6^-. The sign of $g_X g_Y g_Z$ was found to be negative as well for $[NpO_2(NO_3)_3]^-$ [41].

The sign of $g_X g_Y g_Z$ defines the sign of the Berry phase of a pseudo-spin applied in an applied magnetic field [42]. In the case of $\mathcal{S} = 1/2$,

$$\boldsymbol{\mu}^X \boldsymbol{\mu}^Y \boldsymbol{\mu}^Z = i \; g_X g_Y g_Z \mathbf{I}_2 \tag{55}$$

where $\boldsymbol{\mu}^u$ represents the electron magnetic moment $\hat{\mu}_u$ in the basis of the doublet-state components, as defined in section "Representation of the Pseudo-spin Hamiltonian in an Ab Initio Framework" and Eq. (9), and $\mathbf{I}_2$ is the 2×2 identity matrix. This product of matrices is invariant by rotation in the Hilbert spin space and therefore does not need any assignment between the two physical wavefunctions with the $|\mathcal{S}, \mathcal{M}_\mathcal{S}\rangle$ pseudo-spin eigenfunctions. Equation (55) is easily calculated ab initio and gives access to the sign of the product of the three g-factors.

In the case of symmetric molecules, the degree of arbitrariness of the matrix $\mathbf{g}$ can be reduced by imposing symmetry constraints on the pseudo-spin. More specifically, the pseudo-spin may be required to behave under spatial rotations as a spin operator, up to a multiplicative function [43]. The two physical kets $|\lambda, 1\rangle$ and $|\lambda, 2\rangle$ span an irrep Γ of the point group of the molecule, while the two components $|\alpha\rangle$ and $|\beta\rangle$ of a spin $S = 1/2$ span an irrep Γ_S. If one can find a real scalar function ϕ such

$$\begin{aligned} |\lambda, 1\rangle &= \phi\ |\alpha\rangle \\ |\lambda, 2\rangle &= \hat{\theta}|\lambda, 1\rangle = \phi\ |\beta\rangle \end{aligned} \tag{56}$$

where $\hat{\theta}$ is the time-reversal operator and two $|\lambda, 1\rangle$ and $|\lambda, 2\rangle$ are properly defined for the rotations of the molecule. This implies that $\Gamma = \Gamma_\phi \otimes \Gamma_S$ with Γ_ϕ being the irrep of ϕ. Since the pseudo-spin and the multiplet are supposed to have the same degeneracy, Γ_ϕ must be a one-dimensional symmetry species. For example, in the case of the octahedral AnX_6^{q-} complexes with $5f^1$ configuration, the ground state is of symmetry $E_{5/2u}$ and $\Gamma_S = E_{1/2g}$. It follows that $\Gamma_\phi = A_{2u}$ since $E_{5/2u} = A_{2u} \otimes E_{1/2g}$: the decomposition of Eq. (56) is uniquely defined and one can determine the signs of the g-factors [44]. By symmetry, the three principal g-factors are equal and their sign is equal to the sign of the product.

The case of the neptunyl ion, NpO_2^{2+}, is different [45]. The free ion is linear and has a nonzero principal $g_\parallel$-factor in the direction $\parallel$, parallel to the molecular axis. The components $g_\perp$ perpendicular to the axis are zero, however. This means that $g_X g_Y g_Z = 0$, i.e., this product conveys no information about the sign of $g_\parallel$. In $\mathcal{D}_{\infty h}$, the ground state of neptunyl is $E_{5/2u}$ and $\Gamma_S = E_{1/2g}$. There is, however, no one-dimensional irrep satisfying $E_{5/2u} = \Gamma \otimes E_{1/2g}$, and therefore, the decomposition as in Eq. (56) is not possible. With equatorial ligands, the symmetry of the neptunyl is lowered, either to $\mathcal{D}_{3h}$, e.g., in $[NpO_2(NO_3)_3]^-$, or to $\mathcal{D}_{4h}$, e.g., in $[NpO_2Cl_4]^{2-}$. In both cases, the two equatorial $g_\perp$-factors are equal, and $g_\parallel$ has the sign of the product of the three g-factors. This sign is experimentally negative for the first complex [41]. According to Eq. (55), ab initio calculations give a negative sign for the nitrate complex but a positive sign for the chloride. The ground state of $[NpO_2(NO_3)_3]^-$ is of symmetry $E_{1/2}$ and in the $\mathcal{D}_{3h}$ double group $\Gamma_S = E_{1/2}$. The scalar function ϕ belongs either to A_1' or to A_2' since $E_{1/2} = A'_{1(2)} \otimes E_{1/2}$. The decomposition of Eq. (56) leads in both cases to the same negative sign of $g_\parallel < 0$, but one of the solutions gives $g_\perp > 0$ while the other one gives $g_\perp < 0$. In this case, the use of symmetry arguments does not produce a unique sign of $g_\perp$. In the

same way, the ground state of $[NpO_2Cl_4]^{2-}$ is of symmetry $E_{3/2u}$, $\Gamma_S = E_{1/2g}$, and ϕ belongs to either B_{1u} or B_{2u}. Both solutions give $g_\parallel > 0$ but one gives $g_\perp > 0$, while the other one gives $g_\perp < 0$. Therefore, the individual signs of the g-factors of these neptunyl complexes cannot be determined. It has been proposed [43] that the sign of g-factors in the case of an arbitrarily distorted complex could be determined by considering an adiabatic distortion of the complex toward a symmetric system for which the signs are well defined. But the analysis above shows that even in symmetrical cases, the signs of individual g-factors may not be unique.

While the sign of the product of the three g-factors can be related to an observable, namely, the sense of the precession of the magnetic moment around the magnetic field, it appears that it is not possible in general to determine a specific sign of each individual g-factor, even in the case of molecules with high symmetry. Anyhow, the decomposition of Eq. (56) permits constructing a set of doublet components which behave as the components of a spin under the symmetry operations of a molecule.

Selected Case Studies

In this section, selected examples are presented where EPR parameters have been calculated with relativistic two-step complete active-space (CAS) wavefunction methods (treating SO coupling by state interaction) and KS methods, drawing from the authors' research.

In the CAS approaches, first, the wavefunctions are calculated in the absence of magnetic fields, and in a second step, EPR parameters are deduced from the wavefunctions. In two-step approaches, the quality of the wavefunction depends on the quality of the basis set, on the size of the active space, on the introduction of the dynamic correlation, and on the number of states included in the state interaction for the calculation of the SO coupling. In all cases, for a metal in $nd(f)^l$ configuration, minimal active space includes the l electrons in the 5(7) $nd(f)$ orbitals. Such a minimal active space is often sufficient for the description of f elements as far as g-factors and ZFS are concerned. In order to get accurate HFC interactions, the active spaces must allow for spin polarization to take place. For transition metals, also for ZFS and g-factors, the active spaces should be increased with some correlating orbitals, namely, the double shell nd' orbitals and some orbitals of the ligands, namely, the orbitals most involved in the bonding with the metal ion. Except for lanthanides, the inclusion of the dynamical correlation with perturbation theory tends to improve the results. For the state interaction, all the states with the same spin as the ground term of the free ion arising from the $nd(f)^l$ configuration are usually included. The lowest states with $S \pm 1$ should often be included as well.

Once the wavefunctions are calculated, the model space must be chosen. In the case of Kramers doublets, the spin Hamiltonian comprises only the Zeeman term linear in $\mathcal{S}$ and the g-factors are calculated according to Eq. (15). It is illustrated below with the example of neptunyl NpO_2^{2+}. For non-Kramers doublets, a ZFS

parameter must be added as shown below for the plutonyl PuO_2^{2+}. The case $S = 1$ is illustrated with a complex of Ni(II) where the pseudo-spin is very close to the real spin. Two examples are presented for $S = 3/2$: a complex of Co(II) with a ZFS splitting and a octahedral complex of Np(IV) without ZFS splitting but with a large third-order term in the pseudo-spin Hamiltonian. Finally, the $S = 2$ case is illustrated with a high-spin complex of Fe(VI) where there is a low-lying SF state. These calculations are all based on the second-order DKH operator with SO coupling treated by an atomic mean-field integral procedure. This section concludes with selected examples for hyperfine coupling extracted from KS calculations, where the relativistic effects are treated with the help of ZORA.

The neptunyl ion NpO_2^{2+} is a linear complex of Np^{6+} in $5f^1$ configuration [45]. The $5f$ orbitals split due to the interaction with the two oxygen atoms; the three orbitals of symmetry σ and π are antibonding and are destabilized. The remaining four orbitals, δ and ϕ, are nonbonding and are occupied with the single electron. The ground state is of symmetry $E_{5/2u}$ of the $\mathcal{D}_{\infty h}$ group and is close to the $M_J = \pm 5/2$ components of the $J = 5/2$ term of the free ion. Results are summarized in Table 1; $g_\parallel = \pm 4.23$ is close to the $2g_J J = 30/7$ value of the free ion limit for a $|5/2, \pm 5/2\rangle$ doublet. The spin and orbital contributions to the g-factors are determined by turning off the orbital and spin term, respectively, in the Zeeman interaction. The orbital contribution is the largest and opposite to the spin one, as it is the case for the free ion where spin and orbit are in opposite direction since the open shell is less than half filled and $L > S$.

The g-factors of the ground state of NpO_2^{2+} were measured by EPR spectroscopy in the solid state diluted either in $CsUO_2(NO_3)_3$ or in $Cs_2UO_2Cl_4$. The g-factors of the excited states were deduced from the absorption bands in a magnetic field. In the first environment, three nitrate ligands are in the equatorial plane of NpO_2 leading to a local $\mathcal{D}_{3h}$ symmetry, while in the latter, there are four chloride with a local $\mathcal{D}_{4h}$ symmetry. In the first complex, the ϕ orbitals split by interaction with the orbitals of the equatorial ligands and their orbital moment is partially quenched. It gives rise to a magnetic moment in the equatorial direction (see Table 1) dominated by the spin contribution; in this case, spin and orbit are opposite. In $\mathcal{D}_{4h}$, the δ orbitals split by mixing with the orbitals of the ligands quenching their orbital moment. One obtains again a magnetic moment in the equatorial plane almost as large as the axial one. It should be noticed that in this case, the spin and orbital contributions of the axial component have the same sign and are therefore additive. The main effect of the environment is to affect the ratio of δ and ϕ orbitals in the ground state, and it is this ratio which determines the magnetic properties. The effect of the ligands is so large in the chloride environment that there is no relationship anymore with the properties of the free actinyl ion.

As pointed out already, even-electron systems may have doubly degenerate states in presence of rotational symmetry or almost degenerate states with a small energy gap. In this case, the pseudo-spin is $\mathcal{S} = 1/2$. But, while in the spin space the kets behave as $\hat{\Theta}|1/2, \pm 1/2\rangle = \pm|1/2, \mp 1/2\rangle$ under time reversal, with $\hat{\Theta}$ being the time-reversal operator, in the real space $\hat{\Theta}|\lambda, a\rangle = |\lambda, a\rangle$ since there is an even

Table 1 Calculated and experimental g-factors for NpO_2^{2+} and PuO_2^{2+}. g^S and g^L are the spin and orbital contributions to the calculated g-factors (Data taken from Refs. [45,46])

		NpO_2^{2+}		$[NpO_2(NO_3)_3]^-$		$[NpO_2Cl_4]^{2-}$		PuO_2^{2+}		$[PuO_2(NO_3)_3]^-$	
		∥	⊥	∥	⊥	∥	⊥	∥	⊥	∥	⊥
calc	g	± 4.24	0.00	−3.49	± 0.23	1.76	± 1.50	6.09	0.00	5.92	0.00
	g^L	± 5.76	0.00	−4.69	∓ 0.63	1.72	∓ 2.42	9.90	0.00	9.62	0.00
	g^S	∓ 1.52	0.00	1.20	± 0.86	0.04	± 0.92i	−3.80	0.00	−3.68	0.00
exp				3.36;3.405	0.20;0.205	1.32;1.38	1.30			5.32	0.00

number of electrons. The states $|\lambda, a\rangle$ have no magnetic moment, but the magnetic moment arises from the coupling between the two states $\langle\lambda, a|\hat{\boldsymbol{\mu}}|\lambda, b\rangle \neq 0$ if $a \neq b$. It follows that the spin Hamiltonian takes the following form:

$$\hat{\mathcal{H}}_S = g_\parallel \beta_e \hat{\mathcal{S}}_z B_z + \Delta \hat{\mathcal{S}}_x \tag{57}$$

assigning the $|1/2, \pm 1/2\rangle \equiv \frac{1}{\sqrt{2}}(|\lambda, a\rangle \pm i|\lambda, b\rangle)$. The second term of Eq. (57) is a ZFS term linear in $\mathcal{S}$ and is not even with time reversal. In this equation, x and y do not refer to spatial directions and the Hamiltonian is not invariant under rotations. $g_\parallel$ is a number. Although ill defined for symmetries, this Hamiltonian is used to fit EPR spectra of non-Kramers systems, and its parameters can be determined from ab initio calculation.

The plutonyl ion PuO_2^{2+} is the analog of NpO_2^{2+} with a $5f^2$ configuration. The ground state is of 4_g symmetry, close to the $M_J = \pm 4$ of the ground free ion $^3H_4$ term. The calculated g-factor (see Table 1) is close to $2g_J J = 6.4$ of the free ion. The equatorial component is zero, and the spin and orbital contributions to the axial component have opposite signs. The EPR spectrum of plutonyl has been measured diluted in a diamagnetic crystal, forming $[PuO_2(NO_3)_3]^-$ clusters. The equatorial moment remains zero, even with the lowering of the symmetry since there is an even number of electrons in the molecule. In this case, the splitting of the ϕ orbitals has less effect than in neptunyl since the configuration of the ground state remains mostly $\phi^1\delta^1$.

In the complex $[Ni(II)(HIM_2 - py)_2(NO_3)]^+$, the Ni(II) ion ($3d^8$ configuration) is in a pseudo-octahedral environment. Magnetization measurements, HF-HF-EPR studies, and frequency domain magnetic resonance spectroscopy (FDMRS) studies indicated the presence of a very large Ising-type anisotropy with an axial ZFS parameter $D = -10.1 \pm 0.1\,\mathrm{cm}^{-1}$ and a rhombic ZFS parameter of $E = 0.3 \pm 0.1\,\mathrm{cm}^{-1}$ [47]. The spectrum is fitted with an isotropic $g_{\mathrm{iso}} = 2.17$. In an octahedral ligand field, the SR ground state is a $^3A_{2g}$ with a $t_{2g}^6 e_g^2$ configuration. This state is orbitally nondegenerate and triply degenerate for the spin. With SO coupling, it becomes a T_2 state and remains triply degenerate. A distortion of the octahedral environment removes this degeneracy, creating three nondegenerate states. SO coupling is dominated by coupling to a $^3T_{2g}$ state with configuration $t_{2g}^5 e_g^3$. The latter is triply degenerate in octahedral symmetry and also splits in three states when the symmetry is lowered. The model space consists in the three states arising from the ground $^3A_{2g}$ state and $\mathcal{S} = 1$.

According to SO-CASPT2 calculations, the $^3T_{2g}$ state splits in three components at 7,750, 10,088, and 10,504 cm^{-1} above the energy of the ground state. This large splitting is at the origin of the large ZFS in this molecule. According to Eq. (51), one gets $D = -11.53\,\mathrm{cm}^{-1}$ and $E = 0.48\,\mathrm{cm}^{-1}$. The directions of the principal axes of the **d** tensor are depicted on Fig. 3. The norms of the projections of the model wavefunctions in the target space $\mathcal{P}|\lambda, a\rangle$ are 0.99 and their overlap about 10^{-4}.

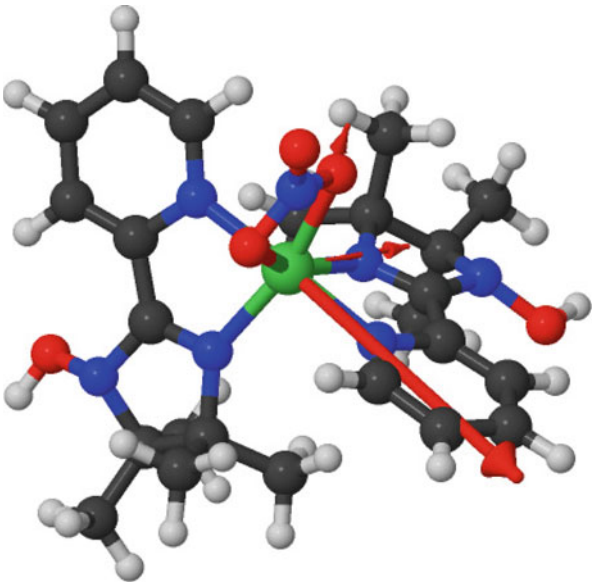


Fig. 3 The $[Ni(II)(HIM_2 - py)_2(NO_3)]^+$ molecule. The *thick arrows* show the direction of the axial magnetization and the *thinner arrows* the two other principal axes of the **d** tensor

It follows that the matrix **C** is close to being orthogonal and close to the rotation matrix $\mathcal{R}$ of Eq. (49). Equations (49) and (51) provide similar results.

The g-factors determined by the projection procedure, Eq. (45), and the matching procedure are the same $g_X = 2.21$, $g_Y = 2.22$, and $g_Z = 2.30$. The principal axes of the **g** tensor are very close to those of the **d** tensor even if the molecule has no symmetry. The g-factors are almost equal and the Zeeman interaction is almost isotropic. This means that the magnetic behavior is dictated by the principal axes of the ZFS tensor **d**. The spin contribution to the g-factors is $g_X^S = g_Y^S = 1.998$, $g_Z^S = 1.992$ and the orbital contributions are $g_X^L = 0.225$ $g_Y^L = 0.222$ and $g_Z^L = 0.306$. As expected for a weak SO case, the spin contribution is by far the largest. But it is isotropic and close to 2. The orbit contribution is smaller, but it brings the departure from the free electron and the anisotropy of the Zeeman interaction.

In $HgCo(NCS)_4$, the Co(II) ion has a $3d^7$ configuration and is in a pseudo-tetrahedral environment. The $[Co(NCS)_4]^{2-}$ molecule belongs to the S_4 point group. In T_d, the ground state would be a fourfold degenerate $^4A_2$, but in the lower symmetry, this state splits in two Kramers doublets. The spin Hamiltonian parameters were deduced from susceptibility data measured with a SQUID in the range 1.7–300 K [48]; D lies between 10 and 11 cm^{-1}, $g_\parallel = 2.168$, and $g_\perp = 2.251$.

The splitting between the two Kramers doublets calculated with SO-CASPT2 is 16.2 cm^{-1}. Due to the symmetry of the molecule, all spin Hamiltonian tensors have the same principal axis, and the spin Hamiltonian takes the following form:

$$\begin{aligned}\hat{\mathcal{H}}_S = &\beta_e\left(g_\perp B_X \hat{\mathcal{S}}_X + g_\perp B_Y \hat{\mathcal{S}}_Y + g_\parallel B_Z \hat{\mathcal{S}}_Z\right)\\ &+ \beta_e\left(g'_\perp B_X \hat{\mathcal{S}}^3_X + g'_\perp B_Y \hat{\mathcal{S}}^3_Y + g'_\parallel B_Z \hat{\mathcal{S}}^3_Z\right)\\ &+ D\left(\hat{\mathcal{S}}^2_Z - \frac{5}{4}\right) + E\left(\hat{\mathcal{S}}^2_X - \hat{\mathcal{S}}^2_Y\right)\end{aligned} \tag{58}$$

The matrix $\mathbf{g}'$ is reduced to its diagonal elements. The ZFS tensor $\mathbf{d}$ is calculated by the effective Hamiltonian technique; the Z axis lies along the C_2 axis, and $D = 8.1\,\mathrm{cm}^{-1}$ and $E = 0$. The validity of the effective Hamiltonian procedure can be assessed by the projection of the $|\lambda, a\rangle$ on the $|\lambda', M\rangle$, more than 0.99 and the overlap between the projected vectors, less than 10^{-4}. The projection technique provides the tensor $\mathcal{A}$ defined in Eq. (45) leading to $g_\parallel = 2.173$ and $g_\perp = 2.255$. The matching technique permits the determination of both linear and cubic terms: one gets $g_\parallel = 2.178$, $g'_\parallel = -0.002$, $g_\perp = 2.263$, $g'_\perp = -0.004$. The Zeeman interaction is mostly isotropic and the cubic term is negligible. The two techniques give close but not similar g-factors due to the neglect of the cubic term in the first one. These values are in very good agreement with the experimental ones. The spin (orbital) contributions are for $g_\perp$ 1.992 (0.271), for $g'_\perp$ −0.0002 (−0.004), for $g_\parallel$ 1.989 (0.189), and for $g'_\parallel$ −0.0002 (−0.002). The spin contribution is again the largest, isotropic, and close to 2. The orbital one is more anisotropic and provides the only contribution to the cubic term. As it was the case for the Ni(II) complex, in the weak SO limit, the anisotropy of the magnetic property is determined by the ZFS interaction.

Another description is to consider the two Kramers doublets independently, each being described by a $\mathcal{S} = 1/2$. One gets for the first Kramers doublet $g_{1\parallel} = 2.18$ and $g_{1\perp} = 4.50$ and for the second one $g_{2\parallel} = 6.52$ and $g_{2\perp} = 0.1$. This description does not give information on the second-order Zeeman interaction between the two Kramers doublets. But in the case of a large ZFS, the transition between the two Kramers doublets cannot be induced even using HFHF EPR, and one only determines the magnetic properties of the ground Kramers doublet. In this scheme, the magnetic anisotropy is described by the Zeeman interaction.

Another example for $\mathcal{S} = 3/2$ is the $NpCl_6^{2-}$ cluster diluted in Cs_2ZrCl_6. The Np^{4+} ion has a $5f^3$ configuration and is in an octahedral environment. The ground state is a quartet $F_{3/2u}$ (using Mulliken's notation, or Γ_8 in Bethe's notation) [49]. In the principal axis system, the spin Hamiltonian takes the form

$$\hat{\mathcal{H}}_S = \beta_e g \hat{\boldsymbol{\mathcal{S}}} \cdot \boldsymbol{B} + \beta_e g'\left(B_X \hat{\mathcal{S}}^3_X + B_Y \hat{\mathcal{S}}^3_Y + B_Z \hat{\mathcal{S}}^3_Z\right) \tag{59}$$

Due to the cubic symmetry, $\hat{\mathcal{H}}^Z_{(1)}$ is isotropic, $\hat{\mathcal{H}}^Z_{(3)}$ is invariant by "changing" the axes by permutation $X \longleftrightarrow Y \longleftrightarrow Z$ but is not isotropic. There are two degrees of freedom, the two scalar g and g' which were deduced from EPR measurements as $g = -0.516$ and $g' = 0.882$. These two numbers define the magnetization of the two pairs of Kramers doublets, $\langle \pm 3/2|\mu^u| \pm 3/2\rangle = \mp(\frac{3}{2}g + \frac{27}{8}g')$ and

$\langle \pm 1/2|\mu^u| \pm 1/2\rangle = \mp(\frac{1}{2}g + \frac{1}{8}g')$. The matching technique permits to find the rotation $\mathcal{R}$ in the model space which assigns the $|\lambda, a\rangle$ to the $|\mathcal{S}, \mathcal{M}_\mathcal{S}\rangle$. One can find two possible assignments as was already pointed out by Bleaney, corresponding to the permutation $\{\Psi_{3/2} \leftrightarrow \Psi_{-1/2}; \Psi_{-3/2} \leftrightarrow \Psi_{1/2}\}$. One of the solutions transforms under rotations as a spin $S = 3/2$, while the other one does not. The solution given above fitting the EPR spectrum does not. From calculations, one gets $g = -0.406$ and $g' = 0.785$. The spin and orbital contributions are 0.027 and −0.460 for g and −0.250 and 1.285 for g'. The accordance with the experimental values is reasonable. The spin and orbital contributions are opposite, the orbital one being the largest as in the free ion. The cubic term is by far non-negligible; it is larger than the linear one and opposite in sign. Due to the high symmetry of the complex, there is no ZFS. However, the cubic term plays an essential role for the Zeeman interaction.

When the complex is distorted, one gets two Kramers doublets with different energies. The two g-tensors are not collinear. Two pseudo-spin models can be used: either one considers two distinct $\mathcal{S} = 1/2$ each with its own g-tensor, or the two g-tensors are not collinear. The two spin spaces are coupled through second-order Zeeman interaction. It is usually the way one describes the low-energy spectra of lanthanide ions. Another way is to describe the model space as an $\mathcal{S} = 3/2$ case. The $\mathbf{g}$ and $\mathbf{g}'$ tensors are not collinear in general, and even if this space is more complete than the previous description, the parameters are less intuitive.

In the [FeLCl] complex with $L = \beta$-diketiminate, the Fe(II) is in $3d^6$ high-spin configuration with a $S = 2$ ground state. EPR measurements in the X band gives a quasidegenerate $M_S = \pm 2$ ground doublet with an axial $g = 10.9$ and a small splitting of $\Delta = 0.35\,\mathrm{cm}^{-1}$. The next component of the spin quintet is estimated to be higher than $150\,\mathrm{cm}^{-1}$ [50]. In a pure spin quintet, the $M_S = \pm 2$ components have a g-factor of 6. The experimental value shows that there is a large orbital contribution due to a partially quenching of the orbital moment and a low-lying state. The CASPT2 calculation shows that without SO coupling, the first excited state is another spin quintet lying $516\,\mathrm{cm}^{-1}$ above the ground state. With the SO coupling, the spin quintet becomes a ground doublet split by $0.34\,\mathrm{cm}^{-1}$, and the other components are 118, 132, and $180\,\mathrm{cm}^{-1}$ above. The g-factors of the ground doublet modeled with a $\mathcal{S} = 1/2$ are given in Table 2. The magnetization is along the Fe-Cl bond with $g = 10.9$ in perfect accordance with the experimental value.

Table 2 g-factors of FeLCl calculated with two model spaces $\mathcal{S} = 1/2$ and $\mathcal{S} = 2$. g^S and g^L are the spin and orbital contributions

		$\mathcal{S} = 1/2$			$\mathcal{S} = 2$					
		$\mathbf{g}$			$\mathbf{g}$			$\mathbf{g}'$		
		x	y	z	x	y	z	x	y	z
calc	g	0.0	0.0	10.903	2.104	1.983	3.108	−0.015	−0.013	−0.094
	g^S	0.0	0.0	7.994	2.000	1.965	2.000	−0.016	−0.012	0.000
	g^L	0.0	0.0	2.909	0.104	0.018	1.108	0.002	−0.001	−0.094
exp	0.0	0.0	10.9							

The spin contribution of 7.99 is the expected value for a spin quintet, and one notices a large orbital contribution due to the low-lying quintet state. The five components issued from the spin quintet may be modeled with a $\mathcal{S} = 2$ pseudo-spin. In the case where all the matrices are collinear, the spin Hamiltonian is the following:

$$\begin{aligned}\hat{\mathcal{H}}_S &= \beta_e \left(g_X B_X \hat{\mathcal{S}}_X + g_Y B_Y \hat{\mathcal{S}}_Y + g_Z B_Z \hat{\mathcal{S}}_Z \right) \\ &+ \beta_e \left(g'_X B_X \hat{\mathcal{S}}_X^3 + g'_Y B_Y \hat{\mathcal{S}}_Y^3 + g'_Z B_Z \hat{\mathcal{S}}_Z^3 \right) \\ &+ D \left(\hat{\mathcal{S}}_Z^2 - 2 \right) + E(\hat{\mathcal{S}}_X^2 - \hat{\mathcal{S}}_Y^2) + D' \left(\hat{\mathcal{S}}_Z^4 - \frac{34}{5} \right) + E' \left(\hat{\mathcal{S}}_X^4 - \hat{\mathcal{S}}_Y^4 \right) \end{aligned} \quad (60)$$

The matrices $\mathbf{g}'$ and $\mathbf{d}'$ are reduced to their diagonal elements. All the spin parameters have been determined. Z is along the Fe − Cl axis. The g-factors and their spin and orbital contributions are summarized in Table 2. For the ZFS tensors, one finds $D = -58.0\,\mathrm{cm}^{-1}$, $E = 2.0\,\mathrm{cm}^{-1}$, $D' = 3.3\,\mathrm{cm}^{-1}$, and $E' = 0.1\,\mathrm{cm}^{-1}$. The tensor of second order is largely dominant. While the rhombic term E models the splitting between the two $M_S = \pm 2$ components, the parameter D' of the term of fourth order is necessary to model the whole spectrum of energy of the five pseudo-spin components. It should be noticed that, due to the large coupling with the excited spin quintet, the projection of the five wavefunctions in the target space is reduced to 0.89 and 0.95. For the Zeeman interaction, as in $Co(NCS)_4$, the linear term is largely dominant and the cubic one is negligible. The spin contribution is isotropic and close to 2, while the orbital contribution is very important in the Z direction, larger than 1, and negligible in the two other ones.

Electronic EPR parameters can be calculated in many different cases. In transition metal complexes, the pseudo-spin is usually the spin of the ground state. In the case of a large ZFS, the model space can be reduced to the lower components, and effective values of g-factors are measured and calculated. When the pseudo-spin is equal to the spin, the spin contribution to the g-factor is isotropic and close to 2. The anisotropy of $\mathbf{g}$ arises from orbital contributions. For large values of the pseudo-spin, the higher-order terms, in $\mathcal{S}^3$ for the Zeeman interaction and in $\mathcal{S}^4$ the ZFS interaction, are much smaller than the $\mathcal{S}$ and $\mathcal{S}^2$ terms. Except in the case of a low-lying excited state, the anisotropy of the magnetic properties arises from the ZFS tensor, the g-factors being almost isotropic.

In the case of complexes with heavy elements, the pseudo-spin is rarely larger than 2. Usually, the spin and orbital components are in opposite directions for less than filled open shells and in the same direction otherwise, as a reminiscence of the free ion. But in the case of the quenching of some orbitals due to the interaction with the ligands, this rule can be skewed for actinides.

Regarding HFC, there are few examples highlighting the role of relativistic effects and finite nuclear volume effects better than ^{199}Hg. Table 3 lists a collection of calculated (DFT, with a hybrid functional for NR and ZORA and a non-hybrid functional for the four-component calculations) and experimental data for

Table 3 Isotropic ^{199}Hg HFC constants, in units of MHz, calculated with different relativistic methods[a]

	HgF	HgAg
NR PN	9,173	2,068
ZORA PN	21,958	3,404
ZORA FN	19,171	3,094
4-component PN	18,927	3,690
4-component FN	16,895	3,285
Expt.	22,163	2,720

[a] *PN* = point nucleus, *FN* = finite nucleus (spherical Gaussian model), *ZORA* = zeroth-order regular approximation 2-component method. For citations of the original data, see Table 3 of Ref. [51]

the radicals HgH and HgAg. The increase of the Hg HFC from NR to relativistic calculations, in particular for HgH, is rather spectacular. The effect is mainly due to SR effects from the relativistic increase of the spin density at the Hg nucleus. The effects from a finite nuclear volume are also very large, roughly on the order of 10 % relative to the total, and decrease the HFC. SO effects on mercury HFC constants tend to be relatively minor because they are dominated by contributions from the Hg 6*s* orbital.

The situation is different for the $5f^1$ complex ^{237}Np and ^{19}F HFC constants of NpF6. Experimentally, the isotropic HFC constants were found to be -1995 MHz for ^{237}Np and -73 MHz for ^{19}F. For Np, the HFC tensor is isotropic due to the octahedral symmetry of the complex. For ^{19}F, there are two unique tensor components, $a_{\parallel}$ in the direction of the Np–F axes and a degenerate pair of $a_{\perp}$ perpendicular to the Np–F axes, with $a_{\parallel} = -132$, $a_{\perp} = -42$ MHz experimentally. The unpaired electron is described at the SR level by a nonbonding Np $5f$ orbital of δ symmetry with respect to the Np–F axes. Without SO coupling, the isotropic Np HFC is calculated to be much too small in magnitude, by a factor of 3 to 5 depending on the computational method. Spin polarization is responsible for the residual Np HFC and most of the fluorine HFC at the SR level. Under the SO interaction, there is a very dominant contribution to the Np HFC from the PSO mechanism. This mechanism can be interpreted as follows: Via SO coupling, the unpaired spin at Np creates a paramagnetic orbital current density in the $5f$ shell which then interacts magnetically with the nuclear spin. The NMR shielding tensors for diamagnetic molecules afford a similar mechanism, except that a paramagnetic orbital current density is induced by the external magnetic field rather than via an unpaired spin and SO coupling.

Figure 4 shows the ^{237}Np HFC constant for NpF6 calculated with HF theory and DFT using Eq. (22). The horizontal axis indicates a parameter λ used to scale the SO Hamiltonian integrals in the ground-state calculations, meaning that $\lambda = 0$ gives the SR limit. The fact that the data exhibit significant curvatures shows that SO effects beyond first order are important and reduce the HFC magnitude. Indeed,

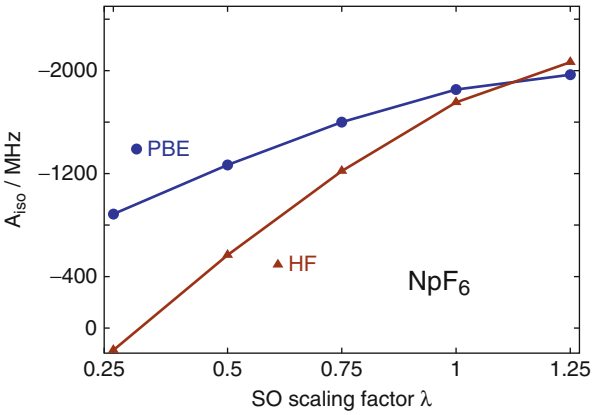


Fig. 4 ^{237}Np HFC constant of NpF6 versus scaling of SO Hamiltonian integrals in the ground-state calculation ($\lambda = 0$ means SR, $\lambda = 1$ means full SO). HF and DFT calculations (PBE functional) of Ref. [30] utilizing Eq. (22)

calculations where SO coupling is treated with a linear-response method to produce the PSO mechanism overestimate the Np HFC by 15 % or more, depending on the functional used for the calculations. Further details and citations of prior studies where such SO scaling has been used can be found in Ref. [30].

For ^{19}F, CASSCF calculations recently indicated [52] that without SO coupling, the HFC tensor anisotropy is not correctly predicted. With SO coupling, the calculations predict $a_{\parallel}$ to be twice as large in magnitude as $a_{\perp}$, with the latter close to experiment. The magnitude of the calculated $a_{\parallel}$ appears to be underestimated. However, the measurements were performed on a solid, with NpF6 diluted in a UF6 host crystal, while the calculations were for an isolated molecule. It is presently unknown precisely how solid-state packing affects the fluorine HFC tensors.

The CAS calculations of g-tensors and ZFS discussed above were performed with the MOLCAS suite of programs. For the benefit of the reader, some details regarding the computations are given here. $\mathbf{NpO_2^{2+}}$, $\mathbf{[NpO_2(NO_3)_3]^-}$, $\mathbf{[NpO_2Cl_4]^{2-}}$, $\mathbf{PuO_2^{2+}}$ and $\mathbf{[PuO_2(NO_3)_3]^-}$: Results are given with ano.rcc basis sets of TZP quality at the SO-CASPT2 level with CAS(7,10) comprising the 7 $5f$ orbitals and the two bonding σ and π orbitals. The environment is described by ECP and point charges. The state interaction for the calculation of the SO coupling is performed with six spin doublets. $\mathbf{[Ni(II)(HIM_2 - py)_2(NO_3)]^+}$: Results are given with ano.rcc basis sets of TZP quality on the Ni and DZP on the ligands at the SO-CASPT2 level with CAS(8,10) comprising the eight $3d$ electrons within a double shell of d orbitals. The state interaction for the calculation of SO coupling is performed with 10 triplets and 15 singlets. Similar calculations are published in Ref. [53]. **$\mathbf{HgCo(NCS)_4}$**: Calculations are performed with ano.rcc basis sets of TZP quality for the $[Co(NCS)_4]^{2-}$ complex. The complex is embedded in point charges describing the rest of the crystal. Hg^{2+} ions are described by ECPs and other ions by point charges within a radius of 10 Å. The cluster is described at the SO-CASPT2 level with 10 spin quartets and 40 spin doublets using a CAS(7,5) (7 electrons in the 5 $3d$ orbitals). $\mathbf{NpCl_6^{2-}}$: Calculations are performed with ano.rcc basis sets of TZP quality for the $[NpCl_6]^{2-}$ complex. The complex is embedded in point charges

and model potentials describing the rest of the crystal. Hg^{2+} ions are described by ECPs and other ions by point charges. The cluster is described at the SO-CASPT2 level with 35 quartets and 84 doublets using a CAS(3,7) (3 electrons in the 7 5*f* orbitals). **LFeCl**: Results are given with ano.rcc basis sets of TZP quality on the Fe and DZP on the ligands at the SO-CASPT2 level with CAS(6,5) comprising the 6 3*d* electrons within the 5 3*d* orbitals. The state interaction for the calculation of SO coupling is performed with 5 quintets, 15 triplets, and 10 singlets.

Summary

There is a large variety of techniques available for ab initio calculations of EPR parameters. Nonetheless, it can be a complicated task to extract EPR spin Hamiltonian parameters from ab initio calculations if SO coupling is strong, if the degeneracy of the state of interest is higher than twofold. In the limit of weak SO coupling, the effective Hamiltonian technique is well established. For strong SO coupling, the wavefunctions should include SO coupling variationally or via state interaction. KS methods can be an effective alternative to wavefunction-based methods.

Acknowledgements J.A. acknowledges support of his research on EPR and NMR parameters of open-shell heavy metal complexes by the US Department of Energy, Office of Basic Energy Sciences, Heavy Element Chemistry program, under grant DE-FG02-09ER16066.

References

1. Atherton NM (1993) Principles of electron spin resonance. Ellis Horwood series in physical chemistry. Prentice Hall, New York
2. Rieger PH (2007) Electron spin resonance. Analysis and interpretation. The Royal Society of Chemistry, Cambridge
3. Harriman JE (1978) Theoretical foundations of electron spin resonance. Academic, New York
4. Abragam A, Bleaney B (1970) Electron paramagnetic resonance of transition ions. Clarendon, Oxford
5. Mohr PJ, Taylor BN, Newell DB (2012) CODATA recommended values of the fundamental physical constants: 2010. Rev Mod Phys 84:1527–1605
6. Butler JE, Hutchison CA Jr (1981) Electron paramagnetic resonance and electron nuclear double resonance of 237-neptunium hexafluoride in uranium hexafluoride single crystals. J Chem Phys 74:3102–3119
7. Griffith JS (1967) Transformations of a spin-Hamiltonian induced by reassignments of ficticious spin states. Mol Phys 12:359
8. Neese F, Solomon EI (1998) Calculation of zero-field splittings, g-values and the relativistic nephelauxetic effect in transition metal complexes. Application to high-spin ferric complexes. Inorg Chem 37:6568
9. Bolvin H (2006) An alternative approach to the g-matrix: theory and applications. Chem Phys Chem 7:1575
10. Chibotaru LF, Ungur L (2012) Ab initio calculation of anisotropic magnetic properties of complexes. I. Unique definition of pseudospin Hamiltonians and their derivation. J Chem Phys 137:064112–22

11. Chibotaru LF (2013) Ab initio methodology for pseudospin hamiltonians of anisotropic magnetic complexes. Adv Chem Phys 153:397
12. Roos BO, Taylor PR, Siegbahn PEM (1980) Chem Phys 48:157
13. Andersson K, Malmqvist PA, Roos BO, Sadlej AJ, Wolinski K (1990) J Phys Chem 94:5483
14. Angeli C, Cimiraglia R, Malrieu JP (2002) J Chem Phys 117:9138
15. Lan TN, Kurashige Y, Yanai T (2014) Toward reliable prediction of hyperfine coupling constants using ab initio density matrix renormalization group method: diatomic $^2\Sigma$ and vinyl radicals as test cases. J Chem Theory Comput 10:1953–1967
16. Fernandez B, Jørgensen P, Byberg J, Olsen J, Helgaker T, Jensen HJA (1992) Spin polarization in restricted electronic structure theory: multiconfiguration self-consistent-field calculations of hyperfine coupling constants. J Chem Phys 97:3412–3419
17. Malmqvist PA, Roos BO (1989) Chem Phys Lett 155:189
18. Vad MS, Pedersen MN, Nørager A, Jensen HJA (2013) Correlated four-component EPR g-tensors for doublet molecules. J Chem Phys 138:214106–214109
19. Kutzelnigg W (1990) Perturbation theory of relativistic corrections 2. Analysis and classification of known and other possible methods. Z Phys D 15:27–50
20. Perdew JP, Ruzsinszky A, Constantin LA, Sun J, Csonka GI (2009) Some fundamental issues in ground-state density functional theory: a guide for the perplexed. J Chem Theory Comput 5:902–908
21. Eriksson LA (1998) ESR hyperfine calculations. In: von Ragué Schleyer P (ed) Encyclopedia of computational chemistry. Wiley, Chichester, pp 952–958
22. Patchkovskii S, Schreckenbach G (2004) Calculation of EPR g–tensors with density functional theory. In: Kaupp M, Bühl M, Malkin VG (eds) Calculation of NMR and EPR parameters. Theory and applications. Wiley-VCH, Weinheim, pp 505–532
23. Kaupp M, Köhler FH (2009) Combining NMR spectroscopy and quantum chemistry as tools to quantify spin density distributions in molecular magnetic compounds. Coord Chem Rev 253:2376–2386
24. Neese F (2009) Prediction of molecular properties and molecular spectroscopy with density functional theory: from fundamental theory to exchange-coupling. Coord Chem Rev 253:526–563
25. Helgaker T, Coriani S, Jørgensen P, Kristensen K, Olsen J, Ruud K (2012) Recent advances in wave function-based methods of molecular-property calculations. Chem Rev 112:543–631
26. Autschbach J, Ziegler T (2003) Double perturbation theory: a powerful tool in computational coordination chemistry. Coord Chem Rev 238/239:83–126
27. Autschbach J (2010) Relativistic effects on magnetic resonance parameters and other properties of inorganic molecules and metal complexes. In: Barysz M, Ishikawa Y (eds) Relativistic methods for chemists. Challenges and advances in computational chemistry and physics, vol 10. Springer, Dordrecht, chap 12, pp 521–598
28. van Lenthe E, Wormer PES, van der Avoird A (1997) Density functional calculations of molecular g-tensors in the zero order regular approximation for relativistic effects. J Chem Phys 107:2488–2498
29. Malkin E, Repisky M, Komorovsky S, Mach P, Malkina OL, Malkin VG (2011) Effects of finite size nuclei in relativistic four-component calculations of hyperfine structure. J Chem Phys 134:044111–8
30. Verma P, Autschbach J (2013) Relativistic density functional calculations of hyperfine coupling with variational versus perturbational treatment of spin-orbit coupling. J Chem Theory Comput 9:1932–1948
31. Schmitt S, Jost P, van Wüllen C (2011) Zero-field splittings from density functional calculations: Analysis and improvement of known methods. J Chem Phys 134:194113
32. van Wüllen C (2009) Magnetic anisotropy from density functional calculations. Comparison of different approaches: Mn12O12 acetate as a test case. J Chem Phys 130:194109
33. Autschbach J (2012) Perspective: relativistic effects. J Chem Phys 136:150902
34. Visscher L, Dyall K (1997) Dirac–Fock atomic electronic structure calculations using different nuclear charge distributions. At Data Nucl Data Tables 67:207–224

35. Van den Heuvel W, Soncini A (2012) NMR chemical shift in an electronic state with arbitrary degeneracy. Phys Rev Lett 109:073001
36. Rodowicz C (2013) Higher-order field-dependent terms in spin Hamiltonians for transition ions implications for high-magnetic field and high-frequency EMR measurements. Nukleonika 58:341
37. Maurice R, Bastardis R, de Graaf C, Suaud N, Mallah T, Guihéry N (2009) Universal approach to extract anisotropic spin hamiltonians. J Chem Theory Comput 11:2977
38. Pryce MHL (1959) Sign of g in magnetic resonance, and the sign of the quadrupole moment of Np237. Phys Rev Lett 3:375
39. Axe JD, Stapleton HJ, Jefries CD (1961) Paramagnetic resonance hyperfine structure of tetravalent Pa231 in Cs2ZrCl6. Phys Rev 121:1630
40. Rigny P, Plurien P (1967) Resonance paramagnetique dans les fluorures complexes d'uranium (V) de type UF6M. J Phys Chem Solids 28:2589
41. Eisenstein JC, Pryce MHL (1965) Electronic structure and magnetic properties of the neptunyl ion. J Res Natl Bur Stand Sect A 69A:217–235
42. Chibotaru LF, Ungur L (2012) Negative g factors, berry phases, and magnetic properties of complexes. Phys Rev Lett 109:246403
43. Chibotaru L, Ceulemans A, Bolvin H (2008) The unique definition of the g tensor of a Kramers doublet. Phys Rev Lett 101:033003
44. Notter FP, Bolvin H (2009) Optical and magnetic properties of the $5f^1 AnX_6^q$-series: a theoretical study. J Chem Phys 130:184310
45. Gendron F, Paez Hernandez D, Notter FP, Pritchard B, Bolvin H, Autschbach J (2014) Magnetic properties and electronic structure of neptunyl(VI) complexes: Wavefunctions, orbitals, and crystal-field models. Chem Eur J 20:7994
46. Gendron F, Pritchard B, Bolvin H, Autschbach J (2014) Magnetic resonance properties of actinyl carbonate complexes and plutonyl(VI)-tris-nitrate. Inorg Chem 53:8577
47. Rogez G, Rebilly JN, Barra AL, Sorace L, Blondin G, Kirchner N, Duran M, van Slageren J, Parsons S, Ricard L, Marvilliers A, Mallah T (2005) Very large ising-type magnetic anisotropy in a mononuclear NiII complex. Angew Chem 44:1876
48. Nelson D, ter Haar LW (1993) Single crystal studies of the zero-field splitting and magnetic exchange interactions in the magnetic susceptibility calibrant HgCo(NCS)4. Inorg Chem 32:182
49. Páez Hernàndez D, Bolvin H (2014) Magnetic properties of a fourfold degenerate state: Np^{4+} ion diluted in Cs_2ZrCl_6 crystal. J Electron Spectrosc Relat Phenom 194:74
50. Andres H, Bominaar EL, Smith JM, Eckert NA, Holland PL, Münck E (2002) Planar three coordinate high spin FeII complexes with large orbital angular momentum: mossbauer, electron paramagnetic resonance, and electronic structure studies. J Am Chem Soc 124:3012
51. Autschbach J (2014) Relativistic calculations of magnetic resonance parameters: background and some recent developments. J Philos Trans A 372:20120489
52. Sharkas K, Autschbach J (2015) Effects from spin-orbit coupling on electron-nucleus hyperfine coupling calculated at the restricted active space level for Kramers doublets. J Chem Theory Comput 11:538
53. Chibotaru L, Ungur L (2012) Ab initio calculation of anisotropic magnetic properties of complexes. I. Unique definition of pseudospin Hamiltonians and their derivation. J Chem Phys 137:064112

Zero-Field Splitting in Transition Metal Complexes: Ab Initio Calculations, Effective Hamiltonians, Model Hamiltonians, and Crystal-Field Models

24

Rémi Maurice, Ria Broer, Nathalie Guihéry, and Coen de Graaf

Contents

Introduction . . . 766
Ab Initio Calculations and Effective Hamiltonians . . . 768
From the Dirac Equation to Contracted Spin-Orbit Configuration Interaction . . . 768
The Treatment of the Spin-Orbit Free Electron Correlation . . . 770
On the Effective Hamiltonian Theory . . . 771
Model Hamiltonians and Effective Hamiltonians . . . 772
Mononuclear Complexes . . . 773
Binuclear Complexes . . . 779
Magneto-Structural Correlations . . . 785
Magneto-Structural Correlations Based on Ab Initio Calculations . . . 785
Magneto-Structural Correlations Based on Crystal-Field Models and Ab Initio Calculations . . . 787
Conclusion . . . 791
References . . . 792

R. Maurice (✉)
SUBATECH, UMR CNRS 6457, IN2P3/EMN Nantes/Université de Nantes, Nantes, France
e-mail: remi.maurice@subatech.in2p3.fr

R. Broer
Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands
e-mail: r.broer@rug.nl

N. Guihéry
Laboratoire de Chimie et Physique Quantiques, IRSAMC/UMR5626, Université de Toulouse 3, Toulouse, France
e-mail: nathalie.guihery@irsamc.ups-tlse.fr

C. de Graaf
Departament de Química Física I Inorgánica, Universitat Rovira i Virgili, Tarragona, Spain

Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain
e-mail: coen.degraaf@urv.cat

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_37

Abstract

Zero-field splitting (ZFS) is one of the essential ingredients for the occurrence of magnetic bistability at a molecular level. It is commonly understood as the loss of degeneracy of the spin components of a spin-orbit free electronic state in the absence of an external magnetic field. The loss of degeneracy finds its origin in the combined action of spin-orbit coupling and an anisotropic crystal field exerted on the magnetic center. Although ZFS was already described in the early days of quantum mechanics, it became a central issue of many theoretical and experimental studies after the discovery of single-molecule magnet behavior about twenty years ago. Moreover, ZFS also plays an important role in the magnetic properties of multiferroic solid state compounds, where electric and magnetic properties are intrinsically coupled and one may induce magnetic phase transitions by applying an external electric field, or vice versa. ZFS is commonly described in terms of model Hamiltonians that are basically introduced on a phenomenological basis. Typically, these model Hamiltonians only contain spin operators, since the ZFS applies by definition only to systems where the ground state orbital configuration is well separated from the other configurations (i.e., no orbital degree of freedom), and hence, the effective description of the low-lying energy levels can be restricted to the spin variables. This chapter aims to fill the gap between ab initio calculations based on the full electronic Hamiltonian and the phenomenological Hamiltonians used to describe ZFS. In this way, the model Hamiltonians can be validated and put on a rigorous foundation. Moreover, we establish magneto-structural correlations and demonstrate that these correlations can be understood within a crystal field reasoning.

Introduction

Zero-field splitting (ZFS) is a concept that is commonly used to indicate the loss of the degeneracy of the spin components of a spin-orbit free (SOF) state in the absence of an external magnetic field. This picture of ZFS is based on the separation of SOF and spin-dependent effects such as spin-orbit coupling (SOC). With this viewpoint, ZFS between the spin components of $S > 1/2$ SOF states can be observed in mononuclear complexes (i.e., complexes with one transition metal (TM) center), when (i) the "crystal" field felt by the TM ion is anisotropic (i.e., does not correspond to a spherical or cubic symmetry) and (ii) when SOC is sufficiently large [56]. Spin-spin coupling (SSC) is also at play [55], but usually to a lesser extent. ZFS can in principle occur in any type of systems; however, it is commonly associated with an orbitally nondegenerate ground SOF state. This chapter focuses on TM complexes and more particularly the $3d$ ones. Mononuclear and binuclear (i.e., with two TM centers) complexes will be considered to introduce the main features of single-ion and molecular or intersite anisotropies.

Although ZFS is a concept that has been known for almost a century, it has triggered new investigations in the last 20 years, among which theoretical studies play a key role. One of the reasons for this revival is the discovery of the single-molecule

magnet (SMM) behavior [26], which was first evidenced in the so-called Mn_{12} molecule [13]. In this system, the low-temperature magnetic behavior is interpreted as coming from the splitting of the spin components of the $S = 10$ SOF ground state. The slow relaxation of the magnetization from the $M_S = -10$ to $M_S = 10$ components and vice versa is observable at reasonable experimental timescales. The relaxation occurs via different mechanisms, such as thermal activation, direct tunneling, or thermally assisted tunneling. More information concerning SMMs can be, for instance, found in the nice review of Gatteschi and Sessoli [26]. Let us just recall that in order to design higher-temperature SMMs, various strategies have been attempted, as, for instance, enlarging the number of TM centers or enlarging the single-ion anisotropies. More particularly for the latter strategy, it became clear that more extensive studies were necessary to understand the properties of "exotic" coordination spheres with pentacoordinated or heptacoordinated metal centers [16, 68, 75, 76] or even low-coordinated cases [7, 89]. Moreover, the search for magneto-structural correlations from ab initio calculations and crystal-field models appeared necessary to eventually guide the synthesis of new coordination complexes, as highlighted by Telser in 2006 [83].

ZFS also plays an important role in the magnetic properties of condensed matter, as, for instance, ionic solids. In these systems, TM centers may be the subjects of local and intersite anisotropies. Low-temperature magnetic properties of extended systems are at least partly driven by the microscopic interactions that lead to ZFS. One may quote magnetic multiferroics [23], for which magnetic and, for instance, electric transitions are coupled, meaning that one can in principle influence the magnetism of the material by the application of an external electric field or the electric polarization by a magnetic field. One common way of tackling the properties of solids is to treat infinite systems by applying periodic boundary conditions (PBCs), i.e., neglecting edge effects and assuming an ideally ordered arrangement of the atoms. However, for reasons that will be discussed later in this chapter, it is practically impossible to accurately compute ZFS when PBCs are considered. Another approach, which is perfectly suited for the computation of local effects in ionic solids, consists of using an embedded cluster to model the material, as was done by Pradipto et al. to study cupric oxide [67] and Maurice et al. for $LiCu_2O_2$ [52]. The interested reader may consult more literature concerning this approach ([33, 34] and references therein), but we already stress that the methods and conclusions that are given in this chapter are directly applicable to solids provided that an embedded cluster approach is followed.

ZFS is often described in terms of model Hamiltonians, which have been (almost) always introduced phenomenologically. Such models are typically spin Hamiltonians, since by definition ZFS applies to systems for which the ground orbital configuration is separated in energy from the others, and hence, the effective description of the lowest-lying states can be restricted to the spin variables. For mononuclear complexes, these Hamiltonians only consider the spin anisotropy of the magnetic center. For polynuclear complexes, two main types of models are widely used, namely, the giant-spin and the multispin models. Both types will be discussed here for binuclear systems to keep the discussion as clear as possible.

This chapter aims at (i) making a bridge between ab initio calculations and model Hamiltonians to validate phenomenological Hamiltonians, (ii) establishing magneto-structural correlations, and (iii) demonstrating that magneto-structural correlations can also be understood in terms of crystal-field models. Therefore, in principle, one can bridge ab initio calculations, model Hamiltonians, and crystal-field models to get a full and intuitive picture of the ZFS in TM complexes. The chapter is organized as follows; first, we describe the ab initio methodologies that are sufficiently accurate to compute ZFS, as well as the effective Hamiltonian theory; second, we discuss the relevance of standard and improved model Hamiltonians that effectively describe the ZFS in mononuclear and binuclear complexes; third, we evaluate magneto-structural correlations, either derived from ab initio calculations performed on model complexes or analyzed within crystal-field model; finally, we conclude on the overall progress made in the last two decades and also give some perspectives.

Ab Initio Calculations and Effective Hamiltonians

In this part, we describe a common strategy to introduce relativistic effects in the calculation of ZFS, the typical way to introduce electron correlation in contracted spin-orbit configuration interaction (c-SOCI) and a way to bridge the resulting spin-orbit wave functions to model Hamiltonians to (i) assess the validity of model Hamiltonians and (ii) extract the model parameter values if appropriate. Apart from the here-discussed approach, other methods have been developed to compute ZFS and extract the model parameter values [6, 15, 25, 61–63, 66, 78], but these are out of the scope of this chapter.

From the Dirac Equation to Contracted Spin-Orbit Configuration Interaction

The time-independent Dirac equation may be written as:

$$\widehat{H}\psi = \left(c\,\hat{p}\hat{\alpha} + m_e c^2\hat{\beta} + \hat{V}\right)\psi = E\psi$$

where c is the speed of light, $\hat{p}$ is the momentum operator, $\hat{\alpha}$ and $\hat{\beta}$ are the Dirac 4×4 matrices, $\hat{V}$ is the one-electron external potential, and ψ is a four-component ($4c$) wave function. The resolution of the Dirac equation leads to two different types of solutions; the upper energy eigenfunctions are usually referred to as the "large" components, and the lower energy ones are known as the "small" components. Since the large components tend to standard spin orbitals in the nonrelativistic limit, these may also be considered as "essentially electronic" solutions. Although the introduction of relativistic effects is in principle more natural in $4c$ frameworks, much effort has been devoted in the last decades to reduce the complexity and derive accurate two-component ($2c$) models.

Various transformations/approximations have been implemented in standard codes, among which we quote the Douglas-Kroll (DK) transformation-based expansions [20, 31, 36] and the zeroth-order regular approximation (ZORA) approach [86, 87]. More recently, exact $2c$ formalisms (X2C) have been proposed [35]. Since many investigations of ZFS in TM complexes make use of the DK transformation, we will here describe briefly how to reach c-SOCI formalisms after this transformation, although c-SOCI schemes may also be derived from other reference $2c$ Hamiltonians.

The DK transformation can in principle lead to the exact energies of the Dirac Hamiltonian if one considers an infinite-order expansion of the one-electron external potential ($\hat{V}$). In practice, the expansion is of course limited to finite order; in most cases, the expansion only contains second-order terms. After the sign correction by Jansen and Hess in the original derivation of the transformation [36], the second-order (and higher) DK expansions are commonly referred to as the Douglas-Kroll-Hess (DKHn) Hamiltonians, where n is the expansion order. Another approximation, called the no-pair transformation, can be then introduced, in which the one-electron kinetic Hamiltonian matrix is diagonalized within a (finite) basis set to form a conventional one-electron basis. Although one can in principle use transformed one-electron and two-electron interactions, a one-component ($1c$) pseudo-relativistic Hamiltonian can be used in scalar relativistic calculations if only the spin-independent one-electron interactions are transformed [72], which means that nonrelativistic two-electron interactions are formally considered. A standard approximation consists of introducing the SOC after a nonrelativistic or scalar relativistic calculation. Assuming that a set of scalar relativistic reference states has been built in the first step, a c-SOCI calculation consists of diagonalizing $\widehat{H} = E_{el} + \widehat{H}_{\mathrm{SOC}}$ within the basis of the spin components of these reference SOF states, in which E_{el} are SOF electronic energies and $\widehat{H}_{\mathrm{SOC}}$ is an appropriate SOC Hamiltonian. More details concerning the choice of the SOF electronic energies will be given in the next section. Let us just mention that it is common practice nowadays to use electronic *energies* coming from a higher level of theory than the one that is used to compute the multireference *wave functions*. So in addition to using the reference wave functions for the off-diagonal SOC matrix elements of $\widehat{H}$, one can "dress" the diagonal elements of a c-SOCI matrix with higher-level electronic energies. Such an approach was proposed by [82] and also Llusar et al. [42].

Due to the local character of the spin-orbit operator, one may neglect interatomic SOCs. Expressions for the atomic SOC Hamiltonian that include one-electron and two-electron interactions adapted to the no-pair DKH2 Hamiltonian can be found elsewhere [72]. If one further applies a mean-field approximation to treat the two-electron part of the atomic SOC Hamiltonian by assuming an atomic one-electron density and adding the resulting mean-fields to the one-electron integrals, the so-called atomic mean-field integral (AMFI) approximation is used [32]. Note that alternative mean-field approximations of the SOC operator have been proposed by Neese [60] and that the atomic approximation can nowadays be avoided even in routine calculations.

The Treatment of the Spin-Orbit Free Electron Correlation

Although c-SOCI calculations can in principle be performed using single-reference SOF states, this is not an optimal approach to compute ZFS since it is in most cases impossible to converge enough excited SOF states to obtain a good representation of the SOC operator. Therefore, it is more appropriate to use multireference SOF states, which are obtained from multiconfigurational self-consistent field (MCSCF) approaches. A second requirement to compute ZFS is that all the spin components of the SOF state(s) of interest have to be coupled to excited components in a balanced way, i.e., one should not introduce a bias in the treatment of the lowest-energy spin-orbit states toward one specific spin-orbit component. This common-sense requirement has two implications in practice; (i) the set of SOF states considered in the first step of the calculation must be rather well thought, since selections based on an energy basis do not always lead to a consistent choice, and (ii) it is convenient to consider state-averaged (SA) orbitals between all the SOF states of interest to control the balance between the SOC excitations that are introduced a posteriori. Although SA orbitals can be calculated for any type of MCSCF multireference SOF states, the appropriate calculation of physical properties requires ensuring some properties of the *wave functions*, which may break down by any space truncation of the CI space that is used for computing the SOF states. This is why most researchers consider the complete active space self-consistent field (CASSCF) method [73] in the first step of a ZFS calculation.

Perhaps it is necessary at this stage to introduce the example of the nearly octahedral d^8 complexes to illustrate how one can consistently choose balanced active spaces and balanced SOCI spaces, i.e., an adequate set of SOF states in the first step of the calculation. If one only considers the d^8 manifold, ten spin-triplet (10T) and 15 spin-singlet (15S) SOF states can be at maximum consideration. Therefore, the easiest way to define balanced spaces consists of choosing an active space of eight electrons in five orbitals, i.e., a CAS(8/5), and in computing 10T and 15S SOF states with the state-averaged CASSCF (SA-CASSCF) method [71]. Note that it is easier to guess which active spaces and which sets of states can be suitable by looking at the reference, most symmetrical situation, i.e., in the octahedral case. One can for instance think of adding the two orbitals associated to the ligand-to-metal "σ" donation or to consider the ground SOF state plus a number of excited states that are consistent with the orbital degeneracies in that symmetry point group ("roots"), leading, for instance, to 4T, 7T, and 2S and 10T and 9S sets of SOF states [44]. In any case, the degeneracy of the first three spin-orbit roots should be strictly maintained in the octahedral situation. When the symmetry is lowered, which is necessary to observe ZFS, one should maintain a balanced treatment of the SOC operator. While it is clear that the active spaces that are consistent for the octahedral situation are transferable to the case of nearly octahedral complexes, the definition of the set of SOF states may be problematic. One can always recommend using the full set of states that can be formed within an 8/5 active space (i.e., 10T and 15S) or check that any of the previously mentioned subsets of it leads to similar averaged occupation numbers for the orbitals from which excitations are formed, i.e., for the three orbitals that correspond to the t_{2g} orbitals in the octahedral case. A similar

reasoning can be applied to any d^n configuration near an ideal geometry that leads to orbitally nondegenerate ground states, as, for instance, in nearly tetrahedral d^7 complexes. In this case, the "full" set of spin-orbit free states consists of ten spin-quartet (10Q) and 40 spin-doublet (40D) SOF states, while it is also safe to consider the 4Q and 7Q subsets [44].

Now that the set of SOF states has been defined, the SOC is computed between the spin components (i.e., M_S components) of these states to form the c-SOCI matrix. However, one needs to further discuss the choice of the SOF electronic energies that appear on the diagonal of the matrix. Since the spin components of the SA-CASSCF SOF states are considered to compute the off-diagonal elements of the c-SOCI matrix, it may appear natural to just consider the SA-CASSCF electronic energies on the diagonal of this matrix. This straightforward choice may however not be the wisest. In mononuclear complexes, satisfactory results may be obtained at this level, although it is also clear that results are in general slightly improved when the diagonal of the c-SOCI matrix is "dressed" with post-CASSCF correlated energies [44]. In binuclear complexes, ZFS can be severely underestimated if SA-CASSCF energies are considered on the diagonal elements of the c-SOCI matrix, as shown by Maurice et al. showed in a very detailed study concerning the ZFS of the first excited spin-triplet block in copper acetate monohydrate [50]. In this case, it was shown via a crystal-field model that the ZFS of interest relates at second order of perturbation to the isotropic magnetic couplings of the orbitally single-excited states. As shown by decades of experience, a proper description of isotropic magnetic couplings is not trivial at post-CASSCF levels [43]. Perturbative approaches such as the complete active space perturbation theory at second order (CASPT2) [4] and the n-electron valence state perturbation theory at second order (NEVPT2) [5] do not fully account for the effect of charge-transfer configurations on the isotropic couplings. Although the description is clearly improved compared to CASSCF, multireference configuration interaction (MRCI) has to be considered for quantitative results. Among the different MRCI schemes, the difference-dedicated configuration interaction (DDCI) methods [57] is one of the most successful ones for computing isotropic couplings. This method considers all the single and double excitations minus the double excitations that create two holes (h) in the inactive orbitals and two particles (p) in the virtual orbitals, usually referred to as the $2h2p$ excitations. Numerical examples will be given in sections "Mononuclear Complexes" and "Binuclear Complexes" for mononuclear and binuclear complexes, respectively, and the comparison with experimental data will be discussed.

On the Effective Hamiltonian Theory

The effective Hamiltonian theory [8, 18] enables one to establish a bi-univocal relation between a sophisticated Hamiltonian (here called the "reference" Hamiltonian) expressed in a large reference space and an effective Hamiltonian working on a truncated space, usually of much smaller dimension, that is called the "model" space. By definition, the eigenfunctions of an effective Hamiltonian are the wave functions of the reference Hamiltonian projected onto the model space, while its eigenvalues are

set to be identical to the energies of the reference Hamiltonian. Note that effective Hamiltonians can be used to design computational approaches [14, 24, 79, 80]. Alternatively, effective Hamiltonians can be used to extract information from wave functions and energies in order to determine the interactions of model Hamiltonians (i.e., the model operators and parameters). By doing so, computational chemistry can (i) assess the validity of the considered model Hamiltonian by checking that the model space is appropriate (the norms of the projections onto the model space must be large enough) and checking that the operators of the model Hamiltonian are the relevant ones (a good one-to-one correspondence between the model Hamiltonian and the effective Hamiltonian matrices must be obtained) and (ii) extract the model parameter values. This approach proved to be particularly appropriate in the field of ZFS for the second purpose, even if alternatives exist, as, for instance, the pseudo-spin approach of Chiboratu and Ungur [15].

Various formulations of effective Hamiltonians have been reported in the literature, among which the [8] and des Cloizeaux [18] ones that will be commented here. The Bloch formulation of the effective Hamiltonian is defined as:

$$\widehat{H}_{\text{Bloch}} = \sum_k |\tilde{\psi}_k\rangle E_k \langle S^{-1}\tilde{\psi}_k|$$

where $\tilde{\psi}_k$ and E_k are the kth projected eigenvector and eigenvalue of the reference Hamiltonian. The projected eigenvectors of the reference Hamiltonian are mutually non-orthogonal, and S^{-1} is the inverse of the overlap matrix between the projected eigenvectors. However, one should note that this formulation does not ensure hermiticity of the effective Hamiltonian, which may be problematic for extracting model parameter values. In order to get Hermitian model Hamiltonians, the des Cloizeaux formalism, which symmetrically orthogonalizes the projected eigenvectors (usually referred to as the Löwdin's orthogonalization in quantum chemistry), can be used. In this formalism, the effective Hamiltonian is defined as:

$$\widehat{H}_{\text{des Cloizeaux}} = \sum_k \left|S^{-\frac{1}{2}}\tilde{\psi}_k\right\rangle E_k \left\langle S^{-\frac{1}{2}}\tilde{\psi}_k\right|$$

Examples of discussions on the validity of model Hamiltonians and extractions of model parameter values are given in the next section.

Model Hamiltonians and Effective Hamiltonians

Model Hamiltonians not only reduces the complexity of a given reference Hamiltonian, the model space being always smaller than the reference space, but also introduce effective parameters with a well-defined physical meaning. In the field of molecular magnetism, the typical simpler Hamiltonians are the well-known phenomenological spin Hamiltonians, in which no track of the orbital part of the wave functions is kept. Although the earliest spin Hamiltonian is the Heisenberg-

Dirac-van Vleck (HDV) [19, 29, 88] one (vide infra), the expression "spin Hamiltonian" was introduced later by Abragam and Pryce [2] in the context of the electron paramagnetic resonance (EPR) spectroscopy. In this part, we will show that (i) the *effective Hamiltonian theory* can be used to project the information contained in c-SOCI wave functions onto a model space consisting of the spin components of one (or several) spin state(s), (ii) question the validity of *model Hamiltonians* and improve them if necessary, and (iii) show that model parameter values in good agreement with experiment can be obtained. We will start by discussing mononuclear complexes prior to dealing with binuclear systems to gradually increase the complexity of the models to be introduced. Note that a basic introduction to the use of effective Hamiltonian theory in relation to model Hamiltonians can be found in a recent textbook [17].

Mononuclear Complexes

Mononuclear complexes with a d^n electronic configuration can be split in two groups, $S = 1$ or $S = 3/2$ systems and $S = 2$ or $S = 5/2$ systems, as the model Hamiltonians that have to be used to describe the full complexity of their ZFS are different. Note that, in a first approximation, the model Hamiltonian for $S = 1$ or $S = 3/2$ systems can be used to describe $S = 2$ or $S = 5/2$ systems. Although this is commonly done in the literature, this approximation may not always be adequate, as will be discussed later. One should also stress that usually, in the $S = 2$ or $S = 5/2$ systems, the SSC can contribute to a significant part of the total ZFS, around 10 % of the total ZFS in manganese(III) complexes [21] and even up to 20 % in manganese(II) complexes [90]. Therefore, one should account for the SSC in the determination of ZFS parameters [61]. A remark is thus worth here. Both SOC and SSC generate second-rank ZFS tensors, and, unless these axes are imposed by symmetry, both effects may independently generate different magnetic axes. Therefore, one should in principle not only perform a c-SOCI calculation but rather diagonalize $\widehat{H} = E_{el} + \widehat{H}_{\mathrm{SOC}} + \widehat{H}_{\mathrm{SSC}}$ prior to applying the effective Hamiltonian theory. This point is also valid for the $S = 1$ and $S = 3/2$ systems, although $\widehat{H}_{\mathrm{SSC}}$ can be more safely neglected in these cases, especially in complexes for which large ZFSs are observed. For mononuclear complexes, we compare computed values to experimental ones when the SSC contribution to the ZFS can be neglected. When this contribution is expected to play a more important contribution, we do not compare to experiment but rather focus on the SOC contribution and on the validity of the model Hamiltonians.

$S = 1$ and $S = 3/2$ Systems

For $S = 1$ and $S = 3/2$ systems, the model Hamiltonian which describes the ZFS of orbitally nondegenerate states is simply [39]:

$$\widehat{H}_{\mathrm{mod1}} = \hat{S}\overline{\overline{D}}\hat{S}$$

Table 1 Analytical interaction matrix corresponding to the ZFS of $S = 1$ systems [44]. X, Y, and Z correspond to the Cartesian axes of an arbitrary axis frame

$\widehat{H}_{\text{mod1}}$	$\|1,-1\rangle$	$\|1,0\rangle$	$\|1,1\rangle$
$\langle 1,-1\|$	$\frac{1}{2}(D_{XX}+D_{YY})+D_{ZZ}$	$-\frac{\sqrt{2}}{2}(D_{XZ}+iD_{YZ})$	$\frac{1}{2}(D_{XX}-D_{YY})+iD_{XY}$
$\langle 1,0\|$	$-\frac{\sqrt{2}}{2}(D_{XZ}-iD_{YZ})$	$D_{XX}+D_{YY}$	$\frac{\sqrt{2}}{2}(D_{XZ}+iD_{YZ})$
$\langle 1,1\|$	$\frac{1}{2}(D_{XX}-D_{YY})+iD_{XY}$	$\frac{\sqrt{2}}{2}(D_{XZ}-iD_{YZ})$	$\frac{1}{2}(D_{XX}+D_{YY})+D_{ZZ}$

where $\hat{S}$ is the spin operator row or column vector and $\overline{\overline{D}}$ the second-rank ZFS tensor. $\overline{\overline{D}}$ is symmetric and only composed of real numbers. Expanding this Hamiltonian and applying it to the $|S, M_S\rangle$ spin component basis allows one to derive the analytical interaction matrix to be compared to the effective Hamiltonian matrix. Although a similar reasoning can be used for $S = 3/2$ systems, we will only discuss in details the $S = 1$ case. Note that the analytical interaction matrix for $S = 3/2$ systems is available elsewhere [44]. The analytical interaction matrix is given in Table 1 for $S = 1$. As mentioned earlier, the analytical interaction matrix is Hermitian (see Table 1), as any other analytical interaction matrix that will be discussed in this chapter.

The c-SOCI calculation delivers wave functions expressed in terms of the spin components of a set of SOF states. To describe the ZFS of an $S = 1$ system with $\widehat{H}_{\text{mod1}}$, the c-SOCI eigenvectors have to be projected onto the spin components of a SOF state, typically the ground state. The norm of the projection can be assessed by looking at the diagonal elements of the overlap matrix between the projected eigenvectors (S in $\widehat{H}_{\text{Bloch}}$ or in $\widehat{H}_{\text{des Cloizeaux}}$). If the (ground) SOF state is well separated in energy from any other SOF state, the norm of the projections is expected to be close to 1. Let us choose the example of the $[\text{Ni(HIM2-Py)}_2\text{NO}_3]^+$ (HIM2-py = 2-(2-pyridyl)-4,4,5,5-tetramethyl-4,5-dihydro-1H-imidazolyl-1-hydroxy) complex (see Fig. 1) [44].

The projected wave functions are obtained from the ab initio ones simply by truncation, i.e. by only the part that concerns the spin components of the Ψ_0 ground SOF states:

$$
\begin{aligned}
|\tilde{\Psi}_1\rangle &= (0.045 + i0.092)\,|\Psi_0,-1\rangle - (0.668 - i0.724)\,|\Psi_0,0\rangle \\
&\quad + (0.096 + i0.037)\,|\Psi_0,1\rangle \\
|\tilde{\Psi}_2\rangle &= -(0.395 - i0.578)\,|\Psi_0,-1\rangle + (0.062 + i0088)\,|\Psi_0,0\rangle \\
&\quad - (0.096 - i0.173)\,|\Psi_0,1\rangle \\
|\tilde{\Psi}_3\rangle &= (0.701 + i0.026)\,|\Psi_0,-1\rangle - (0.090 + i0.037)\,|\Psi_0,0\rangle \\
&\quad - (0.519 + i0.472)\,|\Psi_0,1\rangle
\end{aligned}
$$

The norms of the projected vectors (prior to orthonormalization) are all larger than 0.99, which perfectly legitimates the use of a spin Hamiltonian in this case.

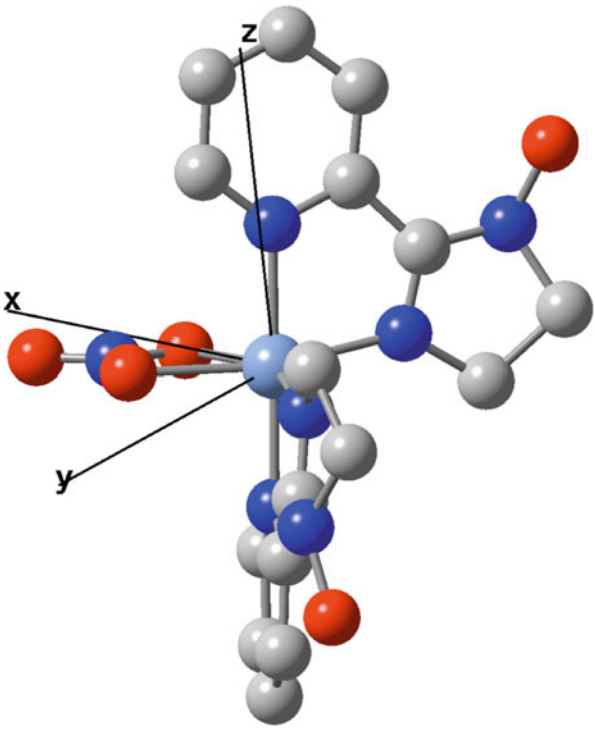


Fig. 1 Ball-and-stick representation of a model of the $[Ni(HIM2\text{-}Py)_2NO_3]^+$ complex and its main magnetic axes [44]. The "external" methyl groups have been modelled by hydrogen atoms; all hydrogen atoms are omitted for clarity

Table 2 Effective interaction matrix corresponding to the ZFS of the $[Ni(HIM2\text{-}Py)_2NO_3]^+$ complex [44]

$\widehat{H}_{\text{des Cloizeaux}}$	$\lvert\Psi_0,-1\rangle$	$\lvert\Psi_0,0\rangle$	$\lvert\Psi_0,1\rangle$
$\langle\Psi_0,-1\rvert$	6.386	$-0.690+i0.376$	$-3.734+i3.134$
$\langle\Psi_0,0\rvert$	$-0.690-i0.376$	0.125	$0.690-i.0376$
$\langle\Psi_0,1\rvert$	$-3.734-i3.134$	$0.690+i.0376$	6.386

However, prior to validating $\widehat{H}_{\text{mod1}}$, other tests are necessary: one must show that (i) the effective and analytical interaction matrices match and (ii) how the extracted tensor component values transform with respect to a change of the axis frame (i.e., that the extracted $\overline{\overline{D}}$ actually transforms as a tensor). The effective interaction matrix that is built with $\widehat{H}_{\text{des Cloizeaux}}$ is represented in Table 2 ($E_1 = 0.00$, $E_2 = 1.529$ and $E_3 = 11.396$, all energies being in cm^{-1}).

By construction, the effective interaction matrix is Hermitian and has the same eigenvalues as the reference Hamiltonian ($\widehat{H}_{\text{ref}} = E_{el} + \widehat{H}_{\text{SOC}}$), and its eigenvectors $\tilde{\Psi}_1$, $\tilde{\Psi}_2$, and $\tilde{\Psi}_3$ are identical to the projected ab initio eigenvectors up to a given complex phase factor. By the term-by-term comparison of $\widehat{H}_{\text{mod1}}$ and $\widehat{H}_{\text{des Cloizeaux}}$, it is immediately clear that both matrices perfectly match, meaning that $\widehat{H}_{\text{mod1}}$ is suited to describe the ZFS in this system, as it turns out to be the case for any $S = 1$ system with an orbitally nondegenerate ground state. Therefore, the second-rank ZFS tensor can be unambiguously extracted. Diagonalization of this tensor leads

to the determination of the magnetic axes X^m, Y^m, and Z^m, as well as the ZFS parameters:

$$D = D_{Z^m Z^m} - \frac{1}{2}\left(D_{X^m X^m} + D_{Y^m Y^m}\right)$$

$$E = \frac{1}{2}\left(D_{X^m X^m} - D_{Y^m Y^m}\right)$$

provided that conventions are applied, i.e., $|D| > 3E$ and $E > 0$ (or, alternatively $E/D > 0$). If one uses the transformation matrix P^{-1} that allows expressing $\overline{\overline{D}}$ in the magnetic axis frame (such that $\overline{\overline{D}}^m = P^{-1}\overline{\overline{D}}P$), one can build again $\widehat{H}_{\text{des Cloizeaux}}$ after computing the c-SOCI solutions in this coordinate system and show that the extracted tensor is diagonal and finally that the same ZFS parameters can be extracted. Therefore, we conclude that $\overline{\overline{D}}$ actually transforms as a second-rank tensor and show that the model Hamiltonian $\widehat{H}_{\text{mod1}}$ is fully valid. In this case, the extracted values for D and E are -10.60 and $0.76\,\text{cm}^{-1}$, respectively, [44] and compare well with the most accurate experimental values (-10.15 and $0.10\,\text{cm}^{-1}$, respectively, from high-field and high-frequency EPR spectroscopy [70]). Other nickel(II) complexes have been studied in a similar way, and a good agreement between theory and experiment is generally observed with c-SOCI [16, 44, 51, 75, 76], while density functional theory methods seem to fail for this high-spin d^8 configuration [41].

A similar reasoning can be applied to $S = 3/2$ complexes, such as nearly tetrahedral cobalt(II) complexes. The D and E parameters cannot be extracted from the eigenvalues of any reference Hamiltonian, since one only has access to the energy difference between the two Kramer's doublets of interest. On the contrary, the application of the effective Hamiltonian theory unambiguously allows extracting the full ZFS tensor, i.e., determining the magnetic axes and the ZFS parameters. As shown elsewhere [44], $\widehat{H}_{\text{mod1}}$ is indeed also perfectly suited to describe the ZFS of $S = 3/2$ ground states. As an example of application, let us consider the $[Co(PPh_3)2Cl_2]$ (Ph = phenyl) complex (see Fig. 2) [44]. In this case, the extracted ZFS parameter values are -14.86 and $0.54\,\text{cm}^{-1}$ for D and E, respectively [44], which also compares well to the experimental values of -14.76 and $1.14\,\text{cm}^{-1}$, respectively [40]. Also, note that the energy difference from the two Kramer's doublet of interest relates to the D and E parameters as follows:

$$\Delta E = 2\sqrt{D^2 + 3E^2}$$

In both nickel(II) and cobalt(II) examples described above, the computational methodology was based on SA-CASSCF calculations with quite large active spaces, small sets of SOF states (4T and 7Q, respectively), and CASPT2 correlated energies [44]. As a recommendation, one may note that enlarging the set of SOF states does not systematically improve the results: on the one hand, more SOC excitations are accounted for, and on the other hand, averaging artifacts are introduced, which may result, for instance, in a poor orbital set for describing the ground state. Therefore,

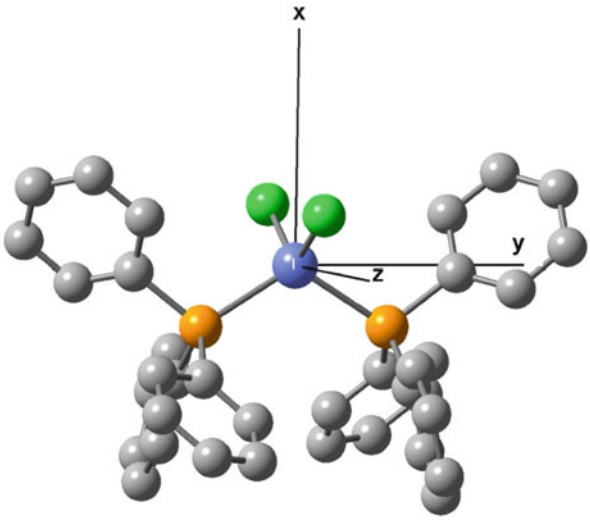


Fig. 2 Ball-and-stick representation of the $[Co(PPh_3)2Cl_2]$ complex and its main magnetic axes [44]. All the hydrogen atoms are omitted for clarity

one should make the compromise of describing as best as possible the stronger SOC interactions, i.e., the balance between the number of possible excitations and the accurate description of the involved states. As a conclusion concerning the ab initio methodology, one may say that c-SOCI methods do not constitute a "black-box machinery" to compute ZFSs.

$S = 2$ and $S = 5/2$ Systems

For $S = 2$ and $S = 5/2$ complexes, additional operators must be introduced in the model Hamiltonian to achieve a complete description of the ZFS:

$$\widehat{H}_{\text{mod2}} = \hat{S}\overline{\overline{D}}\hat{S} + \sum_{q=-4}^{4} B_4^q \widehat{O}_4^q$$

where q may be odd and where the $\widehat{O}_4^q$ operators are extended Stevens operators [1, 3, 77, 81]. This Hamiltonian is valid in any arbitrary axis frame for $S = 2$ and $S = 5/2$ complexes. When the ground SOF state is well separated in energy from the excited states, the fourth-rank spin $\widehat{O}_4^q$ operators have a very small effect on the effective interactions of the model $\widehat{H}_{\text{mod1}}$. Therefore, one can first find the main anisotropy axes by extracting $\overline{\overline{D}}$ from the comparison of the effective interaction matrix and the analytical one that is obtained with $\widehat{H}_{\text{mod1}}$ and then compute the effective interaction matrix in this frame after a second c-SOCI calculation. In this case, the model Hamiltonian reduces to:

$$\widehat{H}_{\text{mod3}} = \sum_{n=0}^{4}\sum_{k=2}^{4} B_k^n \widehat{O}_k^n$$

where k and n must be even and the $\widehat{O}_k^n$ operators are standard Stevens operators. Note that $B_2^0 = D/3$ and that $B_2^2 = E$, i.e., B_2^0 is a second-rank axial ZFS parameter, while B_2^2 is a rhombic one. As shown elsewhere, this two-step procedure leads to the unambiguous extraction of the main magnetic axes and of the five B_k^n parameters (B_2^0, B_2^2, B_4^0, B_4^2, and B_4^4) [45]. Indeed, in the magnetic axis frame, the effective interaction matrix is in almost perfect correspondence with the analytical one derived for $\widehat{H}_{\text{mod3}}$ with only some negligible deviations, typically not larger than 0.01 cm^{-1}. These deviations can be considered as numerical noise and do not significantly alter the extracted B_k^n values. Note that if one wants to neglect the B_4^n parameters, i.e., introduce only second-rank spin operators in the model Hamiltonian, as in $\widehat{H}_{\text{mod1}}$, it is important to check a priori that the $k = 4$ terms are not important to insure that the model and effective interaction matrices match. As will be discussed in section "Magneto-Structural Correlations Based on Crystal-Field Models and Ab Initio Calculations", the fourth-rank Stevens parameters relate to the near-degeneracy of spin components of different SOF states. Therefore, one should not neglect them in such situations, as, for instance, in nearly octahedral manganese(III) complexes [45]. Note that, however, the Jahn-Teller effect tends to largely remove the near-degeneracy between the two lowest orbital configurations in this case, which explains, for instance, why the B_4^n parameters are not crucial to describe the ZFS in the [γ-Mn(acac)$_3$] (acac = acetylacetonato) complex (see Fig. 3) [49], for which the first coordination sphere is tetragonally elongated.

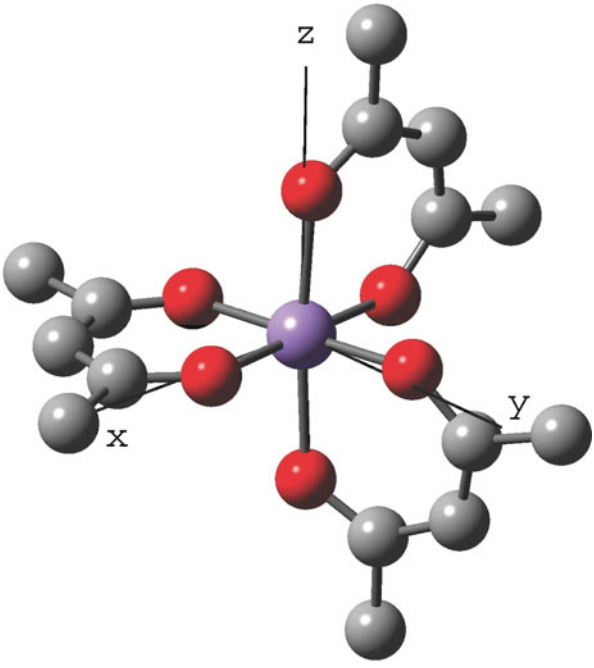


Fig. 3 Ball-and-stick representation of the [γ-Mn(acac)$_3$] complex and its main magnetic axes [49]. All the hydrogen atoms are omitted for clarity

In many cases of experimental interest, $\widehat{H}_{\text{mod1}}$ is perfectly suited to describe the ZFS of $S = 2$ and $S = 5/2$ complexes.

Binuclear Complexes

Prior to introducing the model Hamiltonians that can be used to describe the ZFS in binuclear complexes, it is worth introducing the HDV Hamiltonian [19, 29, 88], which may be expressed in a "multispin" picture, i.e., by considering local spin operators that are to be applied within the basis of local spin components, i.e., within the uncoupled spin basis:

$$\widehat{H}_{\text{HDV}}^{\text{uncoupled}} = J\,\hat{S}_a \cdot \hat{S}_b$$

where J is the isotropic coupling constant and $\hat{S}_a$ and $\hat{S}_b$ are spin operator column vectors. Note that various expressions coexist in the literature, depending on a factor that is applied to this Hamiltonian (here 1, but one may find -1 or more often -2). It can easily be shown that $\widehat{H}_{\text{mod1}}$ can be also written in terms of spin operators that lead to a diagonal analytical interaction matrix if one works within the basis of spin eigenfunctions, i.e., within the coupled spin basis:

$$\widehat{H}_{\text{HDV}}^{\text{coupled}} = \frac{J}{2}\sum_{S}(\hat{S}^2 - \hat{S}_a^2 - \hat{S}_b^2)$$

where S ranges between $|S_a - S_b|$ and $S_a + S_b$, $\hat{S}$ is the spin operator associated to each coupled spin state, and S_a and S_b are the local spins on the a and b sites. Therefore, it is clear that $\widehat{H}_{\text{HDV}}$ splits the coupled spin states, which may further be the subject of ZFS and mixings when anisotropic effective interactions are considered in the model Hamiltonian. Such ZFSs and "spin mixings" can be effectively described in two different ways that work in the coupled and uncoupled basis, respectively, and which are classified as "giant-spin" or "block-spin" models and "multispin" ones.

Giant-Spin and Block-Spin Hamiltonians

If the isotropic coupling constantly plays a much more important role on the effective interaction matrix than the spin mixings, which is usually referred to as the "strong-exchange limit" [9], very simple models can be used to describe the low-energy spectrum. If only one spin "block" has to be described and that its magnetic axis frame is considered, one can use a very simple giant-spin Hamiltonian that is similar to $\widehat{H}_{\text{mod3}}$:

$$\widehat{H}_{\text{giant spin}}^{\text{coupled}} = \sum_{n=0}^{k}\sum_{k=2}^{x} B_k^n \hat{O}_k^n$$

where $x = 2S$ if S is even or $x = 2S - 1$ if S is odd and k and n are even.

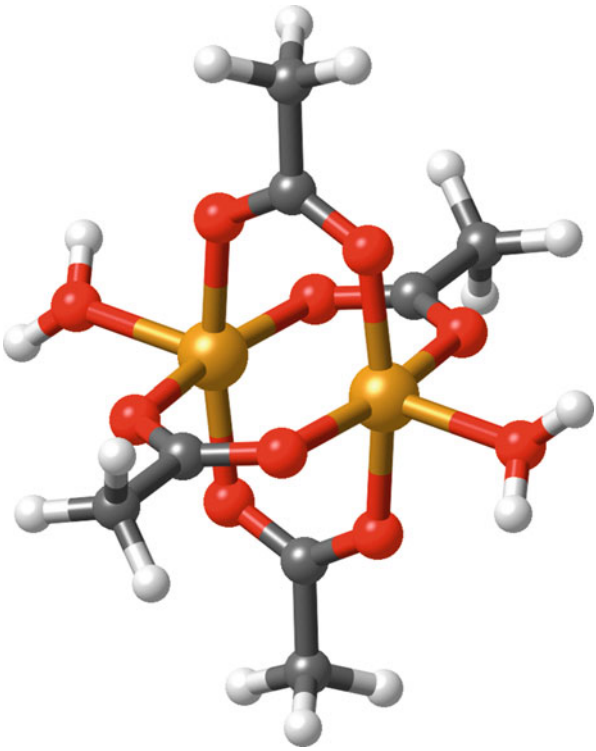


Fig. 4 Ball-and-stick representation of copper acetate monohydrate [50]

The simplest case is the ZFS of an $S = 1$ spin state resulting from the coupling of two local spins $S_a = S_b = 1/2$. One typical example of such a situation is copper acetate monohydrate (see Fig. 4) [50].

In this system, it is crucial to account for both SOC and SSC to compute the ZFS of the excited $^3A_{1u}$ SOF state. As shown by Maurice et al. [50], the treatment of SOC requires a special attention to the correlated energies appearing on the diagonal elements of the $\widehat{H} = E_{el} + \widehat{H}_{\mathrm{SOC}} + \widehat{H}_{\mathrm{SSC}}$ matrix (see Table 3). The reference wave functions were obtained with SA-CASSCF(18/10) calculations (see [50] for more details). As can be seen in Table 3, second-order perturbation theory does not describe sufficiently well the SOF excitation energies, as well as variational approaches with limited configuration interaction spaces (the DDCI1 calculations include in this case only the $1h$ and $1p$ excitations, while the DDCI2 ones also account for the $1h1p$, $2h$, and $2p$ excitations). Although the DCCI2 level appears to lead to an almost converged value of D when only $\widehat{H}_{\mathrm{SOC}}$ is considered, it is essentially due to error cancellations (more details are given on this in [50]). The best value, obtained with the DDCI3 energies, is in exceptional agreement with experiment ($D_{\mathrm{expt.}} = -0.335\,\mathrm{cm}^{-1}$ [65]). The rhombic E parameter value is very small, $0.006\,\mathrm{cm}^{-1}$ with the DDCI3 energies (in good agreement with $E_{\mathrm{expt.}} = 0.01\,\mathrm{cm}^{-1}$), and will not be discussed in details here.

Table 3 Computed axial ZFS D parameters as a function of the correlated energies used on the diagonal of the contracted configuration interaction matrices and of the operators introduced in the Hamiltonian [50]

E_{el}	$\widehat{H}_{\mathrm{SOC}}$	$\widehat{H}_{\mathrm{SSC}}$	$\widehat{H}_{\mathrm{SOC}} + \widehat{H}_{\mathrm{SSC}}$
SA-CASSCF	−0.017	−0.118	−0.137
NEVPT2	−0.026	−0.118	−0.144
DDCI1	0.005	−0.118	−0.115
DDCI2	−0.172	−0.118	−0.291
DDCI3	−0.200	−0.118	−0.319

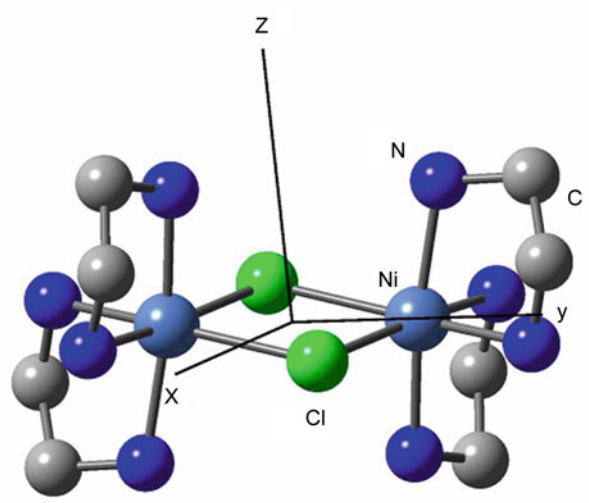


Fig. 5 Ball-and-stick representation of the $[Ni_2(en)_4Cl_2]^{2+}$ complex and its $S = 2$ block main magnetic axes. All the hydrogen atoms are omitted for clarity

One should stress that in this particular case, i.e. the d^9–d^9 configuration, computing the ZFS of the triplet block happened to be particularly challenging, but it may not be the case for other configurations. Another system that has been studied within the giant-spin approach is the $[Ni_2(en)_4Cl_2]^{2+}$ (en = ethylenediamine) complex (see Fig. 5) [46, 47]. Although this system does not fall within the strong-exchange limit, it is possible to build two effective Hamiltonians in the basis of the spin components of the $S = 2$, $S = 1$, and $S = 0$ spin components, one in which spin mixings are set to zero and one in which these spin mixings are allowed [47, 49]. From this theoretical study, it was shown that, in the magnetic axis frame, $\widehat{H}^{\mathrm{coupled}}_{\mathrm{giantspin}}$ can describe the ZFS of both the $S = 2$ and $S = 1$ blocks in the absence of spin mixing and that the spin mixings can be described using additional operators [47] (in this case, due to a symmetry center, the spin mixings concern the $S = 0$ and some $S = 2$ spin components). In other words, one can define a block spin Hamiltonian which describes the isotropic coupling and the ZFSs of the different blocks in the absence of spin mixing, and the spin-mixing effects can actually be introduced inside the different spin blocks, i.e.,

one can use a block-diagonal analytical interaction matrix to describe the entire low-energy spectrum [49]. The comparison with experimental data is complicated since all the studies reported so far neglected some, and usually different, effective interactions in the model Hamiltonians [27, 30, 37, 38]. One can just mention that the best computed value for D_2, i.e., the axial ZFS parameter of the $S = 2$ block, is in good semiquantitative agreement with the experimental one [30] (-3.0 vs.$-1.8\,\mathrm{cm}^{-1}$, respectively).

Another study of the d^8–d^8 configuration, related to the strong-exchange limit, concerned model complexes [74]. It was shown that, contrary to what is often proposed, no simple relations appear between the ZFS parameters of the $S = 2$ and $S = 1$ blocks, which can also be analyzed within a multispin picture (vide infra). Also, it is clear that giant-spin and block-spin Hamiltonians may not be relevant in the weak-exchange limit, i.e., when J becomes negligible, since spin mixings cannot be considered as a perturbation in such a case. In principle, one should in this case consider a multispin model, the extraction of which is far from being straightforward, as will be shown later.

Multispin Hamiltonians

As in section "Mononuclear Complexes", we will introduce progressively the complexity of the multispin Hamiltonian in binuclear complexes. Let us start with the easiest case of two coupled $S = 1/2$ centers, as in the d^9–d^9 configurations, for instance. The model Hamiltonian which is commonly used includes an isotropic coupling terms ($\widehat{H}_{\mathrm{HDV}}$) plus an anisotropy tensor [39]:

$$\widehat{H}^{\text{uncoupled}}_{\text{multispin}(S_a=S_b=1/2)} = J\,\hat{S}_a \cdot \hat{S}_b + \hat{S}_a \overline{\overline{T}}_{ab} \hat{S}_b = J\,\hat{S}_a \cdot \hat{S}_b + \hat{S}_a \overline{\overline{D}}_{ab} \hat{S}_b + \bar{d}_{ab} \hat{S}_a \times \hat{S}_b$$

where $\overline{\overline{T}}_{ab}$ is a second-rank tensor that is neither symmetric nor antisymmetric in the general case, $\overline{\overline{D}}_{ab}$ is the symmetric anisotropy exchange tensor, and $\bar{d}_{ab}$ is the Dzyaloshinskii-Moriya term [22, 58] that can also be referred to as the antisymmetric exchange pseudo-vector. As mentioned earlier, computing $\overline{\overline{D}}_{ab}$ may turn into a real nightmare [50], but the semiquantitative determination of $\bar{d}_{ab}$ is much less demanding, since it can be obtained from CASSCF(2/2) + c-SOCI calculations [48, 67]. The analytical interaction matrix built in the uncoupled basis is represented in Table 4.

One can also express this analytical interaction matrix with the $\overline{\overline{D}}_{ab}$ tensor and the $\bar{d}_{ab}$ pseudo-vector by using the $D_{ii} = T_{ii}$, $D_{ij} = 1/2(T_{ij} + T_{ji})$, $d_X = 1/2(T_{YZ} - T_{ZY})$, $d_Y = 1/2(T_{ZX} - T_{XZ})$, and $d_Z = 1/2(T_{XY} - T_{YX})$ relations [48]. The analytical interaction matrix can be transformed into the coupled basis as follows:

$$\widehat{H}^{\text{coupled}}_{\text{multispin}} = U^T \widehat{H}^{\text{uncoupled}}_{\text{multispin}} U$$

Table 4 Analytical interaction matrix corresponding to $\widehat{H}^{\text{uncoupled}}_{\text{multispin}}$ when $S_a = S_b = 1/2$ [48]. A shortened notation $|M_{S_a}, M_{S_b}\rangle$ is used for the $|S_b, M_{S_a}, S_b, M_{S_b}\rangle$ uncoupled functions. X, Y, and Z correspond to the Cartesian axes of an arbitrary axis frame

$\widehat{H}^{\text{uncoupled}}_{\text{multispin}}$	$\lvert-1/2,-1/2\rangle$	$\lvert-1/2,1/2\rangle$	$\lvert1/2,-1/2\rangle$	$\lvert1/2,1/2\rangle$
$\langle-1/2,-1/2\rvert$	$\frac{1}{4}(J+T_{ZZ})$	$-\frac{1}{4}(T_{ZX}+iT_{ZY})$	$-\frac{1}{4}(T_{XZ}+iT_{YZ})$	$\frac{1}{4}[T_{XX}-T_{YY}+i(T_{XY}+T_{YX})]$
$\langle-1/2,1/2\rvert$	$-\frac{1}{4}(T_{ZX}-iT_{ZY})$	$-\frac{1}{4}(J+T_{ZZ})$	$\frac{1}{4}[2J+T_{XX}+T_{YY}+i(T_{YX}-T_{XY})]$	$\frac{1}{4}(T_{XZ}+iT_{YZ})$
$\langle1/2,-1/2\rvert$	$-\frac{1}{4}(T_{XZ}-iT_{YZ})$	$\frac{1}{4}[2J+T_{XX}+T_{YY}-i(T_{YX}-T_{XY})]$	$-\frac{1}{4}(J+T_{ZZ})$	$\frac{1}{4}(T_{ZX}+iT_{ZY})$
$\langle1/2,1/2\rvert$	$\frac{1}{4}[T_{XX}-T_{YY}-i(T_{XY}+T_{YX})]$	$\frac{1}{4}(T_{XZ}-iT_{YZ})$	$\frac{1}{4}(T_{ZX}-iT_{ZY})$	$\frac{1}{4}(J+T_{zz})$

where U^T is the transpose of the change of basis matrix U. The matrix elements of U are given by the appropriate Clebsch-Gordan coefficients [9]. When $\widehat{H}_{\text{multispin}}$ is expressed in the coupled basis, the $\overline{\overline{D}}_{ab}$ tensor relates to the splitting and mixing of the $S = 1$ block (i.e., its ZFS in the strong-exchange limit), while the $\bar{d}_{ab}$ pseudo-vector introduces spin mixings between the $S = 1$ spin components and the $S = 0$ one. This type of S/S+1 spin mixing is not always symmetry allowed (the symmetry rules are available elsewhere [12]). One may just recall that if the system contains a symmetry center, as it is the case for copper acetate monohydrate, the $\bar{d}_{ab}$ pseudo-vector is null [58]. Therefore, in this system, as in any other d^9–d^9 binuclear system, a simple relation appears between $\overline{\overline{D}}_1$ and $\overline{\overline{D}}_{ab}$ [50]:

$$\overline{\overline{D}}_1 = \frac{1}{2}\overline{\overline{D}}_{ab}$$

By studying model copper(II)-copper(II) complexes, Maurice et al. showed that $\widehat{H}_{\text{multispin}}$ is perfectly valid to describe the isotropic coupling and the ZFSs when $S_a = S_b = 1/2$ [48]. Owing to the effective Hamiltonian theory, all the model parameter values can be theoretically extracted, while it appears complicated to properly distinguish between the symmetric and antisymmetric exchange terms from the outcomes of experiments. Moreover, since one can obtain good semiquantitative estimates of the $\bar{d}_{ab}$ pseudo-vector components from CASSCF(2/2) + c-SOCI calculations [48, 67], antisymmetric exchange is essentially due to the direct SOC between the $S = 1$ and $S = 0$ spin components.

When $S_a = 1$ and $S_b = 1/2$, another term must be added to the phenomenological model Hamiltonian, related to the single-ion anisotropy of site a [39]:

$$\widehat{H}^{\text{uncoupled}}_{\text{multispin}(S_a=1,S_b=1/2)} = J\,\hat{S}_a \cdot \hat{S}_b + \hat{S}_a\overline{\overline{D}}_a\hat{S}_a + \hat{S}_a\overline{\overline{D}}_{ab}\hat{S}_b + \bar{d}_{ab}\hat{S}_a \times \hat{S}_b$$

In this case, both the $\overline{\overline{D}}_a$ and $\overline{\overline{D}}_{ab}$ symmetric tensors affect the $S = 3/2$ block, while $\bar{d}_{ab}$ relates to the $S/S + 1$ spin mixings, as usual.

The situation is drastically complicated when considering the $S_a = S_b = 1$ case, e.g. nickel(II)-nickel(II) complexes. The following model Hamiltonian was used for decades to interpret experimental data [27, 30, 37, 38]:

$$\widehat{H}^{\text{uncoupled}}_{\text{multispin}(S_a=S_b=1)} = J\,\hat{S}_a \cdot \hat{S}_b + \hat{S}_a \overline{\overline{D}}_a \hat{S}_a + \hat{S}_b \overline{\overline{D}}_b \hat{S}_b + \hat{S}_a \overline{\overline{D}}_{ab} \hat{S}_b + \bar{d}_{ab} \hat{S}_a \times \hat{S}_b$$

The validity of this model Hamiltonian was assessed by c-SOCI calculations and the effective Hamiltonian theory in 2010 by Maurice et al. [46]. It was shown that many terms of the effective interaction matrix which were obtained for the $[Ni_2(en)_4Cl_2]^{2+}$ complex were not associated to any model parameter in the analytical interaction matrix. In order to reproduce all features of the effective Hamiltonian matrix, one must actually introduce a symmetric fourth-rank exchange tensor, $\boldsymbol{D}_{aabb}$, in the model Hamiltonian, leading to:

$$\begin{aligned}\widehat{H}^{\text{uncoupled}}_{\text{multispin}(S_a=S_b=1)} = {} & J\,\hat{S}_a \cdot \hat{S}_b + \hat{S}_a \overline{\overline{D}}_a \hat{S}_a + \hat{S}_b \overline{\overline{D}}_b \hat{S}_b + \hat{S}_a \overline{\overline{D}}_{ab} \hat{S}_b \\ & + \hat{S}_a \otimes \hat{S}_a \boldsymbol{D}_{aabb} \hat{S}_b \otimes \hat{S}_b\end{aligned}$$

if $\bar{d}_{ab}$ is null. The extraction of the $\boldsymbol{D}_{aabb}$ components is not straightforward even with all information contained in the 9×9 effective interaction matrix; one should thus consider relations between these components, as done in a study of model complexes [74]. One should thus stress that the spin mixings between the $S = 2$ and $S = 0$ spin components, mentioned in section "Giant-Spin and Block-Spin Hamiltonians", can be interpreted in terms of the parameters of the multispin Hamiltonian after transforming the analytical interaction matrix to the coupled basis: these terms actually relate to all symmetric tensors of $\widehat{H}_{\text{multispin}}$, i.e., $\overline{\overline{D}}_a$, $\overline{\overline{D}}_b$, $\overline{\overline{D}}_{ab}$, and $\boldsymbol{D}_{aabb}$ [47].

Another interesting point which is worth mentioning here is that one may be interested in computing only the local anisotropy tensors, i.e., $\overline{\overline{D}}_a$ and $\overline{\overline{D}}_b$. Various strategies exist; (i) one may replace one of the two magnetic centers by a model potential [46] or by a diamagnetic ion [11, 46, 53], (ii) one may also consider its lowest-energy closed-shell configuration [74], and (iii) another method considers local excitations while keeping the other site in its lowest-energy open-shell configuration [69]. This last strategy is meant to be the most accurate approach. From the study of model complexes, it was shown that the local anisotropy parameters that can be obtained in these various ways are in very close agreement with those obtained from the extraction of the interactions of $\widehat{H}_{\text{multispin}}$ [74]. In conclusion, if one wants to estimate local anisotropy parameters or local magnetic axes, these approaches can be safely considered. It is also worth mentioning here that in the general case, the $S/S + 1$ spin mixings do not arise solely from $\bar{d}_{ab}$: the

mismatch between the local magnetic axes of the local $\overline{\overline{D}}_a$ and $\overline{\overline{D}}_b$ tensors also affects the effective couplings related to these mixings. This can be easily shown by considering $D_a = D_b \neq 0$, $E_a = E_b = 0$, and an angle 2α between coplanar local $Z^m_{a,b}$ axes. If one builds the corresponding model interaction matrix within the uncoupled basis and transforms it to the coupled one, the $S/S + 1$ spin-mixing terms are found proportional to $D_a \sin 2\alpha$ (the details of the derivation are not given here for a sake of simplicity). It is thus clear that these terms vanish for $\alpha = 0$, as it is the case for centrosymmetric complexes. Therefore, the local anisotropy tensors can be affected in the general case by the $S/S + 2$ and to $S/S + 1$ spin mixings. Furthermore, note that one should also never neglect these terms within the weak-exchange limit, contrary to what was done, for instance, to interpret the low-energy spectrum of a cobalt(II)-cobalt(II) complex [64]. From the perspective of modeling, it is not clear yet whether the model Hamiltonian used for the $S_a = S_b = 1$ case is directly applicable to any other configuration. Actually, higher-rank tensors could be necessary to reproduce all ZFS features for higher local spins [54], such as, for Instance, a sixth-rank tensor in the $S_a = S_b = 3/2$ case.

Magneto-Structural Correlations

Magneto-structural correlations are particularly useful for chemists as they give clues to tune the properties of a system and pave the way for the rational design of new magnetic systems with predetermined properties. In the field of ZFS, they can be established from correlations of molecular geometry features with the experimental values of the parameters, as was done by Titiš and Boča in nickel(II) and cobalt(II) mononuclear complexes [84,85]. Here, we will only discuss magneto-structural correlations deduced from the analysis of ab initio results or from combined ab initio/crystal-field studies.

Magneto-Structural Correlations Based on Ab Initio Calculations

Due to the lack of intuition on the role of distortions on the $\bar{d}_{ab}$ pseudo-vector components, Maurice et al. studied the effects of two angular distortions on the DM vector components on model $[Cu_2O(H_2O)_6]^{2+}$ complexes (see Fig. 6) [48]. This study was based on CASSCF(2/2) + c-SOCI calculations and made use of the effective Hamiltonian theory, as mentioned in section "Multispin Hamiltonians".

The norm of the $\bar{d}_{ab}$ pseudo-vector as a function of the ϑ_1 and ϑ_2 deformation angles is represented in Fig. 7.

In this study, the only point for which the $\bar{d}_{ab}$ pseudo-vector is null by symmetry corresponds to the $\vartheta_1 = \vartheta_2 = 0$ case, for which the system possesses a symmetry center. Another interesting point refers to $\vartheta_1 = \pi/2$ and $\vartheta_2 = 0$. In this case, there is no atomic orbital contribution from the copper(II) centers, and as a consequence, the

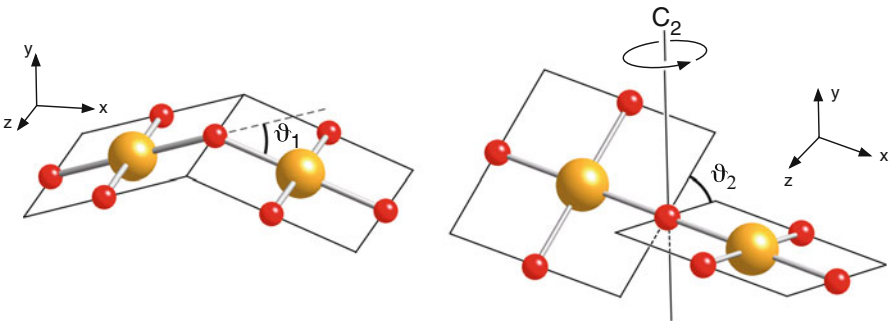


Fig. 6 Ball-and-stick representation of model $[Cu_2O(H_2O)_6]^{2+}$ complexes and the deformation angles that were applied to them [48]. All the hydrogen atoms are omitted for clarity

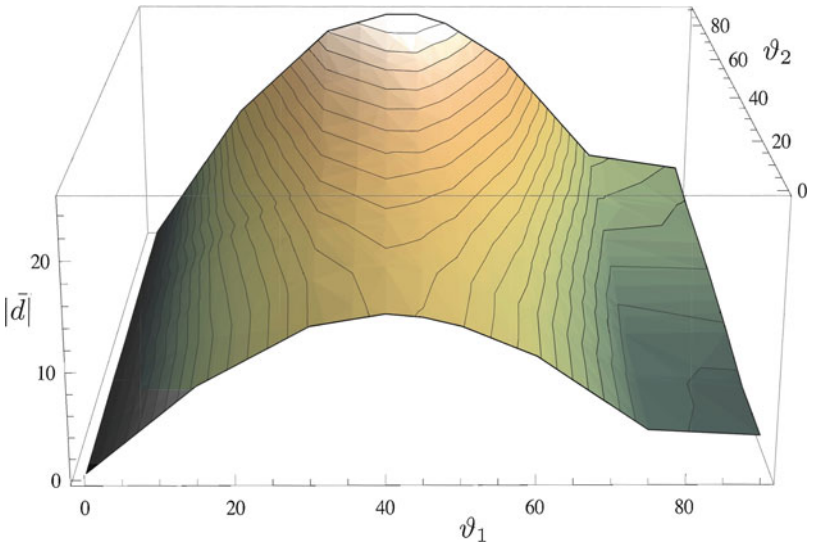


Fig. 7 Norm of the $\bar{d}_{ab}$ pseudo-vector as a function of the ϑ_1 and ϑ_2 deformation angles (see Fig. 6 for the definition of these angles) [48]

only important contribution to this norm comes from the bridging oxygen center. As the $\bar{d}_{ab}$ pseudo-vector is far from being null, in this case ($|\bar{d}_{ab}| = 3.6\,\mathrm{cm}^{-1}$ [48]) the (essentially) closed-shell bridging oxygen contributes to the norm of the $\bar{d}_{ab}$ pseudo-vector, as highlighted by Moskvin in 2007 [59]. Other studies concerning single-ion anisotropies exist in the literature, among which one may quote the extensive one of Gomez-Coca et al. [28], but these will not be discussed here. Instead, we will explore the cases of mononuclear complexes for which ab initio calculations are used to extract crystal-field parameters.

Magneto-Structural Correlations Based on Crystal-Field Models and Ab Initio Calculations

Joint ab initio and crystal-field studies are important to validate the equations derived from the crystal-field theory and to interpret the outcomes of experiments in a simple way. It is commonly practiced on mononuclear complexes, although equations can also be derived in binuclear complexes, as was done, for instance, in the case of copper acetate monohydrate [50]. In this section, we shall illustrate the procedure for the d^8 and d^4 configurations, using ab initio calculations on nearly octahedral $[Ni(NCH)_6]^{2+}$ and $[Mn(NCH)_6]^{3+}$ model complexes. In both cases, we define the axial deformation parameter as:

$$\tau_{ax} = \frac{2d\,(TM, N_Z)}{d\,(TM, N_X) + d\,(TM, N_Y)}$$

and the rhombic deformation as:

$$\tau_{rh} = \frac{d\,(TM, N_Y)}{d\,(TM, N_X)}$$

The mean d(TM,N) distance is 2.054 Å and 2.061 Å for TM = Ni [49] and TM = Mn [45], respectively, while all d(TM,C) parameters are fixed to 1.155 Å and all the d(C,H) ones to 1.083 Å. Minimal CASSCF calculations have been carried out with the five d-orbitals and 8 (Ni) or 5 (Mn) electrons within the active space. Note that in the formulae that are presented here, monoelectronic ζ SOC constants are considered. These constants are always positive and can be converted into polyelectronic λ ones by using the following relation:

$$\lambda = \pm\frac{\zeta}{2S}$$

where S is the total spin of the ground SOF free-ion multiplet.

It is easy to show, as done in the textbook of Abragam and Bleaney [1], that for axially distorted systems:

$$D = -\frac{\zeta^2}{\Delta_1} + \frac{\zeta^2}{\Delta_2}$$

where Δ_1 is the $^3B_{1g} \rightarrow {}^3B_{2g}$ excitation energy and Δ_2 corresponds to the $^3B_{1g} \rightarrow {}^3E_g$ excitation energy. The derivation of this equation is based on a model space containing the spin components of the ground SOF $^3B_{1g}$ state, while the external space consists of the spin components of the lowest two excited SOF states, namely, $^3B_{2g}$ and ^3E_g, which are essentially singly excited states with respect to $^3B_{1g}$. To check the correlation between the ab initio results and the outcomes of the

Table 5 Ab initio Δ_1, Δ_2, and D values and values of D derived from the crystal-field expression (D_{CF}) obtained with $\zeta = 648\,cm^{-1}$ (all values are in cm^{-1}) [49]

τ_{ax}	Δ_1	Δ_2	D	D_{CF}
0.957	8382.4	9692.5	−5.519	−6.855
0.971	8669.7	9559.4	−3.568	−4.564
0.985	8964.0	9416.5	−1.736	−2.279
1.000	9266.9	9266.9	0.000	0.000
1.015	9576.9	9110.8	1.659	2.271
1.029	9895.8	8951.8	3.259	4.530
1.044	10,224.0	8791.1	4.814	6.778

crystal-field model, we substitute the ab initio Δ_1 and Δ_2 values in the expression for D and take the SOC constant of the free Ni^{2+} ion, 648 cm^{-1} (see Table 5). As can be seen, a good correlation appears between the ab initio and the crystal-field D values (D_{CF}): the trend line that passes through the origin, as forced by symmetry, leads to $D_{CF} = 1.32\,D$ with $R^2 = 0.9969$. This shows that accounting for covalency effects by applying a reduction factor of 0.87 brings the D_{CF} values in perfect agreement with the ab initio ones along the whole curve. Therefore, the crystal-field formula presented above is fully supported by c-SOCI calculations that consider the spin components of four SOF triplet states. Note that considering more SOF states in the first step of the calculation does not significantly improve the computed values, meaning that this formula explains most of the ZFS in axially distorted six-coordinated nickel(II) complexes.

In a similar way, one can introduce a rhombic distortion, i.e., an in-plane radial distortion. The crystal-field derivation leads to:

$$D = -\frac{\zeta^2}{\Delta_1} + \frac{\zeta^2}{2\Delta_2} + \frac{\zeta^2}{2\Delta_3}$$

and:

$$E = -\frac{\zeta^2}{2\Delta_2} + \frac{\zeta^2}{2\Delta_3}$$

where Δ_1 correlates with the $^3B_{1g} \rightarrow {}^3B_{2g}$ energy difference in the D_{4h} symmetry point group and where Δ_2 and Δ_3 both correlate with the energy of the $^3B_{1g} \rightarrow {}^3E_g$ excitation in the same symmetry point group [49]. Similarly, it can be shown that these formulae are supported by ab initio calculations [49], corroborating that the outcomes of model complex studies which aim at establishing magneto-structural correlation can be safely explained by crystal-field models.

Before presenting the example of nearly octahedral manganese(III) complexes, it is worth mentioning that such derivations are only valid close to ideal geometries of high symmetry, since it is assumed that the ground and excited SOF wave functions are either (i) not much affected by the distortion or (ii) affected in a way that can be easily modeled. In general, it is always advisable to perform

ab initio calculations and to analyze the nature of the SOF wave functions of interest. For instance, large angular distortions usually result in a mixing of various configurations in such a complex way that pen-and-paper analytical derivations become cumbersome. Moreover, even if analytical derivations can be performed, if the resulting formulae are too complicated, they become pointless in practice for establishing or understanding magneto-structural correlations. In such cases, it is preferable to directly establish the correlations by means of ab initio calculations, as presented in section "Magneto-Structural Correlations Based on Ab Initio Calculations".

Another interesting point concerns the role of the second coordination sphere. Although this effect is traditionally neglected in crystal-field models, it was shown by means of ab initio calculations that the second coordination sphere can play a crucial role on the single-ion anisotropy in some particular cases [11, 51].

Nearly octahedral manganese(III) complexes, corresponding to the d^4 configuration, are particularly interesting as they present nonintuitive ZFSs [45]. Although formulae which rationalize the ZFS of such systems are presented in the book of Abragam and Bleaney [1], this case is often misinterpreted. Ab initio calculations showed that the external space cannot be restricted to quintet-spin SOF state components. Indeed, three triplet-spin roots must also be included in the derivation to obtain accurate crystal-field formulae for this configuration [45]. The use of these newly derived expressions leads to the energies of the ten spin components of the SOF ^{5}E state reported in Table 6. Note that the same wave functions as those obtained by Abragam and Bleaney with five quintet roots were obtained [1]. The trend line $E_{CF} = 1.034\, E_{\text{ab initio}}$ has an R^2 value of 0.9992, meaning that (i) the crystal-field formulae presented in Table 6 are valid and (ii) a reduced effective SOC constant of 346 cm^{-1} has to be used in the crystal-field model to effectively account for covalency.

Similarly to the nickel(II) case, one can derive formulae for the D and E parameters of Mn(III) complexes which belong to the D_{4h} and D_{2h} symmetry point groups. In this case, analytical formulae can also be written down for the parameters appearing in the Stevens fourth-rank operators. However, as shown in [45], one can

Table 6 Analytical crystal-field expressions of the energies, ab initio, and derived crystal-field energies (in cm^{-1}) [45]. For the computation of the crystal-field energies, the free-ion SOC of 352 cm^{-1} and ab initio SOF excitation energies of $\Delta Q = 13{,}993\,\text{cm}^{-1}$ and $\Delta T = 11{,}005\,\text{cm}^{-1}$ have been used

Multiplicity	$E_{\text{analytical}}$	$E_{\text{ab initio}}$	E_{CF}
Singlet	$2\frac{4\Delta Q\zeta^2+3\Delta T\zeta^2}{8\Delta Q\Delta T}$	16.813	17.900
Triplet	$\frac{4\Delta Q\zeta^2+3\Delta T\zeta^2}{8\Delta Q\Delta T}$	8.399	8.950
Doublet	0	0.000	0.000
Triplet	$-\frac{4\Delta Q\zeta^2+3\Delta T\zeta^2}{8\Delta Q\Delta T}$	−8.890	−8.950
Singlet	$-2\frac{4\Delta Q\zeta^2+3\Delta T\zeta^2}{8\Delta Q\Delta T}$	−17.775	−17.900

consider various approximations, i.e., neglecting (i) the Stevens fourth-rank terms (large distortions) and (ii) the degeneracy lift of the excited SOF multiplets of the octahedral situation ($^3T_{1g}$ and $^5T_{2g}$). For the former approximation, one should note that in the case of small distortions, the Stevens fourth-rank terms can be important as they are closely related to the near-degeneracy of the states that originate from the ^5E_g of the octahedron [45]. In the D_{2h} symmetry point group, two configurations mix to form the two lowest SOF states [1]. It is therefore necessary to introduce a mixing parameter δ to express these states as:

$$|\phi_1\rangle = \cos\delta|Q_1\rangle + \sin\delta|Q_2\rangle$$

and:

$$|\phi_2\rangle = -\sin\delta|Q_1\rangle + \cos\delta|Q_2\rangle$$

where Q_1 and Q_2 are the two coupled configurations. This leads, after some pen-and-paper work, to [45]:

$$D = \zeta^2 \cos 2\delta \left[\frac{3}{16\Delta Q} + \frac{1}{4\Delta T}\right]$$

and:

$$E = \zeta^2 |\sin 2\delta| \left[\frac{3}{16\Delta Q} + \frac{1}{4\Delta T}\right]$$

where E is defined positive by convention. The careful reader will notice that, as in the octahedral situation, the contribution of the triplet roots which were added in this derivation is proportional to the effect of the quintet roots that correlate with $^5T_{2g}$ in the octahedron. Finally, we end up with the same formula as the one reported in the book of Abragam and Bleaney [1]:

$$\frac{E}{|D|} = \frac{\sqrt{3}}{3}|\tan 2\delta|$$

To conclude, one should mention that the anisotropy parameters are not enlarged by distorting the first coordination sphere in this configuration, as illustrated by Fig. 8 in the case of axial distortions. The ZFS of complexes belonging to this configuration is nonintuitive, and the combined ab initio and crystal-field model study has proved to be enlightening for experimental applications. Indeed, in this configuration, it is pointless to synthesize complexes with large distortions to enlarge the ZFS parameters, which is notably consistent with the empirical fact that d^4 complexes typically have axial ZFS parameter values ranging between -5 and $5\,cm^{-1}$ [10].

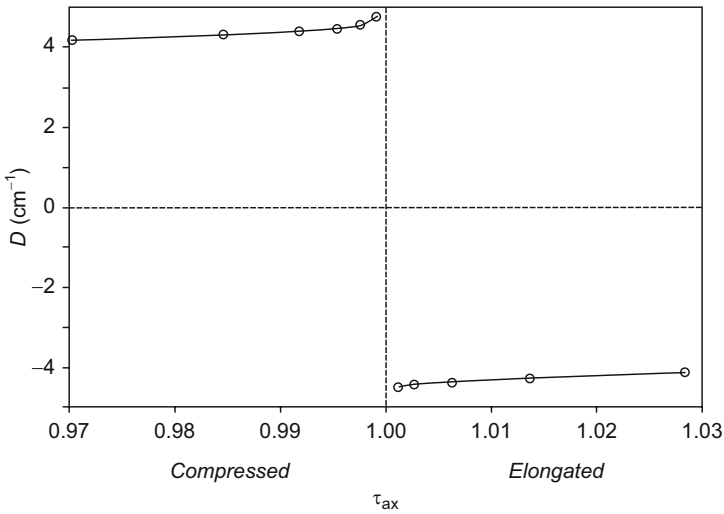


Fig. 8 Ab initio D parameter as a function of the axial distortion in model D_{4h} manganese(III) complexes [45]

Conclusion

In this chapter, we have shown that phenomenological Hamiltonians can be justified or even improved using the effective Hamiltonian theory; this creates a bridge between (supposedly) accurate ab initio calculations and intuitive models. We have also exemplified how the crystal-field theory can be used to rationalize the nature and magnitude of ZFS. These tools allow us to take another step in the direction of the control of magnetic properties as they provide concrete understanding of how to increase the magnetic anisotropy. Magneto-structural correlations have also been established, which may help to design molecules with desired properties. To maximize single-ion anisotropies, researchers followed strategies such as exploring exotic coordination spheres (for instance, pentacoordinate or heptacoordinate complexes) and even low-coordination spheres. Several efforts have been devoted to binuclear and polynuclear systems, more often concerning single-ion anisotropies but also concerning giant-spin and multispin models. Consequently, substantial progress has been made in the understanding of the magnitude and nature of the magnetic anisotropy in TM complexes with a wide range of ZFS parameter values over the last two decades.

Nevertheless, many aspects deserve further studies. For instance, the treatment of large systems is still problematic as (i) it is not clear which method can be used to obtain an optimal balance between accuracy and efficiency and (ii) current models to describe polynuclear complexes may not be complete. Indeed, even in the case of mononuclear complexes, some cases are typically pathological, e.g., when heavy atom ligands are involved. In this situation, it is not clear yet if sum-

over-states or c-SOCI approaches can be safely applied to the computation of the ZFS due to the truncation and state-averaging errors [49]. One may thus prefer to introduce the spin-dependent effects a priori, as done within 2*c* frameworks. Also, some experimental data may have been incorrectly interpreted due to the use of inadequate models to fit the experimental outcomes of various techniques (magnetic susceptibility, magnetization, EPR, etc.). Therefore, there is clearly a need for more extensive studies and developments in the field of molecular magnetism. We hope that this chapter will motivate future work of this kind.

References

1. Abragam A, Bleaney B (1986) Electron paramagnetic resonance of transition ions. Dover Publications, New York
2. Abragam A, Pryce MHL (1951) The theory of the nuclear hyperfine structure of paramagnetic resonance spectra in the copper Tutton salts. Proc R Soc Lond Ser A 206:164–172
3. Altshuler SA, Kozyrev BM (1974) Electron paramagnetic resonance in compounds of transition elements. Wiley, New York
4. Andersson K, Malmqvist P-Å, Roos BO (1992) Second-order perturbation theory with a complete active space self-consistent field reference function. J Chem Phys 96:1218–1226
5. Angeli C, Cimiraglia R, Evangelisti S, Leininger T, Malrieu J-P (2001) Introduction of n-electron valence states for multireference perturbation theory. J Chem Phys 114:10252–10264
6. Aquino F, Rodriguez JH (2005) First-principle computation of zero-field splittings: application to a high valent Fe(IV)-oxo model of nonheme iron proteins. J Chem Phys 123:204902
7. Atanasov M, Zadrozny JM, Long JR, Neese F (2013) A theoretical analysis of chemical bonding, vibronic coupling, and magnetic anisotropy in linear iron(II) complexes with single-molecule magnet behavior. Chem Sci 4:139–156
8. Bloch C (1958) Sur la théorie des perturbations des états liés. Nucl Phys 6:329–347
9. Boča R (1999) Theoretical foundations of molecular magnetism. Elsevier, Amsterdam
10. Boča R (2004) Zero-field splitting in metal complexes. Coord Chem Rev 248:757–815
11. Bogdanov NA, Maurice R, Rousochatsakis I, van ben Brink J, Hozoi L (2013) Magnetic state of pyrochlore $Cd_2Os_2O_7$ emerging from strong competition of ligand distortions and longer-range crystalline anisotropy. Phys Rev Lett 110:127206
12. Buckingham AD, Pyykko P, Robert JB, Wiesenfeld L (1982) Symmetry rules for the indirect nuclear spin-spin coupling tensor revisited. Mol Phys 46:177–182
13. Caneschi A, Gatteschi D, Sessoli R, Barra AL, Brunel LC, Guillot M (1991) Alternating current susceptibility, high field magnetization, and millimeter band EPR evidence for a ground S = 10 state in $[Mn_{12}O_{12}(CH_3COO)_{16}(H_2O)_4] \cdot 2CH_3COOH \cdot 4H_2O$. J Am Chem Soc 113:5873–5874
14. Chang C, Pélissier M, Durand P (1986) Regular two-component Pauli-like effective Hamiltonians in Dirac theory. Phys Scripta 34:394–404
15. Chibotaru LF, Ungur L (2012) Ab initio calculation of anisotropic magnetic properties of complexes. I. Unique definition of pseudospin Hamiltonians and their derivation. J Chem Phys 137:064112
16. Costes J-P, Maurice R, Vendier L (2012) Pentacoordinate Ni^{II} complexes: preparation, magnetic measurements, and ab initio calculations of the magnetic anisotropy terms. Chem Eur J 18:4031–4040
17. de Graaf C, Broer R (2015) Magnetic interactions in molecules and solids. Springer. ISBN:978-3-319-22950-8

18. des Cloizeaux J (1960) Extension d'une formule de Lagrange à des problèmes de valeurs propres. Nucl Phys 20:321–346
19. Dirac PAM (1929) Quantum mechanics of many-electron systems. Proc R Soc Lond Ser A 123:714–733
20. Douglas N, Kroll NM (1976) Quantum electrodynamical corrections to the fine structure of Helium. Ann Phys 82:89–155
21. Duboc C, Ganyushin D, Sivalingam K, Collomb M-N, Neese F (2010) Systematic theoretical study of the zero-field splitting in coordination complexes of Mn(III). Density functional theory versus multireference wave function approaches. J Phys Chem A 114:10750–10758
22. Dzyaloshinskii I (1958) A thermodynamic theory of weak ferromagnetism of antiferromagnetics. J Phys Chem Solids 4:241–255
23. Ederer C, Spaldin NA (2004) A new route to magnetic ferroelectrics. Nat Mater 3:849–851
24. Finley J, Malmqvist P-Å, Roos BO, Serrano-Andrés L (1998) The multi-state CASPT2 method. Chem Phys Lett 288:299–306
25. Ganyushin D, Neese F (2006) First-principles calculations of zero-field splitting parameters. J Chem Phys 125:024103
26. Gatteschi D, Sessoli R (2003) Quantum tunneling of magnetization and related phenomena in molecular materials. Angew Chem Int Ed 42:268–297
27. Ginsberg AP, Brookes RW, Martin RL, Sherwood RC (1972) Magnetic exchange in transition metal complexes. IX. Dimeric nickel(II)- ethylenediamine complexes. Inorg Chem 11:2884–2889
28. Gomes-Coca S, Cremades E, Aliaga-Alcalde N, Ruiz E (2013) Mononuclear single-molecule magnets: tailoring the magnetic anisotropy of first-row transition-metal complexes. J Am Chem Soc 135:7010–7018
29. Heisenberg W (1928) Zur Theorie des Ferromagnetismus. Z Phys 49:619–636
30. Herchel R, Boča R, Krzystek J, Ozarowski A, Durán M, van Slageren J (2007) Definitive determination of zero-field splitting and exchange interactions in a Ni(II) dimer: investigation of $[Ni_2(en)_4Cl_2]Cl_2$ using magnetization and tunable-frequency high-field electron paramagnetic resonance. J Am Chem Soc 129:10306–10307
31. Hess BA (1986) Relativistic electronic structure calculations employing a two-component no-pair formalism with external-field projection operators. Phys Rev A 33:3742–3748
32. Hess BA, Marian CM, Wahlgren U, Gropen O (1996) A mean-field spin-orbit method applicable to correlated wavefunctions. Chem Phys Lett 251:365–371
33. Ibério de Moreira PR, Illas F (2006) A unified view of the theoretical description of magnetic coupling in molecular chemistry and solid state physics. Phys Chem Chem Phys 8:1645–1659
34. Ibério de Moreira PR, Illas F, Calzado CJ, Sanz JF, Malrieu J-P, Ben Amor N, Maynau D (1999) The local character of magnetic coupling in ionic solids. Phys Rev B 59:6593–6596
35. Iliaš M, Saue T (2007) An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. J Chem Phys 126:064102
36. Jansen G, Hess BA (1989) Revision of the Douglas-Kroll transformation. Phys Rev A 39:6016–6017
37. Joung KO, O'Connor CJ, Sinn E, Carlin RL (1979) Structural and magnetic properties of dimeric $[Ni_2(en)_4Cl_2]Cl_2$. Inorg Chem 18:804–808
38. Journaux Y, Kahn O, Chevalier B, Etrouneau J, Claude R, Dworkin A (1978) Evidence for a low temperature phase transition in di-μ-chloro-tetrakis (ethylene diamine) dinickel(II) chloride. Chem Phys Lett 55:140–143
39. Kahn O (1993) Molecular magnetism. VCH Publishers, New York
40. Krzystek J, Zvyagin SA, Ozarowski A, Fiedler AT, Brunold TC, Telser J (2004) Definitive spectroscopic determination of zero-field splitting in high-spin cobalt(II). J Am Chem Soc 126:2148–2155
41. Kubica A, Kowalewski J, Kruk D, Odelius M (2013) Zero-field splitting in nickel(II) complexes: a comparison of DFT and multiconfigurational wavefunction calculations. J Chem Phys 138:064304

42. Llusar R, Casarrubios M, Barandiarán Z, Seijo L (1996) Ab initio model potential calculations on the electronic spectrum of Ni^{2+} doped MgO including correlation, spin-orbit and embedding effects. J Chem Phys 105:5321–5330
43. Malrieu J-P, Caballol R, Calzado CJ, de Graaf C, Guihéry N (2014) Magnetic interactions in molecules and highly correlated materials: Physical content, analytical derivation and rigorous extraction of magnetic Hamiltonians. Chem Rev 114:429–492
44. Maurice R, Bastardis R, de Graaf C, Suaud N, Mallah T, Guihéry N (2009) Universal theoretical approach to extract anisotropic spin Hamiltonians. J Chem Theory Comput 5:2977–2989
45. Maurice R, de Graaf C, Guihéry N (2010) Magnetostructural relations from a combined ab initio and ligand field analysis for the nonintuitive zero-field splitting in Mn(III) complexes. J Chem Phys 133:084307
46. Maurice R, Guihéry N, Bastardis R, de Graaf C (2010) Rigorous extraction of the anisotropic multispin Hamiltonian in bimetallic complexes from the exact electronic Hamiltonian. J Chem Theory Comput 6:55–65
47. Maurice R, de Graaf C, Guihéry N (2010) Magnetic anisotropy in binuclear complexes in the weak- exchange limit: From the multispin to the giant-spin Hamiltonian. Phys Rev B 81:214427
48. Maurice R, Pradipto AM, Guihéry N, Broer R, de Graaf C (2010) Antisymmetric magnetic interactions in oxo-bridged copper(II) bimetallic systems. J Chem Theory Comput 6:3092–3101
49. Maurice R (2011) Zero-field anisotropic spin Hamiltonians in first-row transition metal complexes: theory, models and applications. PhD thesis, Université de Toulouse. Available online at http://thesesups.ups-tlse.fr/1430
50. Maurice R, Sivalingam K, Ganyushin D, Guihéry N, de Graaf C, Neese F (2011) Theoretical determination of the zero-field splitting in copper acetate monohydrate. Inorg Chem 50:6229–6236
51. Maurice R, Vendier L, Costes J-P (2011) Magnetic anisotropy in Ni^{II}–Y^{III} binuclear complexes: On the importance of both the first coordination sphere of the Ni^{II} Ion and the Y^{III} ion belonging to the second coordination sphere. Inorg Chem 50:11075–11081
52. Maurice R, Pradipto A-M, de Graaf C, Broer R (2012) Magnetic interactions in $LiCu_2O_2$: Single-chain versus double-chain models. Phys Rev B 86:024411
53. Maurice R, Verma P, Zadrozny JM, Luo S, Borycz J, Long JR, Truhlar DG, Gagliardi L (2013) Single-ion magnetic anisotropy and isotropic magnetic couplings in the metal– organic framework Fe_2(dobdc). Inorg Chem 52:9379–9389
54. Maurice R, de Graaf C, Guihéry N (2013) Theoretical determination of spin Hamiltonians with isotropic and anisotropic magnetic interactions in transition metal and lanthanide complexes. Phys Chem Chem Phys 15:18784–18804
55. Mceeny R, Mizuno Y (1961) The density matrix in many-electron quantum mechanics. II. Separation of space and spin variables; Spin coupling problems. Proc R Soc Lond Ser A 259:55–577
56. McWeeny R (1965) On the origin of spin-Hamiltonian parameters. J Chem Phys 42:1717–1725
57. Miralles J, Castell O, Caballol R, Malrieu J-P (1993) Specific CI calculation of energy differences: transition energies and bond energies. Chem Phys 172:33–43
58. Moriya T (1960) Anisotropic superexchange interaction and weak ferromagnetism. Phys Rev 120:91–98
59. Moskvin AS (2007) Dzyaloshinsky-Moriya antisymmetric exchange coupling in cuprates: oxygen effects. J Exp Theor Phys 104:913–927
60. Neese F (2005) Efficient and accurate approximations to the molecular spin-orbit coupling operator and their use in molecular g-tensor calculations. J Chem Phys 122:034107
61. Neese F (2006) Importance of direct spin-spin coupling and spin-flip excitations for the zero-field splittings of transition metal complexes: a case study. J Am Chem Soc 128:10213–10222
62. Neese F (2007) Calculation of the zero-field splitting tensor on the basis of hybrid density functional and Hartree-Fock theory. J Chem Phys 127:164112

63. Neese F, Solomon EI (1998) Calculation of zero-field splittings, g-values and the relativistic nephelauxetic effect in transition metal complexes. Application to High-spin ferric complexes. Inorg Chem 37:6568–6582
64. Ostrovsky SM, Werner R, Brown DA, Hasse W (2002) Magnetic properties of dinuclear cobalt complexes. Chem Phys Lett 353:290–294
65. Ozarowski A (2008) The zero-field splitting parameter D in binuclear copper(II) carboxylates is negative. Inorg Chem 47:9760–9762
66. Pederson MR, Khanna SN (1999) Magnetic anisotropy barrier for spin tunneling in $Mn_{12}O_{12}$ molecules. Phys Rev B 60:9566–9572
67. Pradipto A-M, Maurice R, Guihéry N, de Graaf C, Broer R (2012) First-principles study of magnetic interactions in cupric oxide. Phys Rev B 85:014409
68. Rebilly J-N, Charron G, Rivière R, Guillot E, Barra A-L, Serrano MD, van Slageren J, Mallah T (2008) Large magnetic anisotropy in pentacoordinate Ni^{II} complexes. Chem Eur J 14:1169–1177
69. Retegan M, Cox N, Pantazis DA, Neese F (2014) A first-principles approach to the calculation of the on-site zero-field splitting in polynuclear transition metal complexes. Inorg Chem 53:11785–11793
70. Rogez G, Rebilly J-N, Barra A-L, Soraca L, Blondin G, Kirchner N, Duran M, van Slageren J, Parsons S, Ricard L, Marvilliers A, Mallah T (2005) Very large Ising-type magnetic anisotropy in a mononuclear Ni^{II} complex. Angew Chem Int Ed 44:1876–1879
71. Roos BO (2005) In: Multiconfigurational quantum chemistry. Dykstra CE, Frenking G, Kim KS, Scuseria GE (eds) Theory and applications of computational chemistry: the first forty years. Elsevier, Amsterdam, chapter 25, pp 725–764
72. Roos BO, Malmqvist P-Å (2004) Relativistic quantum chemistry: the multiconfigurational approach. Phys Chem Chem Phys 6:2919–2927
73. Roos BO, Taylor PR, Siegbahn PEM (1980) A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. Chem Phys 48:157–173
74. Ruamps R, Maurice R, de Graaf C, Guihéry N (2014) Interplay between local anisotropies in binuclear complexes. Inorg Chem 53:4508–4516
75. Ruamps R, Batchelor LJ, Maurice R, Gogoi N, Jiménez-Lozano P, Guihéry N, de Graaf C, Barra A-L, Sutter J-P, Mallah T (2013) Origin of the magnetic anisotropy in heptacoordinate Ni^{II} and Co^{II} complexes. Chem Eur J 19:950–956
76. Ruamps R, Maurice R, Batchelor LJ, Boggio-Pasqua M, Guillot R, Barra A-L, Liu J, Bendeif E-E, Pillet S, Hill S, Mallah T, Guihéry N (2013) Giant Ising-type magnetic anisotropy in trigonal bipyramidal Ni(II) complexes: Experiment and theory. J Am Chem Soc 135:3017–3026
77. Rudowicz C, Chung CY (2004) The generalization of the extended Stevens operators to higher ranks and spins, and a systematic review of the tables of the tensor operators and their matrix elements. J Phys Condens Matter 16:5825–5847
78. Schmitt S, Jost P, van Wüllen C (2011) Zero-field splittings from density functional calculations: analysis and improvement of known methods. J Chem Phys 134:194113
79. Spiegelmann F, Malrieu J-P (1984) The use of effective Hamiltonians for the treatment of avoided crossings. I. Adiabatic potential curves. J Phys B Atom Mol Phys 17:1235–1257
80. Spiegelmann F, Malrieu J-P (1984) The use of effective Hamiltonians for the treatment of avoided crossings. II. Nearly diabatic potential curves. J Phys B Atom Mol Phys 17:1259–1279
81. Stevens KWH (1952) Matrix elements and operator equivalents connected with the magnetic properties of rare earth ions. Proc R Soc Lond Ser A 65:209–215
82. Teichteil C, Pélissier M, Spiegelman F (1983) Ab initio molecular calculations including spin-orbit coupling. I. Method and atomic tests. Chem Phys 81:273–282
83. Telser J (2006) A perspective on applications of ligand-field analysis: inspiration from electron paramagnetic resonance spectroscopy of coordination complexes of transition metal ions. J Braz Chem Soc 17:1501–1515

84. Titiš J, Boča R (2010) Magnetostructural D correlation in nickel(II) complexes: reinvestigation of the zero-field splitting. Inorg Chem 49:3971–3973
85. Titiš J, Boča R (2011) Magnetostructural D correlations in hexacoordinated cobalt(II) complexes. Inorg Chem 50:11838–11845
86. van Lenthe E, Baerends EJ, Snijders JG (1994) Relativistic total energy using regular approximations. J Chem Phys 101:9783–9792
87. van Lenthe E, Snijders JG, Baerends EJ (1996) The zero-order regular approximation for relativistic effects: the effect of spin-orbit coupling in closed shell molecules. J Chem Phys 105:6505–6516
88. Van Vleck JH (1945) A survey of the theory of ferromagnetism. Rev Mod Phys 17:27–47
89. Zadrozny JM, Atanasov M, Bryan AM, Lin C-Y, Rekken BK, Power PP, Neese F, Long JR (2013) Slow magnetization dynamics in a series of two-coordinate iron(II) complexes. Chem Sci 4:125–138
90. Zein S, Neese F (2008) Ab initio and coupled-perturbed density functional theory estimation of zero-field splittings in Mn^{II} transition metal complexes. J Phys Chem A 112:7976–7983

Relativistic Equation-of-Motion Coupled-Cluster Theory (EOM-CC)

25

Fan Wang

Contents

Introduction 798
Relativistic EOM-CC Theory 801
Summary 819
References 821

Abstract

Equation-of-motion coupled-cluster (EOM-CC) theory can be employed to calculate excitation energies (EE), ionization potentials(IP), as well as electron affinities (EA). The EOM-CC approach at the CC singles and doubles level (CCSD) is able to provide EEs, IPs, and EAs with an error of about 0.1–0.3 eV for single-excitation states or Koopmans states from a reference with a dominant single-reference character. Scalar-relativistic effects can be incorporated straightforwardly in EOM-CC calculations when untransformed two-electron interactions are adopted. On the other hand, time-reversal symmetry and spatial symmetry of double point groups need to be exploited to achieve an efficient implementation when spin-orbit coupling (SOC) is present. Furthermore, including SOC in post-self-consistent field (SCF) treatment could result in a further reduction in computational effort particularly for molecules with low symmetry. Due to effective treatment of orbital relaxation effects by single excitations in the cluster operator, this approach can afford accurate description on SOC effects. It is nontrivial to impose time-reversal symmetry for open-shell reference and broken time-reversal symmetry could result in spurious-level splitting. Open-shell system with one-unpaired electron can be calculated based on EOM-CC

F. Wang (✉)
Institute of Atomic and Molecular Physics, Sichuan University, Chengdu, China
e-mail: wangf44@gmail.com

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_33

for IPs or EAs from a closed-shell reference. In addition, Kramer's degeneracy has to be taken into consideration when calculating properties of systems with an odd number of electrons using EOM-CC for IPs or EAs from a closed-shell reference. EEs, IPs, and EAs for systems containing heavy elements can be obtained reliably based on EOM-CCSD approaches, and SOC splitting is calculated with reasonable accuracy even for double-excitation states.

Keywords

Coupled-cluster theory • Equation-of-motion coupled-cluster theory • Spin-orbit coupling • Excitation energy • Ionization potential • Electron affinity

Introduction

Coupled-cluster theory [1] is one of the most important methods in quantum chemistry to calculate dynamical correlation. Energies and properties of ground states or the lowest state of a certain symmetry with mainly single-reference character can be estimated rather accurately with CC approaches. In addition, total energies provided by CC approaches are size extensive and converge fast with respect to the highest excitation level employed in the cluster operator. The CC approach with the singles and doubles (CCSD) approximation augmented by a perturbative treatment of triple excitations (CCSD(T)) [2] is currently referred to as the "gold standard" of quantum chemistry.

Besides ground states, CC theory can also be used to calculate excited-state energies based on the equation-of-motion CC approach (EOM-CC) [1, 3, 4]. Excitation energies (EE) are determined from eigenvalues of a similarity-transformed Hamiltonian within EOM-CC. Besides excitation energies, ionization energies (IP), electron attachment (EA), double ionization potentials (DIP), as well as double electron attachment (DEA) can also be calculated from eigenvalues of the same similarity-transformed Hamiltonian in various spaces. Recently EOM-CC has even been extended to calculate triple electron-attached states [5]. It is worth noting that IPs and EAs can also be obtained from EOM-EE-CC calculations at a larger computational cost by including a continuum orbital in the basis set and calculating transition energies from or to this continuum orbital [6]. Closely related CC approaches for excitation energies are the linear response CC (LR-CC) approach [4, 7], the Fock-space CC (FSCC) approach [8], the similarity-transformed equation-of-motion CC method (STEOM-CC) [9], and the symmetry-adapted-cluster configuration-interaction (SAC-CI) method [10]. EEs from EOM-CC and LR-CC are the same and these two approaches differ in transition properties. Transition properties based on LR-CC are size intensive, while those with EOM-CC are not. However, it is more involved to calculate transition properties with LR-CC than that with EOM-CC, and recent investigations show that transition dipole moments with these two approaches agree rather well with each other at the CCSD level, while difference will be somewhat larger in lower-level calculations [11].

EEs, IPs, EAs, etc. can also be obtained with FSCC in different sectors of Fock space. In fact, IPs and EAs with EOM-CC are equivalent to those obtained with FSCC, while EEs, DIPs, and DEAs with FSCC differ from those based on EOM-CC. Computational effort for EEs, DIPs, and DEAs with FSCC is smaller than that with EOM-CC, but FSCC results depend on the chosen active space. On the other hand, the STEOM-CC approach closely resembles the intermediate Hamiltonian FSCC approach [12], which is proposed to avoid intruder-state problem in FSCC.

One of the most popular EOM-CC approaches is the EOM-CCSD approach. Error in excitation energies with EOM-CCSD for states dominated by single excitations is usually about 0.1–0.3 eV, and it would be somewhat larger for IPs and EAs of Koopmans type due to a more significant orbital relaxation effect [1]. Another popular EOM/LR-CC approach for excitation energies is the CC2 approach [13] where single-excitation amplitudes are treated as zeroth order, while the double-excitation amplitudes are approximated to be correct to the first order. The CC2 approach is computationally less demanding than EOM-CCSD. Excitation energies with CC2 has been shown to agree even better with those of high- level EOM/LR-CC approaches than EOM-CCSD especially for valence-type excitations of single-reference molecules [14]. However, Rydberg-type excitations are less accurate and CC2 is also found to be sensitive to a strong correlation effect. Besides energies, analytic energy gradient for excited states [15], ionized states [16], and electron-attached states based on EOM-CC approaches have also been implemented, which greatly facilitates determination of equilibrium or transition-state structures and other first-order properties such as dipole moment of corresponding states.

Scalar-relativistic effects can readily be incorporated in CC and EOM-CC calculations especially when untransformed two-electron interactions are employed. On the other hand, it is not as straightforward to include SOC effects in CC and EOM-CC calculations due to spin symmetry breaking and complex arithmetic particularly for low-symmetry molecules. In addition, double point group symmetry and time-reversal symmetry need to be exploited to achieve an efficient implementation. Kramers-restricted closed-shell CCSD [17] and Kramers-unrestricted open-shell CCSD(T) [18] based on four-component Dirac-Coulomb reference have been reported by Visscher et al. The largest Abelian subgroup of the corresponding double point group is exploited in their implementations. Similar to spin adaption in nonrelativistic CC calculations, Kramers adaption is nontrivial for open-shell systems. These works are extended recently to general-order relativistic CC models [19] using the string-based techniques so that calculations based on relativistic CC approach with arbitrary excitation levels can be carried out. The CC approach with SOC using relativistic effective core potentials (RECP) [20] has also been reported by Lee et al. at the CCSD and CCSD(T) levels [21] and by Hirata et al. at general order [22].

A multireference problem shows up for many open-shell systems when SOC is present and CCSD calculations will be problematic for those systems. Fleig et al. reported a general active space (GAS) CC approach of general order within the four-component formalism [23]. This GAS-CC approach is able to tackle multireference

problems such as dissociation of HBr. However, their implementation is based on a large-scale configuration-interaction technique and computational scaling is $n_o^{n+2} n_v^{n+2}$, while it is only $n_o^n n_v^{n+2}$ with ordinary CC approach, where n_o and n_v are the number of occupied and virtual orbitals/spinors, respectively, and n is the highest excitation level in the cluster operator. Their approach can only be applied to systems with less than 12 correlated electrons and with a medium-size basis set. This GAS-CC approach has recently been fulfilled in a Kramers-restricted form with string-based formalism [24]. Computational scaling of this new implementation is the same as that of conventional CC approach.

In the CC approaches mentioned above, SOC is always included at the SCF level, and two-component or four-component spinors are employed in calculation of correlation energies. It has been demonstrated that the computational effort of two-component CC approaches is the same as that of four-component CC methods if the same level of CC approach and the same number of (occupied and virtual) orbitals are considered [25]. Furthermore, the number of floating-point operations of a two-component or four-component Kramers-restricted CCSD method is about 32 times that of a scalar relativistic or nonrelativistic CCSD calculation when a twofold symmetry element exists [17]. For real double point groups, this factor reduces to eight, while it will be significantly larger for quaternion groups.

An alternative way to treat SOC in CC approaches is to include SOC in post-SCF treatment. This method is proposed first by Eliav et al. [26], and it is implemented in an efficient way recently by Wang et al. at CCSD and CCSD(T) levels [27] for closed-shell systems with RECPs. This treatment of SOC is conceptually similar to the one-step spin-orbit configuration-interaction approach. Real spin orbitals are adopted in CC calculations, and an SOC operator represented by a one-electron operator is used. Two-electron integrals are thus identical to those in nonrelativistic or scalar-relativistic calculations. The number of floating-point operations with this approach at the CCSD level for closed-shell systems is about 8–20 times that of nonrelativistic CCSD method even for systems without any symmetry [27]. Time-reversal symmetry and spatial symmetry are exploited in implementation to reduce computational effort. Practical calculations show that the wall time for an SOC-CCSD calculation is about 10–12 times that of a spin-adapted closed-shell nonrelativistic or scalar-relativistic CCSD calculation [28]. Neglecting SOC in the SCF step usually means that orbital relaxation due to SOC is not taken good care of. However, most of orbital relaxation effects can be accounted for via single excitations in the cluster operator with orbital-unrelaxed CC calculations. Results indeed show that this CC approach is able to provide a rather accurate description on SOC effects of even superheavy p-block element compounds [27,29]. Analytic first- [30] and second-order [31] derivatives of total energy have also been reported at the CCSD and CCSD(T) levels for closed-shell molecules with SOC using RECPs. In fact this approach can be extended readily to make use of other relativistic Hamiltonian as long as SOC can be separated out and represented (or approximated) as a one-electron operator. This SOC-CC approach can also be applied to open-shell systems; however, it would be difficult to exploit time-reversal symmetry and spatial

symmetry in this case. Furthermore, a convergence problem rises up for spatially degenerate states since SOC will couple degenerate states.

In fact SOC is usually more important for excited states as well as open-shell systems, and EOM-CC approach can be used to account for SOC effects of these states with high accuracy. The EOM-CC approach with SOC has been reported for EEs, IPs, EAs, and DIPs either with SOC included in post-SCF treatment [32–34] or in the SCF part at CCSD level [35–37] as well as up to CCSDTQ levels [22]. Hubert presented an EOM/LR-CC approach for excitations based on four-component GAS-CC of general order; however, it is based on a CI technique which renders its high computational cost [38]. A different way to treat SOC with EOM-CC is to include SOC perturbatively, and it has been fulfilled by Klein and Gauss using EOM-IP-CCSD [39]. Their method shows promise for SOC splittings of some $^2\Pi$ states. In addition, FSCC in various sectors of Fock space for EEs, IPs, etc. have been implemented [40] at the four-component CCSD level, and it has been extended to intermediate Hamiltonian FSCC approach [41] to avoid intruder-state problem. This chapter will review relativistic EOM-CC theory for EEs, IPs, and EAs based on closed-shell reference with emphasis on treatment of SOC effects.

Relativistic EOM-CC Theory

- Equation-of-motion coupled-cluster theory
  - Basic equations in EOM/LR-CC for EEs, IPs, and EAs

CC equations have first to be solved to calculate EEs, IPs, and EAs with the EOM-CC approach. In CC theory, the wave function is represented by $e^T|\Phi_0\rangle$ where Φ_0 is the reference determinant wave function and is usually chosen as the HF wave function of the corresponding Hamiltonian and T is the cluster operator and defined as the following:

$$T = \sum_{\mu} t_{\mu}\hat{\tau}_{\mu} = \sum_{i,a} t_i^a a_a^+ a_i + \frac{1}{4}\sum_{a,b,i,j} t_{ij}^{ab} a_a^+ a_i a_b^+ a_j + \ldots, \tag{1}$$

where t_μ is the cluster amplitude and $\hat{\tau}$ is the excitation operator, $i, j, k, \ldots$ are indices for occupied orbitals; $a, b, c, \ldots$ are for virtual orbitals; and $p, q, r, \ldots$ denote general orbitals in the reference Φ_0. In CCSD the cluster operator is truncated to the first two terms on the right-hand side (r.h.s.) of Eq. (1). The cluster amplitudes in Eq. (1) are determined by the following equations:

$$< \Phi_\mu | e^{-T} H e^T | \Phi_0 >= 0, \tag{2}$$

where Φ_μ is an excitation determinant defined as $|\Phi_\mu\rangle = \hat{\tau}_\mu|\Phi_0\rangle$. The energy of the reference state is thus

$$E_{CC} =< \Phi_0 | H e^T | \Phi_0 >=< \Phi_0 | e^{-T} H e^T | \Phi_0 > . \tag{3}$$

In the EOM-CC approach, the wave function for an excited state, ionized state, or electron-attached state is represented by $Re^T|\Phi_0\rangle$, where R is the excitation, ionization, or electron-attachment operator if EEs, IPs, or EAs are to be calculated and defined as

$$R = r_0 + \sum_{i,a} r_i^a a_a^+ a_i + \frac{1}{4}\sum_{a,b,i,j} r_{ij}^{ab} a_a^+ a_i a_b^+ a_j + \ldots = \sum_{\nu} r_\nu \hat{\tau}_\nu, \tag{4}$$

$$R = \sum_{i} r_i a_i + \frac{1}{2}\sum_{a,i,j} r_{ij}^{a} a_a^+ a_i a_j + \ldots = \sum_{\nu} r_\nu \hat{\tau}_\nu, \tag{5}$$

$$R = \sum_{a} r^{a} a_a^+ + \frac{1}{2}\sum_{a,b,i} r_{i}^{ab} a_a^+ a_i a_b^+ + \ldots = \sum_{\nu} r_\nu \hat{\tau}_\nu, \tag{6}$$

for EEs, IPs, and EAs, respectively. Note that $\hat{\tau}$ defined in these equations is different from $\hat{\tau}_\mu$ in the cluster operator for IPs and EAs. This R operator is determined using the following eigenvalue equation:

$$e^{-T}HRe^T|\Phi_0\rangle = e^{-T}He^T R|\Phi_0\rangle = \bar{H}R|\Phi_0\rangle = ER|\Phi_0\rangle, \tag{7}$$

where $\bar{H} = e^{-T}He^T$ is the similarity-transformed Hamiltonian, or effective Hamiltonian, and E is the energy of the excited, ionized, or electron-attached state. The fact that R is commutable with the cluster operator T has been used in Eq. (7). In EOM-CCSD calculations, the R operator is truncated to the first three terms for EEs and the first two terms for IPs and EAs. Similar to the CC equations in Eqs. (2) and (3), Eq. (7) is satisfied only in a space spanned by $\{\Phi_0, \Phi_i^a, \Phi_{ij}^{ab}, \ldots\}$ for EEs, $\{\Phi_i, \Phi_{ij}^a, \ldots\}$ for IPs, and $\{\Phi_a, \Phi_i^{ab}, \ldots\}$ for EAs, where Φ_i^a, Φ_{ij}^{ab}, Φ_i, Φ_{ij}^a, Φ^a, and Φ_i^{ab} are defined as 1hole-1particle, 2hole-2particel, 1hole, 2hole-1particle, 1particle, and 2particel-1hole determinants with respect to the reference Φ_0, respectively. This space is defined by summation terms of the R operator in Eqs. (4), (5), and (6). One can easily see from Eqs. (2) and (3) that the reference wave function Φ_0 is also an eigenfunction of the similarity-transformed Hamiltonian with eigenvalue E_{cc} in the space spanned by $\{\Phi_0, \hat{\tau}_\mu\Phi_0\}$ with $\hat{\tau}_\mu$ defined in Eq. (1). It should be noted that eigenvalues of $\bar{H}$ are the same as those of the bare Hamiltonian H in the space spanned by all possible Slater determinants no matter how the cluster operator T is truncated. However, difference between eigenvalues of $\bar{H}$ and those of H rises up when this space is truncated or a truncation on R operator is introduced and truncation in the cluster operator matters in this case.

One can see that the EOM-CC approach closely resembles the CI approach except that $\bar{H}$ is employed in calculations. It is well known that the CI approach usually provides excitation energies that are too high due to a bias toward the ground state at the CISD level. On the other hand, excitation energies with EOM-CCSD are rather accurate for single-excitation states. This is because the matrix elements of the similarity-transformed Hamiltonian $\bar{H}$ between triple-excitation determinants and

single-excitation determinants $\left\langle \Phi_{ijk}^{abc} \middle| \bar{H} \middle| \Phi_l^d \right\rangle$ are of the second order in correlation perturbation, while they are of the first order for the bare Hamiltonian H. Accuracy of energies for excited states dominant by single excitations in EOM-CCSD is thus of second order, while it is only of the first order in CISD [42]. On the other hand, matrix elements of $\bar{H}$ between triple-excitation determinants and double-excitation determinants are of the first order. Excitation energies for states with a significant double-excitation character using EOM-CCSD are thus unreliable since they are only accurate to the first order. An interesting strategy has been proposed by Nooijen to perform a second similarity transformation on $\bar{H}$, the STEOM-CC approach [9], so that matrix elements between double-excitation determinants and single-excitation determinants are of the second order. Excitation energies can be obtained with reasonable accuracy from eigenvalues of the new similarity-transformed Hamiltonian only in the single-excitation space. The resulting equations of the STEOM-CC approach are quite close to those in the intermediate Hamiltonian FSCC approach.

The fact that $\bar{H}$ is a non-Hermitian operator indicates that its eigenfunctions will not be orthogonal to each other and its left-hand side eigenvector will not be complex conjugate to the right-hand side eigenvector of the same eigenvalue. One can solve either the left-hand side eigen-equation or the right-hand side eigen-equation as in Eq. (7) to determine EEs, IPs, or EAs. However, both left-hand side and right-hand side eigenvectors are required to calculate transition properties or properties of target states. The non-Hermitian nature of $\bar{H}$ implies that eigenfunctions of $\bar{H}$ for excited states with the same symmetry as the ground state will not be orthogonal to the ground-state wave function, i.e., Φ_0, and r_0 for excited states with the same symmetry as the reference state would thus be nonzero. Another consequence of the non-Hermitian nature of $\bar{H}$ is that complex eigenvalue would rise up. In actual calculations, complex eigenvalues would be encountered when two states are rather close in energy to each other.

Making use of Eq. (2) the eigenvalue equation in Eq. (7) can readily be written in the following form:

$$\langle \Phi_\nu | \, [\bar{H}, R] \, | \Phi_0 \rangle = (E - E_{CC}) \, \langle \Phi_\nu | \, R \, | \Phi_0 \rangle \, . \tag{8}$$

and the left-hand eigen-equation becomes

$$\langle \Phi_0 | \, L [\bar{H}, \tau_\nu] \, | \Phi_0 \rangle = (E - E_{CC}) \, \langle \Phi_0 | \, L \, | \Phi_\nu \rangle \, . \tag{9}$$

The L operator in Eq. (9) takes the form of a complex conjugate of R operator in Eqs. (4), (5) and (6). These two equations are actually eigen-equation of the matrix $\langle \Phi_0 | \, \hat{\tau}_\nu^+ [\bar{H}, \hat{\tau}_\sigma] \, | \Phi_0 \rangle$ where $\hat{\tau}_\nu$ and $\hat{\tau}_\sigma$ are defined in Eqs. (4), (5), or (6) and eigenvalues of this matrix become excitation energies, ionization energies, or electron affinities of the reference state, instead of the total energy of the corresponding state. In addition, the r_0 term is unnecessary with Eq. (8). The highest excitation level in the R operator may not be the same as that in the T operator. It

should be noted that Eqs. (8) and (9) can only be reached when the number of annihilation operators for the highest excitation level in R is smaller than or equal to that in the cluster operator and the same also holds for creation operators. When this is not the case or when an approximate cluster operator that does not satisfy Eq. (2) is employed in EOM-CC calculations, Eq. (8) will no longer be equivalent to Eq. (7). However, Eq. (8) is still preferred over Eq. (7) to determine EEs, IPs, or EAs in these circumstances since size-intensive EEs, IPs, or EAs can be achieved using Eq. (8).

Equations (8) and (9) can also be obtained based on linear response (LR) theory. In LR theory, a frequency-dependent external potential is introduced, and excitation energies are determined using the fact that the linear response function will have a pole when the frequency of the external field is equal to some excitation energy. The linear response function can be determined from second-order derivatives of the total energy with respect to external perturbation. Time-dependent cluster amplitude satisfies the following equation [7]:

$$i\dot{t}_\mu =< \Phi_\mu|e^{-T(t)}He^{T(t)}|\Phi_0 > + < \Phi_\mu|e^{-T(t)}V(t)e^{T(t)}|\Phi_0 > \tag{10}$$

where $V(t)$ is the external time-dependent potential. The first-order change of the cluster amplitude with respect to a potential V with frequency ω can thus be determined from Eq. (10) as

$$\omega t^1_\mu - < \Phi_\mu|[e^{-T}He^T, T^{(1)}]|\Phi_0 >=< \Phi_\mu|e^{-T}Ve^T|\Phi_0 > \tag{11}$$

where $T^{(1)}$ is the first-order change of the cluster operator. One can easily see that $T^{(1)}$ will be singular when ω is equal to an eigenvalue of the matrix $\langle\Phi_0|\,\hat{\tau}_\nu^+[\bar{H},\hat{\tau}_\sigma]\,|\Phi_0\rangle$. Excitation energies with LR-CC are thus equivalent to those in EOM-CC. IPs or EAs are not as straightforward with LR-CC, but they can be determined as excitation energies to or from a continuum orbital, and the same equations as those in EOM-CC for IPs or EAs can thus be achieved based on LR-CC. Furthermore, when some kind of approximation is introduced in determining the cluster operator, i.e., Eq. (2) or Eq. (10), such as that in CC2, the corresponding approximate equation for EEs can readily be obtained based on LR-CC from the equation that $T^{(1)}$ satisfies, while it is not so easy to achieve such an equation based on EOM-CC. On the other hand, contributions from higher-level excitations not included in the R operator can be introduced more easily with EOM-CC based on perturbation theory.

To facilitate implementation and to reduce computational effort, intermediates F_{pq} and W_{pqrs} are usually introduced which are the one- and two-electron parts of the similarity-transformed Hamiltonian $\overline{H}$, respectively:

$$e^{-T}He^T = E_{CC} + \sum_{p,q} F_{pq}\{a_p^+ a_q\} + \frac{1}{4}\sum_{pq,rs} W_{pqrs}\{a_p^+ a_q^+ a_s a_r\} + \ldots \tag{12}$$

Note that the effective Hamiltonian $\overline{H}$ also contains operators higher than two-body terms. Furthermore, Davidson's algorithm with modification for non-Hermitian matrix is usually adopted to calculate the lowest several or some specified states in EOM-CC calculations. In Davidson's algorithm, approximate eigenvalues and eigenfunctions are first determined by diagonalizing the $\bar{H}$ matrix in a small subspace. This subspace is augmented with error vectors of those approximate eigenfunctions. The $\bar{H}$ matrix in this new subspace is constructed and diagonalized to update the approximate eigenfunctions. This process is iterated until error of the eigenfunction is smaller than a specified criterion. In EOM-CC calculations without SOC, the $\bar{H}$ matrix is non-Hermitian and complex eigenvalues and eigenfunctions may be encounter in this process, although the final results are still real. One can simply treat the real and imaginary parts of the complex eigenfunction as two approximate eigenfunctions with the same eigenvalues.

Computational effort of EOM-CC for an excited state is similar to that for the ground state. On the other hand, it is less demanding in calculating IPs or EAs. Scaling of EOM-CCSD for excitation energies is N^6, same as that of a CCSD calculation, while it is only N^5 for IPs or EAs, where N represents system size. In calculations of IPs and EAs, solving the CC equation becomes the most time-consuming step. Another advantage of EOM-CC for IPs or EAs is that the resulting wave function is free from spin contamination when a closed-shell state is chosen as reference, which is a nontrivial task in CC calculations for open-shell systems. This is important in calculations when SOC is present since SOC will couple states with different spins, and spin contamination would result in spurious splitting of energy levels. Furthermore, by choosing the reference state that can be described by a single- reference method, some states with a significant multiconfiguration character can be treated with EOM-CC for IPs or EAs based on single-reference formalism. Another similar scheme is the EOM-CC method for spin-flip transitions [43], where the singlet state of diradicals can be reached from a triplet reference state. However, spin contamination is inevitable in such calculations.

- Transition and excited-state properties

In the EOM-CC method, the left-hand and right-hand wave functions for the n-th state are represented by $\langle\Phi_0| L_n e^{-T}$ and $e^T R_n |\Phi_0\rangle$, respectively, and the following orthonormalized condition is employed:

$$\langle\Phi_0| L_m R_n |\Phi_0\rangle = \delta_{mn} \tag{13}$$

For ground state, $R_0 = 1$ and $L_0 = 1 + \Lambda$ with Λ being the de-excitation operator introduced in analytic energy gradient for CC theory. r_0 in R_n is usually determined with the orthogonal condition between the left-hand ground-state wave function and the right-hand n-th state wave function using Eq. (13). On the other hand, l_0 for the n-th state is always zero due to orthogonality between the left-hand n-th state wave function and the right-hand ground-state wave function.

Transition properties between an excited state and the ground state or between two excited states in EOM-CC are calculated directly based on the left- and right-hand wave function as the following:

$$\langle m|\hat{A}|n\rangle = \langle \Phi_0 \left| L_m e^{-T} \hat{A} e^{T} R_n \right| \Phi_0 \rangle. \tag{14}$$

In LR-CC theory, transition properties are determined from residue of the response function at the poles. The linear response function as the second-order derivative using the $2n + 1$ rule can be calculated as [7]

$$\begin{aligned} << X, Y >>_{\omega_y} = \frac{d^2 E_{CC}}{dx dy} &= \langle \Phi_0 | \, (1 + \Lambda)[e^{-T} X e^{T}, T^y] \, | \Phi_0 \rangle \\ &+ \langle \Phi_0 | \, \Lambda^y e^{-T} X e^{T} \, | \Phi_0 \rangle \, , \end{aligned} \tag{15}$$

where T^y and Λ^y are the first- order response of the T and Λ operator due to the external perturbation Y with frequency ω_y. T^y is determined by Eq. (11). The time-dependent Λ operator satisfies the following equation when an external time-dependent potential is present:

$$\begin{aligned} -i\dot{\lambda}_\mu = &< \Phi_0 | (1 + \Lambda(t))[e^{-T(t)} H e^{T(t)}, \hat{\tau}_\mu] | \Phi_0 > \\ &+ < \Phi_0 | (1 + \Lambda(t))[e^{-T(t)} V(t) e^{T(t)}, \hat{\tau}_\mu] | \Phi_0 >, \end{aligned} \tag{16}$$

and Λ^y can thus be determined using the following equation:

$$\begin{aligned} -\omega_y \lambda_\mu^y = &< \Phi_0 | (1 + \Lambda)[[e^{-T} H e^{T}, T^y], \hat{\tau}_\mu] | \Phi_0 > \\ &+ < \Phi_0 | \Lambda^y [e^{-T} H e^{T}, \hat{\tau}_\mu] | \Phi_0 > \\ &+ < \Phi_0 | (1 + \Lambda)[e^{-T} Y e^{T}, \hat{\tau}_\mu] | \Phi_0 > . \end{aligned} \tag{17}$$

Transition properties are obtained from residue of the linear response function:

$$\lim_{\omega_y \to \omega_n} (\omega_y - \omega_n) << X, Y >>_{\omega_y} = < 0|X|n >< n|Y|0 > . \tag{18}$$

To calculate residue of linear response functions, residue of T^y and Λ^y needs to be determined first. Residue of T^y can be obtained from Eq. (11):

$$\lim_{\omega_y \to \omega_n} (\omega_y - \omega_n) T^y = \langle \Phi_0 | \, L_n e^{-T} Y e^{T} \, | \Phi_0 \rangle \, R_n. \tag{19}$$

With Eqs. (17) and (19), residue of Λ^y will be written in the following form:

$$\lim_{\omega_y \to \omega_n} (\omega_y - \omega_n) \Lambda^y = \langle \Phi_0 | \, L_n e^{-T} Y e^{T} \, | \Phi_0 \rangle \, Z_n. \tag{20}$$

Z_n in the above equation is a de-excitation operator similar to Λ, and it satisfies the following equation based on Eq. (17):

$$-\omega_n Z_\mu^n =< \Phi_0|(1+\Lambda)[[\bar{H}, R_n], \hat{\tau}_\mu]|\Phi_0 > + < \Phi_0|Z^n[e^{-T} H e^T, \hat{\tau}_\mu]|\Phi_0 > . \tag{21}$$

Residue of the linear response function in LR-CC can thus be determined from the above equations as:

$$\lim_{\omega_y \to \omega_n} (\omega_y - \omega_n) << X, Y >>_{\omega_y} = [\langle\Phi_0|\,(1+\Lambda)[e^{-T} X e^T, R_n]\,|\Phi_0\rangle + \langle\Phi_0|\,Z_n e^{-T} X e^T\,|\Phi_0\rangle]\,\langle\Phi_0|\,L_n e^{-T} Y e^T\,|\Phi_0\rangle\,. \tag{22}$$

Transition properties with LR-CC can be achieved by comparing Eqs. (18) and (22):

$$< 0|X|n >= \langle\Phi_0|\,(1+\Lambda)[e^{-T} X e^T, R_n]\,|\Phi_0\rangle + \langle\Phi_0|\,Z_n e^{-T} X e^T\,|\Phi_0\rangle\,. \tag{23}$$

$$< n|Y|0 >= \langle\Phi_0|\,L_n e^{-T} Y e^T\,|\Phi_0\rangle\,. \tag{24}$$

Transition properties between two excited states are determined in a similar way from residue of a quartic response function, and an equation similar to Eqs. (21) and (23) will be achieved except that $1+\Lambda$ is replaced by L_n and that ω_n in Eq. (21) is replaced by energy difference between the two excited states:

$$< m|X|n >= \langle\Phi_0|\,L_m[e^{-T} X e^T, R_n]\,|\Phi_0\rangle + \langle\Phi_0|\,Z^{mn} e^{-T} X e^T\,|\Phi_0\rangle + \langle\Phi_0|\,(1+\Lambda) e^{-T} X e^T\,|\Phi_0\rangle\,\delta_{mn}\,, \tag{25}$$

$$(\omega_m - \omega_n) Z_\mu^{mn} =< \Phi_0|L_m[[\bar{H}, R_n], \hat{\tau}_\mu]|\Phi_0 > + < \Phi_0|Z^{mn}[e^{-T} H e^T, \hat{\tau}_\mu]|\Phi_0 > . \tag{26}$$

Transition properties with Eqs. (23) or (25) are shown to be size intensive, while this is not the case with Eq. (14). However, an additional equation, i.e., Eqs. (21) or (26), has to be solved to calculate transition properties in LR-EOM, while transition properties based on Eq. (14) can be obtained more easily. In addition, the transition matrix $< m|X|n >$ is not a complex conjugate of $< n|X|m >$ due to the non-Hermitian nature of the effective Hamiltonian, and their geometric mean is usually adopted for the transition matrix element.

Some properties such as dipole moments and nuclear forces are defined by the first-order derivative of total energy with respect to a certain external perturbation. Although finite difference techniques can be adopted to calculate the first-order derivative of total energy, availability of analytic energy gradient greatly facilitates calculations for first-order properties and determination of optimized or transition-state structures. Total energy for an excited, ionized, or electron-attached

state with EOM-CC can be written in the following way [15, 16]:

$$E_n = \langle \Phi_0 | L_n \bar{H} R_n | \Phi_0 \rangle = \langle \Phi_0 | L_n [H, R_n] | \Phi_0 \rangle + \langle \Phi_0 | \bar{H} | \Phi_0 \rangle , \tag{27}$$

where Eq. (13) has been employed. To avoid calculating derivatives of R, L, and T operators with respect to external perturbation, the following energy functional is introduced:

$$\bar{E}_n = \langle \Phi_0 | L_n \bar{H} R_n | \Phi_0 \rangle + \sum_\mu \xi_\mu \langle \Phi_\mu | \bar{H} | \Phi_0 \rangle + \varepsilon (1 - \langle \Phi_0 | L_n R_n | \Phi_0 \rangle), \tag{28}$$

where ξ_μ and ε are Lagrange multipliers. One can readily see that the energy functional is always equal to the EOM-CC energy in Eq. (27) and its derivatives with respect to the introduced Lagrange multipliers are zero. Furthermore, the derivative of the energy functional with respect to l_μ and r_μ will also be zero when ε is equal to the EOM-CC energy. On the other hand, ξ_μ is determined from the requirement that the derivative of the energy functional with respect to the cluster amplitude is zero:

$$\frac{\partial \bar{E}_n}{\partial t_\mu} = 0 = \langle \Phi_0 | L_n [\bar{H}, \hat{\tau}_\mu] R_n | \Phi_0 \rangle + \sum_\mu \xi_\mu \langle \Phi_\mu | [\bar{H}, \hat{\tau}_\mu] | \Phi_0 \rangle . \tag{29}$$

The first term on the right-hand side of Eq. (29) can be further expressed as the following:

$$\langle \Phi_0 | L_n [\bar{H}, \hat{\tau}_\mu] R_n | \Phi_0 \rangle = \sum_q \langle \Phi_0 | L_n \bar{H} | \Phi_q \rangle \langle \Phi_q | R_n | \Phi_\mu \rangle, \tag{30}$$

where Φ_q represents the determinant with excitation level higher than that in the R operator. Equation (29) is the same as the Λ equations used in analytic energy gradient in CC theory except for the constant term, i.e., the term in Eq. (30). The analytic energy gradient of EOM-CC energy is equal to that of the energy functional in Eq. (28) and can be calculated with the following equation:

$$\frac{dE_n}{dx} = \frac{d\bar{E}_n}{dx} = \langle \Phi_0 | L_n e^{-T} \frac{dH}{dx} e^T R_n | \Phi_0 \rangle + \sum_\mu \xi_\mu \langle \Phi_\mu | e^{-T} \frac{dH}{dx} e^T | \Phi_0 \rangle . \tag{31}$$

This expression is in accord with the $2n+1$ rule, which implies that the first-order derivative of a wave function is not required when calculating the first-order derivative of the total energy. One can see from this equation that ξ_μ is introduced to represent response of the R and L operator as well as the cluster operator T due to external perturbation. One-particle and two-particle density matrices for the excited

state are usually introduced to facilitate calculation of analytic energy gradient as the following:

$$D_{pq} = \langle \Phi_0 | L_n e^{-T} a_p^+ a_q e^T R_n | \Phi_0 \rangle + \sum_\mu \xi_\mu \langle \Phi_\mu | e^{-T} a_p^+ a_q e^T | \Phi_0 \rangle , \tag{32}$$

$$\Gamma_{pqrs} = \langle \Phi_0 | L_n e^{-T} a_p^+ a_q^+ a_s a_r e^T R_n | \Phi_0 \rangle + \sum_\mu \xi_\mu \langle \Phi_\mu | e^{-T} a_p^+ a_q^+ a_s a_r e^T | \Phi_0 \rangle . \tag{33}$$

It should be noted that the orbital relaxation effect is not taken into consideration with the above density matrices. In actual calculations, the orbital relaxation effect may be neglected when atomic orbitals do not depend on the perturbation. Properties such as dipole moment of the excited state can thus be determined with density matrices in Eqs. (32) and (33). On the other hand, orbital relaxation effects must be included when basis functions depend on the perturbation. This is the case when nuclear forces are to be calculated since atomic orbital basis functions centered at nuclear positions are customarily used in quantum chemistry.

The analytic energy gradient can be determined alternative based on the second total energy expression in Eq. (27) [44]. Similar to the above treatment, the following energy functional is introduced:

$$\begin{aligned} \bar{E}_n = & \langle \Phi_0 | L_n [\bar{H}, R_n] | \Phi_0 \rangle + \langle \Phi_0 | \bar{H} | \Phi_0 \rangle + \sum_\mu \zeta_\mu \langle \Phi_\mu | \bar{H} | \Phi_0 \rangle \\ & + \varepsilon (1 - \langle \Phi_0 | L_n R_n | \Phi_0 \rangle), \end{aligned} \tag{34}$$

and ζ_μ satisfies the following equation:

$$\begin{aligned} \frac{\partial \bar{E}_n}{\partial t_\mu} = & \langle \Phi_0 | L_n [[\bar{H}, R_n], \hat{\tau}_\mu] | \Phi_0 \rangle + \langle \Phi_0 | [\bar{H}, \hat{\tau}_\mu] | \Phi_0 \rangle \\ & + \sum_\mu \zeta_\mu \langle \Phi_\mu | [\bar{H}, \hat{\tau}_\mu] | \Phi_0 \rangle = 0. \end{aligned} \tag{35}$$

Density matrices and energy gradient can be calculated in a similar manner as in Eqs. (31), (32) and (33) and are equivalent in these two approaches. It should be noted that density matrices of the excited state can be obtained alternatively based on the response theory in Eq. (25) with $m = n$, except that ζ_μ is the summation of Z_μ in Eq. (26) and λ_μ in the Λ operator.

- Implementation of relativistic EOM-CC approach

Basic equations in relativistic EOM-CC are the same as those in nonrelativistic EOM-CC. However, difference in implementation details is significant when SOC

is included. This is mainly due to complex arithmetic and broken spin symmetry. This section will discuss new features in implementation of EOM-CC when SOC is present.

- Time-reversal symmetry and spatial symmetry

Time-reversal symmetry is trivial in closed-shell nonrelativistic or scalar-relativistic calculations, and it maps spin α orbits with spin β orbitals. On the other hand, it could introduce an additional reduction in calculation effort when SOC is present. For closed-shell systems, orbitals (or spinors) can be classified to two sets $\{\psi_p\}$ and $\{\psi_{\bar{p}}\}$ where ψ_p and $\psi_{\bar{p}}$ form a Kramers pair with $\psi_{\bar{p}} = K\psi_p$, and K is the time-reversal operator. For spin orbitals, ψ_p is simply a spin α orbital, while $\psi_{\bar{p}}$ is a spin β orbital. Eigenfunction of a system with time-reversible Hamiltonian and an even number of electrons can always be chosen to be invariant with respect to time reverse, although not necessarily. A Slater determinant composed of a set of Kramers pairs is time reversible, and the requirement that the CC wave function be time reversible indicates that the cluster operator should be commutable with the time-reversal operator. The following relations between different cluster amplitude can thus be achieved [17, 27]:

$$t_i^a = (t_{\bar{i}}^{\bar{a}})^*, \quad t_{\bar{i}}^a = -(t_i^{\bar{a}})^*, \tag{36}$$

$$t_{ij}^{ab} = (t_{\bar{i}\bar{j}}^{\bar{a}\bar{b}})^*, \quad t_{ij}^{\bar{a}\bar{b}} = (t_{\bar{i}\bar{j}}^{ab})^*, \quad t_{i\bar{j}}^{ab} = -(t_{\bar{i}j}^{\bar{a}\bar{b}})^*, \quad t_{ij}^{a\bar{b}} = -(t_{\bar{i}\bar{j}}^{\bar{a}b})^*, \quad t_{i\bar{j}}^{a\bar{b}} = (t_{\bar{i}j}^{\bar{a}b})^*, \tag{37}$$

With these relations, only a certain cases of cluster amplitudes need to be decided and the other cases can be determined either with the above equations or with the permutation symmetry. In calculations without SOC, spin cases of $t_i^a, t_{\bar{i}}^{\bar{a}}, t_{ij}^{ab}, t_{\bar{i}\bar{j}}^{\bar{a}\bar{b}}$, and $t_{i\bar{j}}^{a\bar{b}}$ are required, and the other spin cases such as $t_{\bar{i}}^a, t_{i\bar{j}}^{ab}, t_{ij}^{a\bar{b}}$, and $t_{i\bar{j}}^{ab}$ are zero, since they do not conserve spin. For closed-shell systems without SOC, time-reversal symmetry implies $t_i^a = t_{\bar{i}}^{\bar{a}}$ and $t_{ij}^{ab} = t_{\bar{i}\bar{j}}^{\bar{a}\bar{b}}$. Furthermore, an additional constraint on double-excitation amplitudes exists for wave functions with S = 0: $t_{ij}^{ab} = t_{i\bar{j}}^{a\bar{b}} - t_{i\bar{j}}^{b\bar{a}}$. Spin cases of the cluster amplitudes to be considered in CCSD calculations for closed-shell systems are thus only $\{t_i^a, t_{i\bar{j}}^{a\bar{b}}\}$. On the other hand, when SOC is included in post-SCF treatment, the reference determinant Φ_0 is the Hartree-Fock wave function of the scalar-relativistic Hamiltonian, and spin orbitals are employed in calculations. Much more spin cases of the cluster amplitudes need to be calculated since SOC will couple states with different spins. In a recent work [27], amplitudes of spin cases $\{t_i^a, t_{\bar{i}}^a, t_{ij}^{ab}, t_{i\bar{j}}^{a\bar{b}}, t_{i\bar{j}}^{ab}, t_{ij}^{a\bar{b}}, t_{\bar{i}\bar{j}}^{ab}\}$ are determined in CCSD calculations with SOC. On the other hand, it is not an easy task to impose time-reversal symmetry for open-shell systems in CC calculations with SOC. This is similar to spin adaption for open-shell systems in nonrelativistic or scalar-relativistic

CC calculations. Kramers-unrestricted calculations are usually employed in such calculations.

Exploitation of spatial symmetry can reduce computational effort significantly. This reduction can be as high as h^2 in electron-correlation calculations, where h is the order of the point group. In almost all correlation calculations without SOC, Abelian single point groups, i.e., D_{2h} and its subgroups, are adopted. When SOC is present, double point group symmetry has to be exploited. Molecular spinors will transform according to the Fermion irreducible representations (IR) of the corresponding double point group. Fermion IRs of the double point group is closely related to time-reversal symmetry, but special care has to be taken to construct symmetry functions that transform as basis of fermion IRs of the double point group and form Kramers pairs at the same time. Two-electron integrals $< pq||rs >$ will vanish unless $\Gamma^*{}_p \otimes \Gamma^*{}_q \otimes \Gamma_r \otimes \Gamma_s$ contains the totally symmetric representation of the molecular point group [17, 19], where Γ_p represents IR of spinor p. Similar rules hold for cluster amplitudes and amplitude t_{ij}^{ab} will be nonzero only when $\Gamma^*{}_a \otimes \Gamma^*{}_b \otimes \Gamma_i \otimes \Gamma_j$ contains the totally symmetric representation. This rule stems from the requirement that the symmetry of the wave function is determined by its Slater determinant part and the cluster operator should be totally symmetric Abelian double point groups, e.g., S_n^*, C_n^* are employed to facilitate determination of the direct product of these IRs [45]. For groups with a mirror-plane or a twofold axis, a symmetry function and its Kramers pair belong to different IRs or different columns of a same IR. This means the cluster amplitudes with odd number of bars in Eqs. (36) and (37) will be zero. Furthermore, for systems containing D_2 or C_{2v} as subgroups, a symmetry function and its Kramers pair can be chosen to belong to different columns of a two-dimensional IR and the involving integrals as well as amplitudes can thus be made real. However, these groups are non-Abelian groups. This problem can be solved by choosing a proper Abelian subgroups of the non-Abelian group [45], where a symmetry function and its Kramers pair belong to different IRs in the Abelian subgroup. Property of the non-Abelian group can still be exploited with this scheme.

When SOC is only included in post-SCF treatment, molecular orbitals that transform according to IRs of the corresponding single point group are employed. Furthermore, the wave function of a system with an even number of electrons will also transform according to boson IRs of the double point group when SOC is present. It is thus possible to exploit symmetry of Abelian single groups, i.e., D_{2h} and its subgroups in such calculations, and the fermion IRs of the corresponding double group are not required. The cluster operator should be totally symmetric and its symmetry is the direct product between symmetry of its spin part and symmetry of its spatial part. This indicates that the spatial part symmetry for a specified spin case of cluster operator should be equal to symmetry of its spin part. In the following, symmetry for a certain spin case of the t-amplitude actually refers to symmetry of its spatial part, i.e., the direct product of IRs of the involving orbitals. Based on the theory of tensor operators for angular momentum and time-reversal symmetry for cluster operators in Eqs. (36) and (37), symmetry for different

spin cases of the cluster amplitudes for D_{2h} and its subgroups can be determined as the following [28]: the real part and imaginary part of t-amplitudes with an even number of β spin orbital indices, i.e., $t_i^a, t_{ij}^{ab}, t_{i\bar{j}}^{a\bar{b}}, t_{\bar{i}\bar{j}}^{ab}$ transform as the totally symmetric representation and Γ_{l_z}, respectively; while for t-amplitudes with an odd number of β spin orbital indices, i.e., $t_{\bar{i}}^a, t_{i\bar{j}}^{ab}, t_{ij}^{a\bar{b}}$, their real part and imaginary part should transform as Γ_{l_y} and Γ_{l_x}, respectively, where Γ_{l_x}, Γ_{l_y}, and Γ_{l_z} are the IR of the angular momentum operator along x, y, and z direction, respectively. This symmetry rule will also apply to other quantities such as density matrices as well as intermediates. Using the fact that the direct product of Γ_{l_x} and Γ_{l_y} is Γ_{l_z} for Abelian single point groups, one can deduce that real and imaginary parts of the amplitudes or the intermediates will belong to a different IR for D_{2h}, C_{2v}, and D_2 groups since Γ_{l_z} is not the totally symmetric representation for these groups. This means that the amplitudes or the intermediates will be either real or pure imaginary for D_{2h}, C_{2v}, and D_2 groups. This is consistent with the case when SOC is introduced in SCF calculations. However, this property is not employed directly in implementations. The real and imaginary parts of the amplitudes or intermediates for different spin cases can be calculated separately based on their own symmetry without scarifying efficiency

It is difficult to impose time-reversal symmetry on the reference wave function of an open-shell system, and time-reversal symmetry in EOM-CC calculations with closed-shell reference will be discussed here. In EOM-CC for excitation energies of closed-shell systems, the r-amplitudes in Eq. (4) satisfy similar relations as in Eqs. (36) and (37), and r_0 must be a real number since one can always require the corresponding wave function to be invariant under time reverse. On the other hand, in EOM-CC calculations for IPs or EAs with a closed-shell reference, the target state has an odd number of electrons. The wave function for such a system and its Kramers pair will thus be orthogonal and degenerate to each other, i.e., Kramers degeneracy. For the wave function of the form $Re^T|\Phi_0\rangle$ and its time-reversal $R'e^T|\Phi_0\rangle = KRe^T|\Phi_0\rangle$ in EOM-IP-CC calculations with closed-shell reference, r-amplitude of different cases in R and R' satisfies the following relations [28]:

$$r_i' = -(r_{\bar{i}})^*, r_{\bar{i}}' = (r_i)^*, \tag{38}$$

$$r_{ij}'^a = -(r_{\bar{i}\bar{j}}^{\bar{a}})^*, r_{i\bar{j}}'^a = -(r_{j\bar{i}}^{\bar{a}})^*, r_{ij}'^{\bar{a}} = (r_{\bar{i}\bar{j}}^a)^*, r_{i\bar{j}}'^{\bar{a}} = (r_{j\bar{i}}^a)^*,$$

$$r_{\bar{i}\bar{j}}'^a = -(r_{ij}^{\bar{a}})^*, r_{\bar{i}\bar{j}}'^{\bar{a}} = (r_{ij}^a)^*. \tag{39}$$

The r-amplitudes in EOM-EA-CC calculations satisfy similar relations. Exploitation of time-reversal symmetry in EOM-EE-CC calculations is similar to that in CC calculations and time-reversal symmetry will reduce the cases of the r-amplitudes to be calculated. An additional advantage by making use of time-reversal symmetry in EOM-EE-CC calculations is that the matrix element of $\bar{H}$ between two time-reversible wave functions Ψ_1 and Ψ_2 $:< \Psi_1|\bar{H}|\Psi_2 >$ will be real regardless of point group symmetry of the system. Making use of the fact that $\bar{H}$ is invariant

under time reverse, this matrix element is an inner product between two time-reversible functions and should be real since this inner product is also invariant under time reverse. The real nature of the $\bar{H}$ matrix does not mean the r-amplitudes are real. Furthermore, saving in computational effort is mainly due to exploitation of time-reversal symmetry for the r-amplitudes instead of the real nature of the $\bar{H}$ matrix since a matrix with a rather small dimension is diagonalized in Davidson's algorithm. On the other hand, approximate eigenvectors of the real $\bar{H}$ matrix will usually be real and their error vectors to augment the subspace will thus be invariant under time reverse. In EOM-IP-CC or EOM-EA-CC calculations, however, making use of Kramer's symmetry means only one partner of a Kramer's pair needs to be calculated and its time reverse is determined readily with Eqs. (38) and (39). For double point groups with a mirror-plane or a twofold axis, a symmetry function and its Kramers pair are classified according to two different IRs or two different columns of the same IR and time-reversal symmetry can be exploited simply by only calculating states that transform according to one of the two IRs. Note that for systems containing D_2 or C_{2v} as subgroups, the fact that the two Kramers partners belong to two columns of the same IR implies that the $\bar{H}$ matrix will be real. On the other hand, a symmetry function and its time-reversal state belong to the same IR for ${C_1}^*$ and ${C_i}^*$ groups. Time-reversal symmetry for these groups can be exploited in calculating the action of $\bar{H}$ on trial vectors: $\bar{H}R|\Phi_0\rangle$, which is the most time-consuming part in EOM-CC calculations. When $\bar{H}R|\Phi_0\rangle$ is obtained, the action of $\bar{H}$ on time reverse of the trial wave function $KR|\Phi_0\rangle$ can be determined easily using the similar relations as in Eqs. (38) and (39) based on $\bar{H}KR|\Phi_0\rangle = K\bar{H}R|\Phi_0\rangle$. Furthermore, one should ensure that the Kramers partner of each eigenvector is included in the obtained eigenvectors when diagonalizing the $\bar{H}$ matrix in a subspace.

As for spatial symmetry of the r-amplitude in EOM-CC calculations, symmetry of the target state is determined by direct product between symmetry of the R operator and that of the reference state. For closed-shell reference, symmetry of the R operator will thus be equal to that of the target state. When SOC is included in SCF part and spinors are employed in post-SCF calculations, the r-amplitude r_{ij}^{ab} in EOM-EE-CC will be nonzero only when ${\Gamma^*}_a \otimes {\Gamma^*}_b \otimes \Gamma_i \otimes \Gamma_j$ is equal to the symmetry of the R operator. A similar rule also applies to r-amplitudes in EOM-IP/EA-CC calculations. Exploitation of symmetry in EOM-CC calculations closely resembles that in CC calculations except that the cluster operator is totally symmetric, while symmetry of the R operator depends on the symmetry of the target state and reference state.

On the other hand, when SOC is included in post-SCF treatment, symmetry of the r-amplitude in EOM-EE-CC calculations can be determined in a similar way as that of the cluster operator with closed-shell reference. For excited states that can be classified according to IR Γ_I, spatial symmetry for the corresponding r-amplitudes of different spin cases is determined as the following [34]: for r-amplitudes with an even number of β spin orbital indices, e.g., r_i^a, r_{ij}^{ab}, $r_{i\bar{j}}^{a\bar{b}}$, and $r_{\bar{i}\bar{j}}^{ab}$, their real and imaginary part should transform as Γ_I and $\Gamma_{l_z} \otimes \Gamma_I$, respectively; for r-amplitudes

with an odd number of β spin orbital indices, i.e., $r_{\bar{i}}^{a}$, $r_{i\bar{j}}^{ab}$, and $r_{ij}^{a\bar{b}}$, their real and imaginary part should transform as $\Gamma_{l_y} \otimes \Gamma_I$, and $\Gamma_{l_x} \otimes \Gamma_I$, respectively. Determining symmetry of different spin cases of r-amplitude is a little bit more complicated in EOM-IP/EA-CC calculations since the target state will transform according to Fermion IR of the double point group. Furthermore, the $D_{2h}{}^*$, $D_2{}^*$, and $C_{2v}{}^*$ double groups are non-Abelian groups, and their fermion IRs are two-dimensional. Unlike the case with SOC included in the SCF part where complex Abelian groups such as C_n and S_n group with $n > 2$ can be employed, real spin orbitals are always adopted and complex Abelian groups are avoided. To circumvent this problem, one can make use of symmetry of the effective Hamiltonian matrix elements, i.e., intermediates, with different spin cases [32]. The symmetry rule for different spin cases of the intermediates is the same as that of the cluster amplitudes. The real and imaginary part of the matrix elements of $\bar{H}$ between two determinants with an even number of spin β orbitals or two determinants with an odd number of spin β orbitals will be nonzero unless the direct product of the spatial symmetry of the two determinants is totally symmetric or Γ_{l_z}, respectively. On the other hand, the real and imaginary part of the matrix elements of $\bar{H}$ between one determinant with an even number of spin β orbital indices and another determinant with an odd number of spin β orbital indices will vanish unless the direct product of their spatial symmetry is Γ_{l_y} and Γ_{l_x}, respectively. With this property, spatial symmetry for the r-amplitudes in EOM-IP/EA-CC calculations with SOC included in post-SCF treatment can be determined as the following: if a certain r-amplitude with an even number of spin β orbital indices is chosen to be real and its spatial symmetry is Γ_I, spatial symmetry for real and imaginary part of r-amplitudes with an even number of β spin orbital indices will be Γ_I and $\Gamma_{l_z} \otimes \Gamma_I$, respectively; on the other hand, spatial symmetry for the real and imaginary part of r-amplitudes with an odd number of β spin orbital indices will be $\Gamma_{l_y} \otimes \Gamma_I$ and $\Gamma_{l_x} \otimes \Gamma_I$, respectively. Single point group symmetry can thus be exploited even in calculating states that transform according to Fermion IRs of double point groups. It should be noted that Γ_I is only spatial symmetry of the real part of r-amplitudes with an even number of spin β orbital indices, and symmetry of the target state cannot be determined by it since one can always multiply the wave function with some phase factor. There is no such freedom in EOM-EE-CC calculations because of the requirement that the target wave function should be invariant under time reverse. The same IPs or EAs will be obtained no matter which Γ_I is chosen for groups without inversion center. On the other hand, Γ_I and symmetry of the target state will be both of g-symmetry or u-symmetry for groups with inversion center. Note that the above argument only applies to D_{2h} and its subgroups.

- Excited state and transition properties in EOM-IP/EA-CC method

Excited-state properties are usually calculated as derivative of total energy with respect to corresponding perturbations. Density matrices defined in Eqs. (32) and (33) are introduced to calculate excited-state properties. Transition and excited-state

properties based on EOM-EE-CC method with SOC can be obtained in basically the same way as those in nonrelativistic or scalar-relativistic cases. On the other hand, they are more involved in EOM-IP/EA-CC calculations when SOC is present. One always has to take into account the degenerate Kramers pair in these calculations. For operators that are commutable with a time-reversal operator such as dipole moment, their expectation values in a state are the same as those with its Kramers partner: $< \Psi|A|\Psi > = < K\Psi|A|K\Psi >$, and the matrix element of this type of operators between a Kramers pair will be zero: $< \Psi|A|K\Psi > = 0$. In EOM-IP/EA-CC method, $< \Psi|A|\Psi >$ will be a complex number since $|\Psi >$ and $< \Psi|$ are not a complex conjugate to each other and is a complex conjugate to $< K\Psi|A|K\Psi >$. The expectation value for operator A is determined from the real part of $< \Psi|A|\Psi >$. However, time-reversal symmetry cannot be fully exploited in this way. It would be advantageous to calculate the expectation value of such operators with density matrix defined from $1/2(< \Psi|A|\Psi > + < K\Psi|A|K\Psi >)$ since time-reversal symmetry can be made use of. The average of density matrices corresponding to the Kramers pair should thus be used:

$$D^0_{pq} = (\langle\Psi| a_p^+ a_q |\Psi\rangle + \langle K\Psi| a_p^+ a_q |K\Psi\rangle)/2, \tag{40}$$

This strategy can be employed in calculating nuclear forces in the EOM-IP/EA-CC approach with SOC. On the other hand, for operators that are odd with time reverse, $AK = -KA$, such as angular momentum and spin, their expectation values in a state are opposite to those with its Kramers partner: $< \Psi|A|\Psi > = - < K\Psi|A|K\Psi >$. Furthermore, matrix element for this type of operators between a Kramers pair will generally be nonzero. The degenerate perturbation theory is required in this case. One has to calculate $< \Psi|A|\Psi >$, $< K\Psi|A|\Psi >$ and to diagonalize the matrix of this operator in the $\{\Psi, K\Psi\}$ space in order to fully determine expectation value of this operator. To exploit time-reversal symmetry, density matrices with the following form that are invariant under time reverse are calculated:

$$\begin{aligned} D^1_{pq} &= i(\langle\Psi| a_p^+ a_q |\Psi\rangle - \langle\Psi| a_p^+ a_q |\Psi\rangle)/2, \\ D^2_{pq} &= i(\langle K\Psi| a_p^+ a_q |\Psi\rangle + \langle\Psi| a_p^+ a_q |K\Psi\rangle)/2, \\ D^3_{pq} &= (\langle\Psi| a_p^+ a_q |K\Psi\rangle - \langle K\Psi| a_p^+ a_q |\Psi\rangle)/2 \end{aligned} \tag{41}$$

where D^1_{pq}, D^2_{pq}, and D^3_{pq} are used to determine $< \Psi|A|\Psi >$, the real and imaginary part of $< K\Psi|A|\Psi >$, respectively. It should be noted that if the ξ_μ term in Eqs. (32) and (33) is not considered in calculating density matrices, i.e., neglecting the response of the cluster and the R, L operators with respect to external perturbation, saving in computational effort would not be achieved by exploitation of time-reversal symmetry. However, excited-state properties are calculated with higher accuracy when the ξ_μ term is taken into account. Exploitation

of time-reversal symmetry can result in reduction in computational effort when solving equations of ξ_μ. That is the reason why calculating three density matrices listed in Eq. (41) is preferred over calculating two density matrices corresponding to $\langle\Psi|\, a_p^+ a_q\,|\Psi\rangle$ and $\langle\Psi|\, a_p^+ a_q\,|K\Psi\rangle$ in determination of the expectation value for such operators. The above scheme can be used in calculating EPR g-tensors based on the EOM-IP/EA-CC approach with SOC where the energy derivative with respect to external magnetic field is required.

Things are more complicated when transition properties are to be calculated. The following relations can be obtained for operators that are even with time reverse: $< \Psi_1|A|\Psi_2 >=< K\Psi_1|A|K\Psi_2 >^*, < \Psi_1|A|K\Psi_2 >= - < K\Psi_1|A|\Psi_2 >^*$. On the other hand, transition matrix elements for operators that are odd with time reverse satisfies $< \Psi_1|A|\Psi_2 >= - < K\Psi_1|A|K\Psi_2 >^*, < \Psi_1|A|K\Psi_2 >=< K\Psi_1|A|\Psi_2 >^*$. It should be noted that a linear combination of an eigenfunction and its time reverse is also an eigenfunction for the same state due to Kramers degeneracy. One can thus choose a proper linear combination so that $< \Psi_1|A|K\Psi_2 >$ is equal to zero. The transition property for operator A between two states can thus be determined using $\sqrt{|\langle\Psi_1|\, A\,|\Psi_2\rangle|^2 + |\langle\Psi_1|\, A\,|K\Psi_2\rangle|^2}$. It can be calculated either based on left-hand and right-hand eigenfunction of $\bar{H}$ as in Eq. (14) or based on linear response theory directly. If linear response theory is to be employed, equations involving Z_μ as in Eq. (21) have to be solved. To exploit time-reversal symmetry, all the four density matrices defined in Eqs. (40) and (41) are to be calculated.

- Results

To illustrate performance of the EOM-EE/IP/EA-CC approach with SOC, EEs, IPs, and EAs for some 6s- and 6p-block atoms and cations are listed in Table 1 together with experimental data. In these calculations, the small-core energy-consistent pseudopotentials with one-electron SOC operator developed by the Stuttgart/Cologne groups [20] are employed to describe relativistic effects, and the aug-cc-pwCVQZ basis is chosen. The correlation-consistent polarized weighted core-valence basis set is adopted not only to account for core-valence correlation but also to provide a reasonable description on SOC effects. In these calculations, SOC is included in CC and EOM equations with closed-shell reference and the HF wave function of the scalar-relativistic Hamiltonian is used as the reference determinant. All electrons that are not treated via ECPs are included in post-SCF calculations.

According to this table, excitation energies of Hg and Rn with EOM-EE-CCSD are in good agreement with experimental results, and their errors are within 0.1 eV. Excited states for Hg are represented based on LS coupling scheme, while a jj coupling scheme is adopted for Rn. SOC splittings for Hg are obtained from difference between excited-state energies and average of $^3P_0$, $^3P_1$, and $^3P_2$ states. On the other hand, they are represented by energy difference between two neighbor states for Rn. One can also see that the calculated splittings are in accord with experimental data. Results for the ground state of Hg^+ and Rn^+ are their EAs, and

Table 1 Excitation energies, ionization energies, and electron affinities from some atoms using the EOM-CCSD approach with SOC. Data in parentheses are splitting of states due to SOC (Unit: eV)

Atoms		Results				
Hg		$^3P_0$	$^3P_1$	$^3P_2$	$^1P_1$	
	EOM-EE-CCSD	4.641(−0.495)	4.858(−0.278)	5.402(0.266)	6.765(1.629)	
	Exp.[a]	4.667(−0.514)	4.886(−0.295)	5.461(0.280)	6.704(1.523)	
Rn		$(3/2, 1/2)_2$	$(3/2, 1/2)_1$	$(1/2, 1/2)_0$	$(1/2, 1/2)_1$	
	EOM-EE-CCSD	6.804	7.003(0.199)	10.607(3.604)	10.645(0.038)	
	Exp.[a]	6.772	6.942(0.170)	10.660(3.718)	10.794(0.134)	
Hg^+		$^2S_{1/2}(5d^{10}6s^1)$	$^2D_{5/2}(5d^96s^2)$	$^2D_{3/2}(5d^96s^2)$	$^2P_{1/2}(5d^{10}6p^1)$	$^2P_{3/2}(5d^{10}6p^1)$
	EOM-IP-CCSD	10.412	4.368	6.196(1.828)	7.974	9.014(1.040)
	Exp.[a]	10.437	4.403	6.270(1.867)	6.383	7.514(1.131)
Rn^+		$^2P_{3/2}(6s^26p^5)$	$^2P_{1/2}(6s^26p^5)$			
	EOM-IP-CCSD	10.768	3.760			
	Exp.[a]	10.748	3.831			
Au		$^2S_{1/2}(5d^{10}6s^1)$	$^2D_{5/2}(5d^96s^2)$	$^2D_{3/2}(5d^96s^2)$	$^2P_{1/2}(5d^{10}6p^1)$	$^2P_{3/2}(5d^{10}6p^1)$
	EOM-EA-CCSD	9.139	5.044	6.570(1.526)	4.697	5.191(0.494)
	Exp.[a]	9.226	1.136	2.658(1.522)	4.632	5.105(0.473)
Tl		$^2P_{1/2}(6s^26p^1)$	$^2P_{3/2}(6s^26p^1)$	$^2S_{1/2}(6s^27s^1)$	$^2P_{1/2}(6s^27p^1)$	$^2P_{3/2}(6s^27p^1)$
	EOM-EA-CCSD	6.079	0.921	3.258	4.214	4.333(0.119)
	Exp.[a]	6.108	0.966	3.283	4.235	4.359(0.124)

[a]Moore [46]

Table 2 Fine-structure splitting for some diatomic hydrides with a $^2\Pi$ ground state using various approaches (Unit: cm^{-1})

	Exp.	EOMIP-CCSD	TDDFT	Mk-MRCC[a]	MR-CISD[b]	P-EOMIP-CCSD[c]
OH	139.2	140.1	143.9	135.1	137.0	137.3
SH	377.0	375.1	369.6	375.2	377.6	370.2
SeH	1,764.4	1,749.0	1,700.5	1,707.9		1,679.0
TeH	(3,830)	3,912.7	3,536.4			

[a]Muck and Gauss [48]
[b]Berning et al. [49]
[c]Klein and Gauss [50]

those for the other states are EEs from EOM-IP-CCSD calculations using the neutral atoms as references. The obtained EAs for these two anions agree rather well with experimental values with an error of less than 0.04 eV. Excitation energies and SOC splitting for Koopmans states, i.e., the $^2D_{5/2}$ and $^2D_{5/2}$ states for Hg^+ and the $^2P_{1/2}$ state of Rn^+, are calculated to be consistent with experimental data. On the other hand, the $5d^{10}6p^1$ configuration of Hg^+ can only be reached from the reference by ionizing one 6s electron and exciting another one to the 6p orbital. Excitation energies for this configuration are thus overestimated significantly. However, SOC splitting from this state can still be calculated rather accurately. IPs corresponding to the ground state and EEs for the other states are listed for Au and Tl based on EOM-EA-CCSD using Au^+ and Tl^+ as references. Calculated IPs and EEs for a principle electron-attached state are in accord with experimental data once again, and errors are within 0.1 eV. For the $5d^9 6\,s^2$ configuration of Au, which are reached by attaching one electron to the 6s orbital and exciting another one from the 5d to 6s orbital, excitation energies are overestimated by about 4 eV. On the other hand, SOC splitting between the two states that originated from this configuration are still calculated with high accuracy.

To further demonstrate accuracy of the EOM-CCSD approach in describing SOC splitting, the fine-structure splitting of the ground $^2\Pi$ state for OH, SH, SeH, and TeH at their experimental equilibrium bond lengths is calculated with EOM-IP-CCSD and listed in Table 2. Results with other approaches as well as experimental data are also listed for comparison. The exact two-component (X2C) Hamiltonian adopted in the calculation is described in Ref. [47] with a molecular mean field treatment on the two-electron SOC operator. The ANO basis set in uncontracted form is chosen for the involving atoms. Corresponding closed-shell anions of these hydrides are employed as references, and SOC is included in post-SCF treatment in these calculations. According to results in Table 2, splittings of these states with EOM-IP-CCSD are in excellent agreement with experimental results and the EOM-IP-CCSD approach are the best approach among the methods listed in this table in describing SOC splitting of this state. In fact, pilot calculations on splitting of $^2\Pi$ ground state for other diatomic hydrides indicate that results with similar accuracy

can be achieved using the EOM-IP/EA-CCSD approach. It should be noted that a large basis set is necessary to achieve SOC splitting with such high accuracy.

Summary

EOM-CC approach is a black-box approach and EEs, IPs, as well as EAs can be obtained with relative ease. Accuracy of EOM-CCSD results depends on reliability of CCSD approach for the reference state as well as contribution of single excitations to the target state. Results with reasonable accuracy can usually be achieved for single- excitation states or Koopmans states from a reference dominant by a single-reference character. In addition, SOC splitting of degenerate states can be calculated with reasonable accuracy using the EOM-CCSD approach even for states with a significant double-excitation character. It should be noted that a large basis set is usually required to obtain highly accurate SOC splittings even for light elements. As for heavy elements, many basis sets are designed only for scalar-relativistic calculations, and additional steep p functions need to be included or a basis set with core-valence correlation should be used to properly account for SOC effects. Significant saving in computational effort can be achieved by exploitation of time-reversal symmetry and spatial symmetry. However, it is nontrivial to impose time-reversal symmetry for an open-shell reference state. In addition, including SOC in post-SCF treatment in EOM-CC calculations can lead to further reduction in computational effort, which is especially the case for systems with low symmetry. SOC effects can still be calculated with high accuracy even for superheavy element compounds since the single excitations in a cluster operator are able to account for most of orbital relaxation effects. Computational effort for an excited state is similar to that for the ground state, while calculating an ionized state or an electron-attached state is much faster. Comparing with nonrelativistic or scalar-relativistic EOM-CC calculations, it is computationally much more demanding when SOC is included. The computational time for EOM-CC approach with SOC included in post-SCF treatment is usually about 10 times that of nonrelativistic or scalar-relativistic calculations. Furthermore, more states have to be determined since SOC will split degenerate states. On the other hand, due to the use of Abelian group symmetry and broken spin symmetry when SOC is present, assignment of the obtained states is more involved than that in nonrelativistic or scalar-relativistic calculations.

Although EOM-CC calculations with SOC can be applied to open-shell reference, it is still recommended that closed-shell reference states be employed in such calculations. Calculations based on open-shell reference are usually computationally more demanding since it is not easy to impose Kramers symmetry. This may also lead to fictitious splitting of Kramers degenerate states. In addition, a convergence problem would be encountered or multireference approaches have

to be used for some open-shell systems. On the other hand, EOM-CC for IPs and EAs are only applicable to systems with one unpaired electron when closed-shell references are adopted. For systems with two unpaired electrons, they can be reached based on EOM-CC for double ionization or double electron attachment. The most popular EOM-CC approach, i.e., the EOM-CCSD approach, may lead to large errors for states with significant double-excitation character. A higher excitation level is required to achieve a reliable description on those states. Moreover, triples are also necessary to obtain results with higher accuracy. However, the EOM-CCSDT approach with a computational scaling of N^8 is much too demanding. Some approximate approaches such as CC3 [4] and EOM-CCSDT-3 have been introduced in nonrelativistic calculations. In these approximate approaches, the contribution of triple excitations, which scales as N^7, is determined iteratively and it is still rather demanding. In the most successful approach to account for triples of ground state, the CCSD(T) approach, triples are treated in a perturbative way and the N^7 step only needs to be calculated once. Similar treatments on triples in EOM-CC calculations have also been proposed [51]. Including contribution of triples approximately in EOM-CC calculations with SOC would lead to a more accurate description on SOC splittings and is critical for double-excitation states. On the other hand, the EOM-CCSD approach is still not feasible even for medium-size molecules due to its computational scaling of N^6. CC2 is one of the most popular CC approaches in determining excitation energies with reasonable accuracy and a computational scaling of N^5. Other approximate EOM-CCSD approaches also exist such as the EOM-CCSD (2) approach [52], where the cluster operator is approximated with MP2 theory. In addition, CIS(D) [53] can also be considered as an approximate CC approach based on EOM-CCS with a perturbative treatment on double excitations. SOC may be treated in an approximate way, for example, only included in single-excitation amplitudes, which will result in a significant reduction in solving the CCSD equations. It would be interesting to investigate performance of these approaches in describing SOC effects for excited states, ionized states, as well as electron-attached states. Another possible future development is to include SOC in EOM-CC calculations based on perturbation theory or analytic derivative techniques. This should be applicable to processes involving SOC for light elements, such as spin-crossing reactions and phosphorescence. Moreover, the analytic energy gradient for excited states, ionized states, or electron-attached states is also of great importance in locating minimum energy structures or transition-state structures of heavy element compounds, as well as calculating other first-order properties, such as EPR g-tensor when SOC is already considered in calculating wave functions. Effects of SOC on structures have been ignored in most calculations mainly due to a lack of available analytic energy gradient when SOC is present. In fact, it could be important when several states with different symmetries are close to each other, even if the system does not contain very heavy elements. Recent investigation indeed shows that the geometries of ground and the lowest excited states for CH_3I^+ are

affected to a large extent by SOC [54] due to coupling between the close-lying $^2A'$ and $^2A''$ states. Availability of analytic energy gradient would greatly facilitate investigation of structures for such molecules.

References

1. Bartlett RJ, Musial M (2007) Coupled-cluster theory in quantum chemistry. Rev Mod Phys 79:291–352. doi:10.1103/RevModPhys.79.291
2. Raghavachari K, Trucks GW, Pople JA, Head-Gordon M (1989) A fifth-order perturbation comparison of electron correlation theories. Chem Phys Lett 157:479–483. doi:10.1016/S0009-2614(89)87395-6
3. Bartlett RJ (2012) Coupled-cluster theory and its equation-of-motion extensions. WIREs Comput Mol Sci 2:126–138. doi:10.1002/wcms.76
4. Sneskov K, Christiansen O (2012) Excited state coupled cluster methods. WIREs Comput Mol Sci 2:566–584. doi:10.1002/wcms.99
5. Musiał M, Olszówka M, Lyakh DI, Bartlett RJ (2012) The equation-of-motion coupled cluster method for triple electron attached states. J Chem Phys 137:174102. doi:10.1063/1.4763354
6. Stanton JF, Gauss J (1999) A simple scheme for the direct calculation of ionization potentials with coupled-cluster theory that exploits established excitation energy methods. J Chem Phys 111:8785–8788. doi:10.1063/1.479673
7. Christiansen O, Jørgensen P, Hättig C (1998) Response functions from Fourier component variational perturbation theory applied to a time-averaged quasienergy. Int J Quantum Chem 68:1–52. doi:10.1002/(SICI)1097-461X(1998)68:1<1::AID-QUA1>3.0.CO;2-Z
8. Kaldor U (1991) The Fock space coupled cluster method:theory and application. Theor Chim Acta 80:427–439. doi:10.1007/BF01119664
9. Nooijen M, Bartlett RJ (1997) A new method for excited states: similarity transformed equationof-motion coupled-cluster theory. J Chem Phys 106:6441–6448. doi:10.1063/1.474000
10. Nakatsuji H, Hirao K (1978) Cluster expansion of the wavefunction. Symmetry-adapted-cluster expansion, its variational determination, and extension of open-shell orbital theory. J Chem Phys 68:2053–2065. doi:10.1063/1.436028
11. Kánnár D, Szalay PG (2014) Benchmarking coupled cluster methods on valence singlet excited states. J Chem Theory Comput 10:3757–3765. doi:10.1021/ct500495n
12. Meissner L (1998) Fock-space coupled-cluster method in the intermediate Hamiltonian formulation: model with singles and doubles. J Chem Phys 108:9227–9235. doi:10.1063/1.476377
13. Christiansen O, Koch H, Jørgensen P (1995) The second-order approximate coupled cluster singles and doubles model CC2. Chem Phys Lett 243:409–418. doi:10.1016/0009-2614(95)00841-Q
14. Schreiber M, Silva-Junior MR, Sauer SPA, Thiel W (2008) Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. 128:134110. doi:10.1063/1.2889385
15. Stanton JF, Gauss J (1995) Analytic energy derivatives for the equation-of-motion coupled-cluster method: algebraic expressions, implementation and application to the S1 state of HFCO. Theor Chim Acta 91:267–289. doi:10.1007/BF01133076
16. Stanton JF, Gauss J (1994) Analytic energy derivatives for ionized states described by the equation-of-motion coupled cluster method. J Chem Phys 101:8938–8944. doi:0.1063/1.468022
17. Visscher L, Dyall KG, Lee TJ (1995) Kramers-restricted closed-shel1 CCSD theory. Int J Quantum Chem 29:411–419. doi:10.1002/qua.560560844

18. Visscher L, Lee TJ, Dyall KG (1996) Formulation and implementation of a relativistic unrestricted coupled-cluster method including noniterative connected triples. J Chem Phys 105:8769–8776. doi:10.1063/1.472655
19. Nataraj HS, Kállay M, Visscher L (2010) General implementation of the relativistic coupled-cluster method. J Chem Phys 133:234109. doi:10.1063/1.3518712
20. Dolg M, Cao XY (2012) Relativistic pseudopotentials: their development and scope of applications. Chem Rev 112:403–480. doi:10.1021/cr2001383
21. Lee HS, Han YK, Kim MC, Bae C, Lee YS (1998) Spin-orbit effects calculated by two-component coupled-cluster methods: test calculations on AuH, Au2, TlH and Tl2. Chem Phys Lett 293:97–102. doi:10.1016/S0009-2614(98)00760-X
22. Hirata S, Yanai T, Harrison RJ, Kamiya M, Fan PD (2007) High-order electron-correlation methods with scalar relativistic and spin-orbit corrections. J Chem Phys 126:024104. doi:10.1063/1.2423005
23. Fleig T, Sørensen LK, Olsen J (2007) A relativistic 4-component general-order multi-reference coupled cluster method: initial implementation and application to HBr. Theor Chem Acc 118:347–356. doi:10.1007/s00214-007-0265-y
24. Sørensen LK, Olsen J, Fleig T (2011) Two- and four-component relativistic generalized-active-space coupled cluster method: implementation and application to BiH. J Chem Phys 134:214102. doi:10.1063/1.3592148
25. Saue T, Visscher L (2003) Four-component electronic structure methods for molecules. In: Kaldor U, Wilson S (eds) Theoretical chemistry and physics of heavy and superheavy elements. Kluwer Academic, Dordrecht, p 211
26. Eliav E, Kaldor U, Hess BA (1998) The relativistic Fock-space coupled-cluster method for molecules: CdH and its ions. J Chem Phys 108:3409. doi:10.1063/1.475740
27. Wang F, Gauss J, van Wüllen C (2008) Closed-shell coupled-cluster theory with spin-orbit coupling. J Chem Phys 129:064113. doi:10.1063/1.2968136
28. Tu ZY, Yang DD, Wang F, Guo JW (2011) Symmetry exploitation in closed-shell coupled-cluster theory with spin-orbit coupling. J Chem Phys 135:034115. doi:10.1063/1.3611052
29. Kim I, Park YC, Kim H, Lee YS (2012) Spin–orbit coupling and electron correlation in relativistic configuration interaction and coupled-cluster methods. Chem Phys 395:115–121. doi:10.1016/j.chemphys.2011.05.002
30. Wang F, Gauss J (2008) Analytic energy gradients in closed-shell coupled-cluster theory with spin-orbit coupling. J Chem Phys 129:174110. doi:10.1063/1.3000010
31. Wang F, Gauss J (2009) Analytic second derivatives in closed-shell coupled-cluster theory with spin-orbit coupling. J Chem Phys 131:164113. doi:10.1063/1.3245954
32. Tu ZY, Wang F, Li XY (2011) Equation-of-motion coupled-cluster method for ionized states with spin-orbit coupling. J Chem Phys 136:174102. doi:10.1063/1.4704894
33. Yang DD, Wang F, Guo JW (2012) Equation of motion coupled cluster method for electron attached states with spin–orbit coupling. Chem Phys Lett 531:236–241. doi:10.1016/j.cplett.2012.02.014
34. Wang ZF, Tu ZY, Wang F (2014) Equation-of-motion coupled-cluster theory for excitation energies of closed-shell systems with spin-orbit coupling. J Chem Theory Comput 10:5567–5576. doi:10.1021/ct500854m
35. Chaudhuri RK, Panda PK, Das BP (1999) Relativistic coupled-cluster-based linear response theory for ionization potentials of alkali-metal and alkaline-earth-metal atoms. Phys Rev A 60:246–252. doi:10.1103/PhysRevA.60.246
36. Pathak H, Sahoo BK, Das BP, Vaval N, Pal1 S (2014) Relativistic equation-of-motion coupled-cluster method: application to closed-shell atomic systems. Phys Rev A 89:042510. doi:10.1103/PhysRevA.89.042510
37. Pathak H, Ghosh A, Sahoo BK, Das BP, Vaval N, Pal1 S (2014) Relativistic equation-of-motion coupled-cluster method for the double-ionization potentials of closed-shell atoms. Phys Rev A 90:010501(R). doi:10.1103/PhysRevA.90.010501

38. Hubert M, Olsen J, Loras J, Fleig T (2013) General active space commutator-based coupled cluster theory of general excitation rank for electronically excited states: implementation and application to ScH. J Chem Phys 139:194106. doi:10.1063/1.4827638
39. Klein K, Gauss J (2008) Perturbative calculation of spin-orbit splittings using the equation-of-motion ionization-potential coupled-cluster ansatz. J Chem Phys 129:194106. doi:10.1063/1.3013199
40. Visscher L (2001) Formulation and implementation of the relativistic Fock-space coupled cluster method for molecules. J Chem Phys 115:9720–9726. doi:10.1063/1.1415746
41. Infante I, Gomes ASP, Visscher L (2006) On the performance of the intermediate Hamiltonian Fock-space coupled-cluster method on linear triatomic molecules: the electronic spectra of NpO+2, NpO2+2, and PuO2+2. J Chem Phys 125:074301. doi:10.1063/1.2244564
42. Schirmer J, Mertins F (2010) Review of biorthogonal coupled cluster representations for electronic excitation. Theor Chem Acc 125:145–172. doi:10.1007/s00214-009-0597-x
43. Krylov AI (2008) Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: the Hitchhiker's guide to Fock space. Ann Rev Phys Chem 59:433–462. doi:10.1146/annurev.physchem.59.032607.093602
44. Kállay M, Gauss J (2004) Calculation of excited-state properties using general coupled-cluster and configuration-interaction models. J Chem Phys 121:9257–9269. doi:10.1063/1.1805494
45. Visscher L (1996) On the construction of double group molecular symmetry functions. Chem Phys Lett 253:20–26. doi:10.1016/0009-2614(96)00234-5
46. Moore CE (1949/1952/1958) Atomic energy levels. National Bureau of Standards Circular 467, U.S. Government Printing Office, Washington, DC, vols. 1–3
47. Li ZD, Suo BB, Zhang Y, Xiao YL, Liu WJ (2013) Combining spin-adapted open-shell TD-DFT with spin–orbit coupling. Mol Phys 111:3741–3755. doi:10.1080/00268976. 2013. 785611
48. Muck LA, Gauss J (2012) Spin-orbit splittings in degenerate open-shell states via Mukherjee's multireference coupled-cluster theory: a measure for the coupling contribution. J Chem Phys 136:111103
49. Berning A, Schweizer M, Werner HJ, Knowles P, Palmieri P (2000) Spin-orbit matrix elements for internally contracted multireference configuration interaction wavefunctions. Mol Phys 98:1823
50. Klein K, Gauss J (2008) Perturbative calculation of spin-orbit splittings using the equation-of motion ionization-potential coupled-cluster ansatz. J Chem Phys 129:194106
51. Manohar PU, Krylov AI (2008) A noniterative perturbative triples correction for the spin-flipping and spin-conserving equation-of-motion coupled-cluster methods with single and double substitutions. J Chem Phys 129:194105. doi: 10.1063/1.3013087
52. Stanton JF, Gauss J (1995) Perturbative treatment of the similarity transformed Hamiltonian in equationofmotion coupled-cluster approximations. J Chem Phys 103:1064–1076. doi:10.1063/1.469817
53. Head-Gordon M, Rico RJ, Oumi M, Lee TJ (1994) A doubles correction to electronic excited states from configuration interaction in the space of single substitutions. Chem Phys Lett 219:21–29. doi:10.1016/0009-2614(94)00070-0
54. Kim J, Ihee H, Lee YS (2011) Spin–orbit ab initio study of two low-lying states of chloroiodomethane cation. Theor Chem Acc 129:343–347. doi:10.1007/s00214-010-0849-9

High-Accuracy Relativistic Coupled-Cluster Calculations for the Heaviest Elements

26

Ephraim Eliav, Anastasia Borschevsky, and Uzi Kaldor

Contents

Introduction . . . 826
Methodology . . . 828
The Relativistic Hamiltonian . . . 828
The One-Electron Equation . . . 829
Electron Correlation: Fock-Space (FS) and Intermediate Hamiltonian (IH) Coupled Cluster . . . 831
Applications . . . 834
Experimental-Computational Collaborative Research: The IPs of At and Lr . . . 835
Ground State of Rutherfordium: Interplay of Relativity and Correlation . . . 836
Gold and Roentgenium (E111): Local Maximum of Relativistic Effects . . . 838
Cn (E112) and Eka-Tl (E113): What Chemistry? . . . 840
Fl (Element 114): Exotic Chemical Properties? . . . 841
Element 118: Can a Rare Gas Have Positive Electron Affinity? How Important Is QED? . . . 844
Beginning of the Eighth Period: Relativistic Effects Increase . . . 845
Electron Affinities of E119 and Other Alkali Atoms: Accuracy at the 5 meV Level . . . 846
Eka-actinium (E121) vs. Its Homologs La and Ac: When Is the Breit Term Important? . . . 847
Thorium and Eka-thorium (E122): Different Level Structure . . . 848
Summary . . . 851
References . . . 853

E. Eliav • U. Kaldor (✉)
School of Chemistry, Tel Aviv University, Tel Aviv, Israel
e-mail: ephraim@tau.ac.il; ; kaldor@tau.ac.il

A. Borschevsky
Van Swinderen Institute for Particle Physics and Gravity, University of Groningen, AG Groningen, The Netherlands
e-mail: A.Borschevsky@rug.nl

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_34

Abstract

High-accuracy calculations of atomic properties of the heaviest elements, up to element 122, are reviewed. The properties discussed include ionization potentials, electron affinities, and excitation energies, which are associated with the spectroscopic and chemical behavior of these elements and are therefore of considerable interest. Accurate predictions of these quantities require high-order inclusion of relativity and electron correlation, as well as large, converged basis sets. The Dirac-Coulomb-Breit Hamiltonian, which includes all terms up to second order in the fine-structure constant α, serves as the framework for the treatment; higher-order Lamb shift terms are considered in some selected cases. Electron correlation is treated by the Fock-space coupled cluster method, enhanced by the intermediate Hamiltonian scheme, allowing the use of large, converged model (P) spaces. The quality of the calculations is assessed by comparison with available experimental information. Very good agreement is obtained, usually within a few hundredths of an eV, and similar accuracy is expected for the superheavy elements (SHEs), with $Z \geq 104$, for which experimental values are scarce. Many of the properties predicted for these species differ significantly from what may be expected by straightforward extrapolation of lighter homologs, demonstrating that the structure and chemistry of SHEs are strongly affected by relativity. The major scientific challenge of the calculations is to find the electronic structure and basic atomic properties of the SHE and assign its proper place in the periodic table. Significant recent developments include joint experimental-computational studies of the ionization energies of At and Lr, with excellent agreement of experiment and theory. For Lr, calculations were required not only for comparison with experiment; the extraction of the ionization potential from experimental data depended on reliable estimates of atomic excitation energies, obtainable from theory.

Keywords

Heavy elements • Relativity • Electron correlation • Coupled cluster methods

Introduction

The experimental study of heavy and superheavy elements presents considerable difficulties. Many of these elements are rare, toxic, or radioactive. Most actinides, starting with Np, and all transactinides or superheavy elements (SHEs), with $Z \geq 104$, are produced in accelerators, many of them in minute quantities or with short lifetimes [1], limiting their experimental study. Atomic and chemical studies of the heaviest elements shed light on the lower part of the periodic table, revealing trends which may differ from those shown by lighter elements. Their properties cannot always be estimated by simple extrapolation of the relevant group in the periodic table. One fundamental aim of experiment is to reveal the similarities and

differences in the chemistry of the heaviest elements and their lighter homologs. Relativity and in some cases quantum electrodynamics (QED) are known to strongly affect their valence electron shells and, hence, their chemical behavior, causing significant deviations from the trends exhibited by the group. In particular, the relative energies of orbitals may be affected, modifying the electron configurations of atomic ground and excited states. While some chemistry of these elements can be studied in single-atom experiments, spectroscopy usually requires larger amounts of the relevant species. Important atomic properties, such as ionization potentials (IP), electron affinities (EA) and excitation energies (EE), have not yet been measured for the heavier elements, and theoretical investigations are the only source of this information. This situation may be expected to change in the near future, as indicated by several joint experimental-theoretical efforts involving the spectroscopy of these elements reported in recent years. The excitation spectrum of Fm, using a weighable quantity of an isotope with half-life of 20 h, was measured and calculated in 2003 [2], with very good agreement of the two approaches. More recent joint research established the IP of At in 2013 [3]. Of particular interest is the 2015 publication of the IP of Lr, the heaviest actinide [4], using an isotope with a half-life of 27 s and a slow production rate of one atom every few seconds. In this case, theory was important not only for comparing experimental and calculated IPs; the extraction of the experimental IP from the measured data relied on the 2007 high-precision calculation of the excitation energies [5]. Excellent agreement between experimental and computed IPs was found (see section "Experimental–Computational Collaborative Research: The IPs of At and Lr"). Theory can reveal the contributions of electron correlation, relativistic, and QED effects to different atomic and chemical properties. These effects are often large and nonadditive, and their contributions are complicated by the multireference character of many electronic states, presenting severe challenges to theoretical and computational methodology.

Theoretical investigations provide a very powerful, relatively inexpensive, radiation free, and in many cases the sole alternative to experimental research of SHEs. Indeed, advanced state-of-the-art computational methods, such as those discussed below, can nowadays provide theoretical predictions of atomic properties with an accuracy comparable or even superior to experiment. The high-level first-principles approaches designed to benchmark the heavy and superheavy elements should be based on size-extensive, size-consistent, and balanced treatment of the nondynamic electron correlation, due to virtual states with energy close to the state studied, and dynamic correlation, coming from all other virtual states. In particular, they should include relativity right from the outset, at a level consistent with the treatment of electron correlation.

The currently known nuclear, atomic, and chemical properties of the heaviest elements were summarized in a recent book [6]. Reviews of theoretical aspects are also available [7]. The RTAM searchable database [8] provides an up-to-date list of relativistic calculations for atoms and molecules.

Methodology

The methods used below have been presented earlier. Here we give only a short account of the basic equations which provide the framework for the relativistic atomic structure calculations described below. More extensive presentations may be found in [9].

The Relativistic Hamiltonian

The relativistic many-electron Hamiltonian cannot be written in closed form; it may be derived perturbatively from quantum electrodynamics [10]. The simplest analytical form is the Dirac-Coulomb (DC) Hamiltonian, where the nonrelativistic one-electron terms in the Schrödinger equation are replaced by the one-electron Dirac operator h_D,

$$H_{\mathrm{DC}} = \sum_i h_D(i) + \sum_{i<j} 1/r_{ij}, \quad h_D = c\boldsymbol{\alpha}\cdot p + \beta c^2 + V_{\mathrm{nuc}}. \tag{1}$$

$\boldsymbol{\alpha}$ and β are the four-dimensional Dirac matrices, and V_{nuc} is the nuclear attraction operator, with the nucleus modeled as a point or finite-size charge. Only the one-electron terms of the DC Hamiltonian are relativistic, and the two-electron repulsion remains in the nonrelativistic (noncovariant) form. The lowest-order correction to the two-electron term in Coulomb gauge is the frequency-independent Breit operator

$$B_{12} = -\frac{1}{2}[\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2 + (\boldsymbol{\alpha}_1\cdot\boldsymbol{r}_{12})\cdot(\boldsymbol{\alpha}_2\cdot\boldsymbol{r}_{12})/r_{12}^2]/r_{12}, \tag{2}$$

giving rise to the Dirac-Coulomb-Breit (DCB) Hamiltonian

$$H_{\mathrm{DCB}} = \sum_i h_D(i) + \sum_{i<j}(1/r_{ij} + B_{ij}). \tag{3}$$

All equations are in atomic units. The negative-energy solutions of these Hamiltonians are eliminated by the projector operator Λ^+, yielding the no-virtual-pair approximation (NVPA) [11] with the projected Hamiltonians H^+_{DC} and H^+_{DCB},

$$H^+_{\mathrm{DCB}} = \sum_i h_D(i) + \Lambda^+\sum_{i<j}(1/r_{ij} + B_{ij})\Lambda^+. \tag{4}$$

H^+_{DCB} is correct to second order in the fine-structure constant α and is expected to give highly accurate results for all neutral and weakly ionized atoms [12]. Higher quantum electrodynamic (QED) terms are required for strongly ionized species and

may also be necessary for very high-Z elements. An up-to-date review of Lamb shift energy calculations in superheavy atoms may be found in [13].

The One-Electron Equation

The no-virtual-pair DCB Hamiltonian (4) is often used as a starting point for variational or many-body relativistic calculations [14]. The computational procedure is similar to the nonrelativistic case, with the Hartree-Fock orbitals replaced by the four-component Dirac-Fock-Breit (DFB) functions. The spherical symmetry of atoms leads to the separation of the one-electron equation into radial and spin-angular parts. The radial four-spinor has the so-called large component $P_{n\kappa}$ in the upper two places and the small component $Q_{n\kappa}$ in the lower two. The quantum number κ (with $|\kappa| = j + 1/2$) comes from the spin-angular equation, and n is the principal quantum number, which counts the solutions of the radial equation with the same κ. The DFB equation has the form

$$F_\kappa \phi_{n\kappa} = \varepsilon_{n\kappa} \phi_{n\kappa}, \quad \phi_{n\kappa} = \begin{pmatrix} P_{n\kappa}(r) \\ Q_{n\kappa}(r) \end{pmatrix}, \tag{5}$$

where the one-electron DFB operator F_κ is

$$F_\kappa = \begin{pmatrix} V_{\rm nuc} + U^{LL} & c\Pi_\kappa + U^{LS} \\ c\Pi_\kappa^+ + U^{SL} & V_{\rm nuc} + U^{SS} - 2c^2 \end{pmatrix}, \tag{6}$$

$$\Pi_\kappa = -d/dr + \kappa/r, \quad \Pi_\kappa^+ = d/dr + \kappa/r. \tag{7}$$

The superscripts L, S refer to the large and small components, respectively, and $V_{\rm nuc}$ is the nuclear attraction potential. The point-charge model employed in light atom calculations gives rise to significant errors for heavy elements, and the finite extent of the nucleus must be considered. Several models for the distribution of the nuclear charge are commonly used in the literature.In the uniform charge distribution model, for example, the charge of a nucleus of atomic mass A is distributed uniformly over a sphere with radius $R = 2.2677 \times 10^{-5} A^{-1/3}$. The nuclear potential for a nucleus with charge Z is then

$$V_{\rm nuc} = \begin{cases} -Z/r & \text{for } r > R \\ -(Z/2R)(3 - r^2/R^2) & \text{for } r \le R. \end{cases} \tag{8}$$

Two other commonly used nuclear potentials employ either a Fermi or Gaussian charge distribution. Dirac-Fock calculations for all elements up to Mt (E109) [15] found that significant differences exist between point-charge and finite nuclei, while

Table 1 Effect of nuclear model on DF calculated properties of Fm (E100). Values from Ref. [15]. Absolute and percent differences are shown. The finite nuclear models used include the Fermi, Gaussian, and uniformly charged sphere distributions. Energies in hartree, distances in bohr units

	Largest differences between	
	Point and finite nuclei	Finite nuclear models
Total energy	42(0.12 %)	0.53(0.0015 %)
$1s$ orbital energy	16(0.3 %)	0.20(0.0038 %)
$7s$ orbital energy	6×10^{-4}(0.3 %)	7×10^{-6}(0.0034 %)
$\langle r\rangle$ of $1s$	1.2×10^{-4}(0.24 %)	1.5×10^{-6}(0.0030 %)
$\langle r\rangle$ of $7s$	0.010(0.24 %)	1.2×10^{-4}(0.0029 %)
$\langle r^{-1}\rangle$ of $1s$	1.6(1.1 %)	0.024(0.016 %)
$\langle r^{-1}\rangle$ of $7s$	9×10^{-4}(0.3 %)	1.1×10^{-5}(0.004 %)

the three finite nuclei models give very similar values. Table 1 lists some results for atomic Fm. The numerical effect of choosing a different finite nuclear model is two orders of magnitude smaller than the change caused by going from a finite to a point-charge nucleus. These results are typical, indicating that a finite nucleus should be used, but the choice of the exact model is secondary.

The terms U^{LL}, U^{LS}, etc. in the DCB operator (6) represent the one-body mean-field potential, which approximates the two-electron interaction in the Hamiltonian. Similar procedures appear generally in SCF methods. Here [Eq. (6)], the U interactions include the Breit term (2), in addition to the electron repulsion $1/r_{ij}$. The radial functions $P_{n\kappa}(r)$ and $Q_{n\kappa}(r)$ may be obtained by numerical integration or by expansion in a basis (for more details, see [16]). Since the Dirac Hamiltonian is not bound from below, a "variational collapse" may occur if the correct boundary conditions are not properly considered, where admixture of negative-energy solutions may yield energies much below experimental. To avoid this collapse and to keep the variational minimax procedure stable, the basis sets used for expanding the large and small components must obey "kinetic balance" [17]. In the nonrelativistic limit ($c \to \infty$), the small component should be related to the large component by

$$Q_{n\kappa}(r) = (2c)^{-1} \Pi_{\kappa}^{+} P_{n\kappa}(r), \tag{9}$$

where Π_{κ}^{+} is defined in (7). The simplest way to ensure kinetic balance is to derive the small-component basis functions from those used to span the large component by

$$\chi_{\kappa j}^{S} = \Pi_{\kappa}^{+} \chi_{\kappa j}^{L}. \tag{10}$$

It has been shown [18] that G-spinors, with functions spanned in Gaussian-type functions chosen according to (10), satisfy kinetic balance for finite c values if the nucleus is modeled as a uniformly charged sphere.

Electron Correlation: Fock-Space (FS) and Intermediate Hamiltonian (IH) Coupled Cluster

The relativistic coupled cluster (CC) approach in global (usually Gaussian) basis sets has been first implemented in the 1990s [19] and has since become one of the most powerful and generally applicable electron correlation methods for bound states. The CC method is an all-order, size-extensive, systematic, and very accurate many-body approach, as has been shown in many benchmark applications of 4-component relativistic CC methods to heavy and superheavy atoms and molecules. Multireference variants of relativistic 4-component CC methods capable of handling quasi-degeneracies, which are important for open-shell heavy atomic and molecular systems, have also been developed in recent years. In particular, the multireference Fock-space or valence-universal CC scheme [20,21] is applicable to systems with a variable number of particles and is an ideal candidate for merging with QED theory to create an infinite-order size-extensive covariant many-body method applicable to systems with variable numbers of fermions and bosons [9]. The FSCC gave remarkable agreement with experiment for many ground and excited state properties of heavy systems [9,22]. This success makes the scheme a useful tool for reliable predictions of the structure and spectrum of superheavy elements, which are difficult to access experimentally.

The Dirac-Coulomb-Breit Hamiltonian H^+_{DCB} may be rewritten in second-quantized form [11,23] in terms of normal-ordered products of spinor creation and annihilation operators $\{r^+s\}$ and $\{r^+s^+ut\}$,

$$H = H^+_{\mathrm{DCB}} - \langle 0|H^+_{\mathrm{DCB}}|0\rangle = \sum_{rs} f_{rs}\{r^+s\} + \frac{1}{4}\sum_{rstu}\langle rs||tu\rangle\{r^+s^+ut\}, \tag{11}$$

where

$$\langle rs||tu\rangle = \langle rs|tu\rangle - \langle rs|ut\rangle, \tag{12}$$

$$\langle rs|tu\rangle = \int d\mathbf{x}_1 d\mathbf{x}_2 \Psi^*_r(\mathbf{x}_1)\Psi^*_s(\mathbf{x}_2)(r^{-1}_{12} + B_{12})\Psi_t(\mathbf{x}_1)\Psi_u(\mathbf{x}_2). \tag{13}$$

Here f_{rs} and $\langle rs||tu\rangle$ are, respectively, elements of the one-electron Dirac-Fock-Breit and the antisymmetrized two-electron Coulomb-Breit interaction matrices over Dirac four-component spinors. In this re-formulation, the effect of the projection operators Λ^+ is taken over by normal ordering, denoted by the curly braces in (11), which requires annihilation operators to be moved to the right of creation operators as if all anticommutation relations vanish. The Fermi level is set at the top of the highest occupied positive-energy state, while the negative-energy states are simply ignored.

The development of a general multi-root multireference scheme for treating electron correlation effects usually starts from a consideration of the Schrödinger equation for a number (d) of target states,

$$H\Psi^\alpha = E^\alpha\Psi^\alpha \ , \alpha = 1, \ldots, d. \tag{14}$$

The physical Hamiltonian is divided into two parts, $H = H_0 + V$, so that V is a small perturbation to the zero-order Hamiltonian H_0, which has known eigenvalues and eigenvectors, $H_0|\mu\rangle = E_0^{\mu}|\mu\rangle$.

The case of exact or quasi-degeneracy, occurring in many open-shell heavy compound systems, involves the equality or near equality of some energy values E_0^{α}. By adopting the NVPA approximation, a natural and straightforward extension of the nonrelativistic open-shell CC theory emerges. This multireference valence-universal Fock-space coupled cluster approach is presented here briefly; a more detailed description may be found in [20, 21]. The FSCC method defines and calculates an effective Hamiltonian in a d-dimensional model space $P = \sum |\mu\rangle \langle\mu|$, $\mu = 1, .., d$, comprising the most strongly interacting zero-order many-electron wave functions. All other functions are in the complementary Q-space, so that $P + Q = 1$. All d eigenvalues of H_{eff} coincide with the relevant eigenvalues of the physical Hamiltonian,

$$H_{\text{eff}}\Psi_0^{\alpha} = E^{\alpha}\Psi_0^{\alpha},\ \alpha = 1, \ldots, d. \tag{15}$$

The functions $\Psi_0^{\alpha} = \sum_{\mu} c_{\mu}^{a}|\mu\rangle$, with $\alpha = 1, \ldots, d$, describe the projections $P\Psi^{\alpha}$, which constitute the major part of Ψ^{α}. The effective Hamiltonian has the form [24]

$$H_{\text{eff}} = PH\Omega P,\ H_{\text{eff}} = H_0 + V_{\text{eff}}. \tag{16}$$

Ω is the normal-ordered wave operator, mapping the eigenfunctions of the effective Hamiltonian onto the exact ones, $\Omega\Psi_0^{\alpha} = \Psi^{\alpha}$, $\alpha = 1, \ldots, d$. It satisfies intermediate normalization, $P\Omega P = P$. The effective Hamiltonian and wave operator are connected by the generalized Bloch equation, which may be written for a complete model space P in the compact linked form [24]

$$Q[\Omega, H_0]P = Q(V\Omega - \Omega H_{\text{eff}})_{\text{linked}}P. \tag{17}$$

Ω is parametrized exponentially in the coupled cluster method. A particularly compact form is obtained with the normal-ordered form $\Omega = \{\exp(S)\}$.

The Fock-space approach starts from a reference state (closed-shell in our applications, but other single-determinant functions may also be used), correlates it, and then adds and/or removes electrons one at a time, recorrelating the whole system at each stage. The sector (m, n) of the Fock space includes all states obtained from the reference determinant by removing m electrons from designated occupied orbitals, called valence holes, and adding n electrons in designated virtual orbitals, called valence particles. The practical current limit is $m + n \leq 2$, although higher sectors have also been considered [25]. The excitation operator S, defined by the exponential parametrization of Ω, is partitioned into sector operators $S = \sum_{m\geq 0}\sum_{n\geq 0} S^{(m,n)}$. This partitioning allows for partial decoupling of the open-shell

CC equations according to the so-called subsystem embedding condition [20]. The equations for the (m,n) sector involve only S elements from sectors (k,l) with $k \leq m$ and $l \leq n$, so that the very large system of coupled nonlinear equations is separated into smaller subsystems, which are solved consecutively: first, the equations for $S^{(0,0)}$ are iterated to convergence, the $S^{(1,0)}$ (or $S^{(0,1)}$) equations are then solved using the known $S^{(0,0)}$, and so on. This separation, which is exact, reduces the computational effort significantly. The effective Hamiltonian (16) is also partitioned by sectors. An important advantage of the method is the simultaneous calculation of a large number of states.

The FSCC equations for a particular (m,n) sector of the Fock space are derived by inserting the normal-ordered wave operator into the Bloch equation (17). The final form of the FSCC equation for a complete model space includes only *connected* terms [24],

$$Q[S_l^{(m,n)}, H_0]P = Q\{(V\Omega - \Omega H_{\text{eff}})_l^{(m,n)}\}_{\text{conn}} P, \tag{18}$$

$$H_{\text{eff}}^{(m,n)} = P(H\Omega)_{\text{conn}}^{(m,n)} P. \tag{19}$$

After converging the FSCC equation (18), the effective Hamiltonian (19) is diagonalized, yielding directly the low-lying transition energies. The effective Hamiltonian in the FSCC approach has diagonal structure with respect to the different Fock-space sectors. From (19) it follows that two Fock-space sectors belonging to a common Hilbert space, with the same net number of particles, e.g., (0,1) and (1,2), do not mix even if they have strongly interacting states. This means that important nondynamic correlation effects are only approximated. The mixed-sector CC [26] can eliminate this problem.

The FSCC equation (18) is solved iteratively, usually by the Jacobi algorithm. As in other CC approaches, denominators of the form $(E_0^P - E_0^Q)$ appear, originating in the left-hand side of the equation. The well-known intruder state problem, appearing when some Q states are close to and strongly interacting with P states, may lead to divergence of the CC iterations. To avoid this problem, a generalization of the effective Hamiltonian, the intermediate Hamiltonian (IHFSCC) approach [27, 28], has been developed in recent years. It eliminates intruder state problems caused by energy overlap of P and Q spaces, which spoils the convergence of the CC iterations, and allows thereby much larger model spaces, with enhanced applicability and accuracy. An additional advantage of the ability to use extended model spaces may be reducing the need for including high excitation levels in the formalism. High excitations (triple and higher) with large contributions are usually limited to a small group of virtual orbitals. If such orbitals are brought into P, all excitations involving them are included to infinite order by diagonalizing the effective Hamiltonian, avoiding the need for the (usually expensive) treatment of their contribution to dynamical correlation. Many of the calculations reported below resort to this approach.

Applications

Relativistic and correlation effects are known to strongly influence the structure and properties of heavy elements and to be nonadditive, as demonstrated, e.g., in our early work on the gold [29] and lanthanum [30] atoms. The reasons for these large effects are largely understood: the spatial distribution of the relativistic orbitals differs significantly from that of nonrelativistic counterparts (s and p orbitals undergo contraction, whereas d and f orbitals expand); see, however, the discussion in Ref. [31]. These changes affect not only the orbital energies; they also modify the electron correlation, thereby making relativistic and correlation corrections nonadditive. Additional complications and challenges arise from the size of the systems, as well as the often close energetic proximity of many electronic states.

Many calculations using the NVPA Fock-space CC method have been carried out over the last 20 years for heavy and superheavy atomic and molecular systems, with dozens of transition energies calculated per system. Most atomic results agreed with available experiment within a few hundredths of an eV. Molecular applications are less precise, due to their lower symmetry, which limits the molecular basis sets. Still, our calculations of heavy molecular systems, including actinide compounds, yield state-of-art benchmark molecular parameters. A more detailed description may be found in the original publications and the recent reviews [9].

The spherical symmetry of atoms, which leads to angular decomposition of the wave function and coupled cluster equations, is used at both the Dirac-Fock-Breit [23] and CC [29, 32] stages of the calculation. The energy integrals and CC amplitudes, which appear in the Goldstone-type diagrams defining the CC equations, are decomposed in terms of vector coupling coefficients, expressed by angular momentum diagrams, and reduced Coulomb-Breit or S matrix elements, respectively. The reduced equations for single- and double-excitation amplitudes are derived using the Jucys-Levinson-Vanagas theorem [24] and solved iteratively. This technique makes possible the use of large basis sets with high l values, as a basis orbital gives rise to two functions at most, with $j = l \pm 1/2$, whereas in Cartesian coordinates the number of functions increases rapidly with l. Typically we go up to h ($l = 5$) or i ($l = 6$) orbitals, but higher orbitals (up to $l = 8$) have also been used. To account for core-polarization effects, which may be important in many systems, we correlate at least the two outer shells, usually 20–50 electrons, but as many as 119 electrons were correlated for the anion of element 118 [33]. Finally, uncontracted Gaussians are used, since contraction leads to problems in satisfying kinetic balance and correctly representing the small components. On the other hand, it has been found that high-energy virtual orbitals have little effect on the transition energies calculated, since these orbitals have nodes in the inner regions of the atom and correlate mostly the inner-shell electrons, which we do not correlate anyway. These virtual orbitals, with energies above 80 or 100 hartree, are often eliminated from the CC calculation, constituting, in effect, a post-SCF contraction.

The applications presented below involve heavy and superheavy atomic systems and show the power of the relativistic FSCC approach.

Experimental-Computational Collaborative Research: The IPs of At and Lr

Collaborative determinations of the IPs of At [3] and Lr [4] have been reported recently. These are more difficult experimentally than the earlier case of Fm [2], where an isotope with half-life of 20.1 h was available vs. 7.2 s for ^{199}At and 27 s for ^{256}Lr. The IP of At was measured by resonance ionization laser ion source (RILIS) studies, giving the experimental value IP=9.317510(8) eV, compared with the MCSCF value of 9.24(15) eV and the CC result of 9.315(12) eV [3].

The experimental determination of the Lr IP is made even more difficult by the very slow production rate of the element. It was measured by a surface ionization process taking place on a tantalum surface at very high temperature (2700 and 2800 K). The Lr atoms coming out of the accelerator, one atom every few seconds, were trapped in cavities on the surface, and the ionized atoms were extracted and mass-separated. The value of the IP was determined from the measured thermoionization data using the thermodynamic partition function Q, which depends on atomic excitation energies [4]. These energies are not available experimentally; they were obtained from the 2007 IHFSCC work of Borschevsky et al. [5]. The excitation spectrum of Lu, the lighter homolog of Lr, was also calculated in that work, giving an excellent agreement with experiment, with a mean absolute error of 0.05 eV for the lowest 20 excitations; similar accuracy may be expected for Lr. To obtain such accuracy, basis sets and model spaces were augmented to convergence. The final basis included $37s31p26d21f16g11h6i$ Gaussian spinors; the P_m included $2s2p2d$ orbitals, and the full P included all the combinations of $6s5p4d2f1g$ orbitals.

The IHFSCC calculations assign the $7s^27p_{1/2}$ configuration as the ground state of Lr, in contrast to the $6s^25d_{3/2}$ ground state of Lu. This is due to the strong relativistic effects, which push the $7p_{1/2}$ orbital below the $6d$. Thus, while the excitation energy in Lu corresponding to the $5d_{3/2} \rightarrow 6p_{1/2}$ transition is 4136 cm^{-1} [34], in Lr the order of the two states is reversed, and the predicted energy of the $6d_{3/2} \rightarrow 7p_{1/2}$ transition is -1436 cm^{-1} (Table 2).

The IHFSCC calculations provide a prediction of the Lr IP (Table 2). As the experimental IP is extracted relying on transition energies obtained by the same approach, comparison between the experimental and computed IP values may raise questions. Moreover, single-reference CCSD(T) is a better scheme for the IPs in question, which involve no quasi-degeneracy, unlike the excited states. An independent calculation of the Lr IP was therefore carried out, using the relativistic single-reference coupled cluster method with single, double, and perturbative triple excitations [SRCCSD(T)]. This was the approach applied in the study of the At IP discussed above. For Lr we used a basis set of $25s23p15d14f6g3h$ Gaussians and corrected for the Breit contribution and the Lamb shift. The calculations confirmed the earlier identification of $7s^27p_{1/2}$ as the ground state of the atom and put the Lr IP at 4.963(15), in excellent agreement with the newly measured value of $4.96^{+0.08}_{-0.07}$ eV. This IP is one of the lowest in the periodic table, lower than that of Na (but higher

Table 2 Calculated IPs and excitation energies of Lu and Lr (cm^{-1}); experimental values for Lu are taken from Ref. [34]

Lu				Lr		
State		IHFSCC	Exp.	State		IHFSCC
Ionization potential						
$6s^25d$	$^2D_{3/2}$	42836	43762	$7s^27p$	$^2P_{1/2}$	39466
Excitation energies						
$6s^25d$	$^2D_{5/2}$	1945	1994	$7s^26d$	$^2D_{3/2}$	1436
$6s^26p$	$^2P_{1/2}$	4080	4136		$^2D_{5/2}$	5106
	$^2P_{3/2}$	7383	7476	$7s^27p$	$^2P_{3/2}$	8413
$6s^27s$	$^2S_{1/2}$	23730	24126	$7s^28s$	$^2S_{1/2}$	20118
$6s^28p$	$^2P_{1/2}$	30457	29430	$7s^28p$	$^2P_{1/2}$	26111
	$^2P_{3/2}$	30930	30489		$^2P_{3/2}$	27501
$6s^26d$	$^2D_{3/2}$	31929	31542	$7s^27d$	$^2D_{3/2}$	28118
	$^2D_{5/2}$	32040	31714		$^2D_{5/2}$	28385
$6s^28s$	$^2S_{1/2}$	33978	34610	$7s^29s$	$^2S_{1/2}$	30119
$6s^25f$	$^2F_{5/2}$	36595	36633	$7s^29p$	$^2P_{1/2}$	32295
	$^2F_{7/2}$	36595	36644		$^2P_{3/2}$	32840
$6s^28p$	$^2P_{1/2}$	36005	36809	$7s^26f$	$^2F_{5/2}$	32949
	$^2P_{3/2}$	36119	37131		$^2F_{7/2}$	32950

than K), confirming the large relativistic stabilization and spatial contraction of the 7s shell. The publication of this research in Nature [4] prompted a wide response, in particular regarding the structure of the periodic table, with suggestions that Lr and its lighter homolog Lu should be put in group 3, below Sc and Y. Ongoing and future progress in experimental techniques will probably lead to more collaborations of this kind.

Ground State of Rutherfordium: Interplay of Relativity and Correlation

The nature of the rutherfordium ground state has been a subject of interest for a long time. Rf is the first transactinide; in analogy with the lighter group-4 elements, it should have the ground-state configuration $[Rn]5f^{14}6d^27s^2$. Early considerations and calculations [35] suggested that the relativistic stabilization of the $7p_{1/2}$ orbital would yield a $7s^27p_{1/2}^2$ ground state, and MCSCF calculations including all possible distributions of the four external electrons in the $6d$, $7s$ and $7p$ orbitals indicated a $6d7s^27p$ ground state, with the lowest state of the $6d^27s^2$ configuration higher by 0.5 or 0.24 eV. Dynamic correlation, largely omitted from these MCSCF calculations, has been shown to play a significant role in determining atomic excitation energies, reducing the average error in FSCC calculations of Pr^{3+} excitation energies by a factor of four relative to prior MCSCF results [36].

Table 3 Excitation energies (EE) and ionization potentials (IP) in Rf^+ and Rf (eV) [37]

	Rf^+		Rf	
	$7s^2 6d_{3/2}$	$7s^2 6d_{5/2}$	$7s^2 6d^2$	$7s^2 7p6d$
	$^2D_{3/2}$ IP	$^2D_{5/2}$ EE	$^3F_2$ IP	$^3D_2$ EE
MCSCF	13.47		5.30	−0.24
MCSCF				−0.50
CCSD				
$l \leq 2$[a]	13.37	0.79	5.15	−0.60
$l \leq 3$[a]	13.95	0.82	5.65	−0.11
$l \leq 3$	14.05	0.87	5.76	0.03
$l \leq 4$	14.20	0.90	5.90	0.17
$l \leq 5$	14.34	0.92	5.99	0.25
$l \leq 6$	14.37	0.92	6.01	0.27
$l \leq 5$[b]	14.34	0.87	5.99	0.27

[a] $5f$ electrons not correlated
[b] With Breit interaction

The FSCC method was therefore applied to Rf [37]. Starting from Rf^{2+} with the closed-shell configuration $[Rn]5f^{14}7s^2$, two electrons were added, one at a time, in the $6d$ and $7p$ orbitals, to form the low-lying states of Rf^+ and Rf. A large basis set of $34s24p19d13f8g5h4i$ G-spinors was used, and the external 36 electrons were correlated, leaving only the $[Xe]4f^{14}$ core uncorrelated. A series of calculations, with increasing l values in the virtual space, was performed to assess the convergence of the results. Some of the calculated transition energies are shown in Table 3. Others may be found in the original publication [37].

The salient feature of the calculated transition energies is their monotonic behavior with the amount of correlation included. The correlation of the $5f$ electrons and the gradual inclusion of higher l virtual orbitals all increase the four transition energies in Table 3, as well as those not shown here. The MCSCF results fall invariably between the d and f limits. This makes sense, since the MCSCF function optimizes the orbitals and CI coefficients in a space including configuration state functions which correspond to all possible distributions of the four external electrons in the $6d$, $7s$, and $7p$ orbitals. Nondynamic correlation, resulting from interactions of configurations relatively close in energy, is thus described very well; in contrast, the dynamic correlation, which is more difficult to include and requires many thousands of configurations, is less well described and leads to the failure in predicting the Rf ground state. The identity of the ground state is determined by the sign of the excitation energy in the last column of Table 3. A negative energy means that the $7s^2 7p6d$ configuration is lower than $7s^2 6d^2$ and is therefore the ground state. From the calculations reported in Ref. [37], we estimate the CCSD converged value for this energy at 0.30–0.35 eV, making the $7s^2 6d^2$ state the ground state of atomic Rf. State-of-the-art experiments with Rf confirmed [38] that the chemistry of the atom is similar to that of Hf, which has a $6s^2 5d^2$ ground state.

This example shows the intricate interplay of relativity and correlation. It is well known that relativity stabilizes the p vs. d orbitals, while correlation has often

the opposite effect. When both effects are important and the result not obvious a priori, one must apply methods, such as relativistic CC, which treat relativity and correlation simultaneously to high order.

Gold and Roentgenium (E111): Local Maximum of Relativistic Effects

The gold atom exhibits very large relativistic effects on its chemical and physical properties, due to the contraction and stabilization of the $6s$ orbital. The compactness of the atom relative to its neighbors leads to a local maximum in relativistic effects, called by Pyykkö the "gold maximum." Nonrelativistic calculations lead to large errors, including the reversal of the two lowest excited states [39]. Gold was therefore selected as the first heavy atom to be treated by the CC method [29]. Two closed-shell states can be used as starting points for the Fock-space treatment, defining the (0,0) sector, namely, Au^+ or Au^-. Electrons are then added or removed according to the schemes

$$Au^+(0,0) \rightarrow Au(0,1) \rightarrow Au^-(0,2), \tag{20}$$

$$Au^-(0,0) \rightarrow Au(1,0) \rightarrow Au^+(2,0). \tag{21}$$

The basis consisted of $21s17p11d7f$ Gaussian spinors, and correlated shells included $4spdf5spd6s$. Table 4 shows the nonrelativistic, Dirac-Coulomb, and Dirac-Coulomb-Breit total energies of the two ions. As expected, relativistic effects are very large, over 1100 hartree. The nonadditivity of relativistic and correlation corrections to the energy, discussed above, is apparent in Table 4. The correlation energy of the $4spdf5spd6s$ electrons increases significantly (in absolute value), by nearly 0.1 hartree, upon going from the nonrelativistic to the relativistic Hamiltonians, due to orbital contraction.

The various transition energies of the gold atom and its ions are shown and compared with experiment [40] in Table 5. The nonrelativistic results have errors of several eV. The CC values, on the other hand, are highly accurate, with an average error of 0.06 eV. More recent (unpublished) values, obtained with a larger $31s26p21d17f14g11h8i$ basis, are even closer to experiment, giving 9.17 eV for the IP and 2.29 eV for the EA. The inclusion of the Breit interaction has a small

Table 4 Total energies of the closed-shell systems Au^+ and Au^- (hartree), with nonrelativistic, DC, and DCB Hamiltonians. Correlation includes $4s$ and higher shells

	Au^+		Au^-	
	Uncorrelated	Correlation	Uncorrelated	Correlation
NR	−17863.46301	−1.29756	−17863.68392	−1.37018
DC	−19029.01322	−1.36150	−19029.32077	−1.46436
DCB	−19007.42385	−1.36430	−19007.73063	−1.46690

Table 5 CCSD transition energies in Au (eV). IP is the ionization potential, EA denotes electron affinity, and EE is excitation energy relative to the ground state. FS denotes fine-structure splittings

			NR	DC	DCB	Expt. [40]
IP	$5d^{10}6s$	$^2S_{1/2}$	6.981	9.101	9.086	9.22
EE	$5d^96s^2$	$^2D_{5/2}$	5.301	1.115	1.150	1.136
	$5d^96s^2$	$^2D_{3/2}$	5.301	2.661	2.669	2.658
	$5d^{10}6p_{1/2}$	$^2P_{1/2}$	3.313	4.723	4.720	4.632
	$5d^{10}6p_{3/2}$	$^2P_{3/2}$	3.313	5.193	5.184	5.105
FS		^2D	0	1.546	1.519	1.522
		^2P	0	0.470	0.466	0.473
EA	$5d^{10}6s^2$	$^1S_0$	1.267	2.278	2.269	2.31

Table 6 CCSD EEs, EAs, and IPs of element 111 (eV)

Transition		DC	DCB	NR
EE	$6d^97s^2\ ^2D_{5/2} \rightarrow 6d^97s^2\ ^2D_{3/2}$	2.719	2.687	0
	$6d^97s^2\ ^2D_{5/2} \rightarrow 6d^{10}7s\ ^2S_{1/2}$	3.006	2.953	−5.430
EA	$6d^97s^2\ ^2D_{5/2} \rightarrow 6d^{10}7s^2\ ^1S_0$	1.542	1.565	6.484
IP	$6d^97s^2\ ^2D_{5/2} \rightarrow 6d^87s^2\ ^3D_4$	10.57	10.60	22.98
	$6d^97s^2\ ^2D_{5/2} \rightarrow 6d^97s\ ^3D_3$	12.36	12.33	0.92
	$6d^97s^2\ ^2D_{5/2} \rightarrow 6d^{10}\ ^1S_0$	15.30	15.23	−0.44

effect on the excitations energies, apart from some improvement of the fine-structure splittings.

A major relativistic effect in the gold atom is the stabilization of the 6*s* orbital. This is manifested by the energy separation between the $5d^{10}6s\ ^2S$ ground state and the $5d^96s^2\ ^2D$ excited state. Looking at group 11 (coinage metal) atoms, the ^2D excitation energies of Cu are 1.389 ($J = 5/2$) and 1.642 ($J = 3/2$) eV, increasing to 3.749 and 4.304 eV for Ag [40]. Were it not for relativity, one would expect even higher energies for Au. Indeed, nonrelativistic CCSD (Table 5) puts the ^2D energy at 5.301 eV above the ^2S ground state, in line with expectations. Relativistic effects reduce this value radically, giving 1.150 and 2.669 eV for the excited ^2D sublevels, within 0.015 eV of experiment [40]. Even larger stabilization may be expected for the next member of the group, element 111. The question arose whether this stabilization would be sufficient to move the $^2D_{5/2}$ level below the $^2S_{1/2}$ and to make it the ground state of the atom.

Calculations were carried out as for gold above, starting with the Rg^- anion as reference [41]. As expected, very large relativistic effects are observed, demonstrated by the large contraction of the 7*s* orbital of neutral Rg (Fig. 1). The 7*s* orbital energy of the anion goes down from −0.018 to −0.136 hartree, while the 6*d* goes up from −0.355 to −0.186 ($j = 3/2$) and −0.080 ($j = 5/2$) hartree. Atomic energies also show dramatic changes (see Table 6). Of particular interest to us is the $6d^97s^2\ ^2D_{5/2}$ state, predicted by nonrelativistic CCSD to lie 5.43 eV above the $6d^{10}7s\ ^2S$ state, but reduced relativistically to 3 eV *below* the ^2S, thus becoming

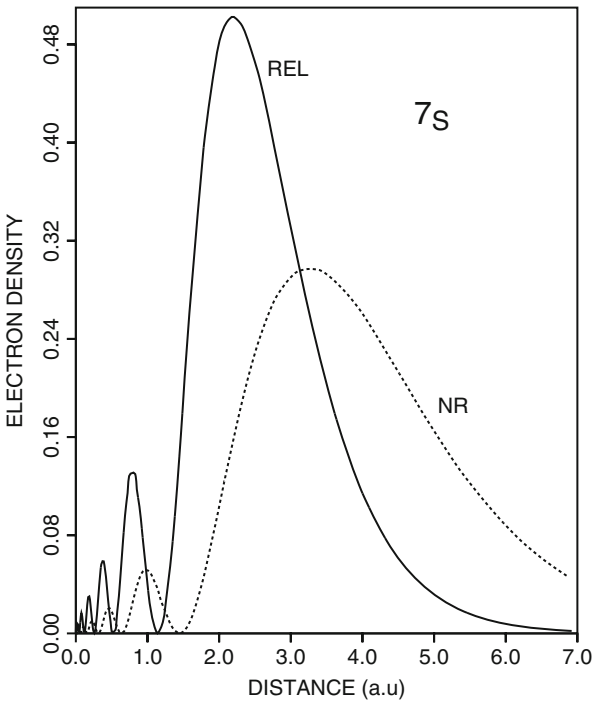


Fig. 1 Relativistic and nonrelativistic densities of Rg (element 111) 7*s* orbital

the ground state. Note that the IPs of the atom show relativistic effects of 12–15 eV! Similar effects were observed in Cn (E112) [42] and eka-thallium (E113) [43], described below.

Cn (E112) and Eka-Tl (E113): What Chemistry?

The chemistry of mercury is determined by the simple electronic structure of its atom and ions. The IPs and EEs were reproduced with high accuracy by relativistic coupled cluster calculations [42]. The ground-state configuration of the Hg atom is $[\mathrm{Xe}]4f^{14}5d^{10}6s^2$, with a relatively high 4.67 eV excitation energy to the lowest $6s6p\ \ ^3P_1$ level. The excited configurations of the mono- and dication may be described by ionization of one or two 6*s* electrons of the neutral atoms, so that no low-energy excited states exist. Application of the same method to Cn gave a very different picture [42]. The ground state of Cn^+ was found to be $6d^97s^2$, like that of the isoelectronic Rg (see above), with an excitation energy to the $6d^{10}7s$ level of only 1.6 eV, compared to 3.0 eV for Rg. The dication shows an even more complex behavior, with strong mixing of the $6d^87s^2$, $6d^97s$, and $6d^{10}$ configurations (see Table 7). The ground state is $6d^87s^2$, but two excited states with dominant $6d^97s$ contributions are found within 0.2 eV. Since both open *s* and *d*

Table 7 IPs and EEs of Cn (E112) and its ions (eV)

	State		Energy
Cn, ground state $6d^{10}7s^2\ {}^1S_0$			
IP			11.97
Cn^+, ground state $6d^97s^2\ {}^2D_{5/2}$			
IP			22.49
EE	$6d^{10}7s$	${}^2S_{1/2}$	1.60
	$6d^97s^2$	${}^2D_{3/2}$	3.14
Cn^{2+}, ground state $6d^87s^2\ J=4$			
EE	$6d^87s^2, 6d^97s$	$J=2$	0.05
	$6d^9({}^2D_{5/2})7s$	${}^3D_3$	0.19
	$6d^{10}, 6d^87s^2$	${}^1S_0$	0.79
	$6d^97s, 6d^87s^2$	$J=2$	1.39

shells can be present, the chemistry of this element may show characteristics of either main group or transition elements.

A similar picture emerges from studies of Tl and eka-Tl (E113) [43]. Here it is interesting to note the high electron affinity (0.68 eV) predicted for E113. The reliability of this value is supported by the history of the Tl EA, which was quoted at the time as 0.2±0.2 eV [44]. Our calculations gave 0.4 eV, in contrast with earlier CI [45] and MCSCF [46] values of 0.27–0.29 eV. Later experiments [47] gave 0.377(13) eV, in very good agreement with our value. The high electron affinity of E113 raises the possibility of $E113^-$ compounds. Another major difference between Tl and eka-Tl is the much reduced energies for $d^n(s+p)^m \rightarrow d^{n-1}(s+p)^{m+1}$ excitations in the latter. In particular, the $6d^97s^2$ level of $E113^{2+}$ is only ∼0.1 eV above the $6d^{10}7s$ ground state [43]; the corresponding energy in Tl^{2+} is 8.2 eV. It is therefore possible that the $E113^{2+}$ ion will form molecules resembling both Cu^{2+} ($3d^9$) and Tl^+ ($6s^2$) compounds. For further details see [43].

Fl (Element 114): Exotic Chemical Properties?

Here we compare the properties of Fl with those of lead and lighter group-4 elements. The intermediate Hamiltonian method is used to allow large P spaces and higher accuracy [48]. Two series of calculations were carried out. The first started from the closed-shell M^{4+} ion, where M stands for Pb or Fl, adding two electrons in the Fock-space scheme

$$M^{4+}(\text{sector 0}) \rightarrow M^{3+}(\text{sector 1}) \rightarrow M^{2+}(\text{sector 2}). \tag{22}$$

Thirty-two electrons $[(n-1)s^2(n-1)p^6(n-1)d^{10}(n-2)f^{14}$, with $n=6$ for Pb and 7 for Fl] are correlated in the reference state. We were interested in the few lowest states only for these sectors, so a relatively small P_m was taken, including only states with ns and np electrons added to the reference. The full P space was

considerably larger, including in addition the next four s, four p, four d, three f, and two g orbitals. In this case, the traditional FSCC converges with a model space constructed from the ns and np orbitals, and the advantage of IHFSCC lies in enhancing accuracy by allowing the larger P space.

The closed-shell M^{2+} ns^2 state is taken as reference in the second Fock-space sequence,

$$M^{2+}(\text{sector } 0) \rightarrow M^{+}(\text{sector } 1) \rightarrow M(\text{sector } 2). \tag{23}$$

Here 34 electrons are correlated in the reference state. Basis sets going to $l = 8$ were used, with $35s26p21d16f11g9h9i7k7l$ Gaussian orbitals. The FSCC iterations converge only when the np orbitals serve as the sole valence particles, i.e., only ns^2np^2 states could be obtained. The IHFSCC method allows many more valence orbitals and, consequently, many more states. The P_m for lead included all states constructed from the $7s$, $8s$, $6p$, $7p$, and $6d$ orbitals; P included, in addition, states with occupied $9s$–$12s$, $8p$–$11p$, $7d$–$9d$, $5f$–$7f$, and $5g$ orbitals. For Fl (E114) we were interested in fewer states, so that P_m was smaller, with $8s$, $7p$, and $8p$ orbitals; $9s$–$13s$, $9p$–$12p$, $7d$–$10d$, $6f$–$8f$, and $5g$ orbitals complete the P space.

The ionization potentials and low excitation energies of the lead and Fl atoms and cations are reported, together with Pb experimental values [40], in Table 8. For easy comparison, terms are listed in the LS coupling as in Moore's tables [40], although the validity of this coupling scheme for Pb, and even more so for Fl, is questionable; the only good quantum number in the reported levels (and in the calculation procedures) is J. All Pb values are within a few hundredths of an eV of experiment. This close agreement indicates that the Fl values should provide good predictions for the electronic spectrum of the superheavy element. Many more transition energies are given in the original paper [48], showing similar agreement. It is noteworthy that many Pb states can be obtained accurately by the IHFSCC method (the average error for the 22 P_m excitations is 602 cm^{-1} or 1.6 %), while only the $6s^26p^2$ states are accessible by traditional FSCC. The predicted transition energies of Fl given in the same table show IPs higher by 0.2–0.3 eV than those reported by Seth et al. [49]. Our values are probably more accurate, due to the inclusion of the Breit term and the use of a larger basis and a much extended model space, which more than compensates for the perturbative inclusion of triples. The transition energies reported here should be roughly as accurate as those of Pb, with errors of a few hundredths of an eV, and should therefore provide good predictions for future experimental values.

The most prominent feature of Table 8 is that the ionization and excitation energies of Fl are much higher than corresponding Pb values. Table 9 collects the first four ionization potentials of all group-14 elements (experimental values [50] for C–Pb, calculated for Pb and Fl), and Fig. 2 shows the first IPs of these elements. The usual trend of IPs decreasing for heavier atoms holds from C to Sn, but is reversed from Pb on, with large increases for Fl, sufficient to make its IPs larger than those of Si. This reversal is due to the relativistic stabilization of the valence s and $p_{1/2}$

Table 8 IPs and EEs of lead and Fl (eka-lead) cations and neutral species (cm^{-1}). n is 6 for Pb and 7 for Fl

	Transition			Pb		Fl
				Expt. [40]	Calc.	Calc.
M^{3+}. Ground state $(n-1)d^{10}ns\ ^2S_{1/2}$						
IP	$\rightarrow$	$(n-1)d^{10}$	$^1S_0$	341350	341748	373208
EE	$\rightarrow$	$(n-1)d^{10}np_{1/2}$	$^2P_{1/2}$	76158	76839	86676
	$\rightarrow$	$(n-1)d^{10}np_{3/2}$	$^2P_{3/2}$	97219	97803	138047
M^{2+}. Ground state $(n-1)d^{10}ns^2\ ^1S_0$						
IP	$\rightarrow$	$(n-1)d^{10}ns$	$^2S_{1/2}$	257592	257617	288256
EE	$\rightarrow$	$(n-1)d^{10}nsnp_{1/2}$	$^3P_0$	60397	60396	73135
	$\rightarrow$		$^3P_1$	64391	64451	79832
	$\rightarrow$	$(n-1)d^{10}nsnp_{3/2}$	$^3P_2$	78985	78964	119597
	$\rightarrow$		$^1P_1$	95340	95716	135170
	$\rightarrow$	$(n-1)d^{10}np_{1/2}^2$	$^3P_0$	142551	143412	166645
M^{+}. Ground state $(n-1)d^{10}ns^2np_{1/2}\ ^2P_{1/2}$						
IP	$\rightarrow$	$(n-1)d^{10}ns^2$	$^1S_0$	121243	121077	136074
EE	$\rightarrow$	$(n-1)d^{10}ns^2np_{3/2}$	$^2P_{3/2}$	14081	13885	39355
	$\rightarrow$	$(n-1)d^{10}ns^2(n+1)s$	$^2S_0$	59448	59253	71993
Neutral M. Ground state $(n-1)d^{10}ns^2np_{1/2}^2\ ^1S_0$						
IP	$\rightarrow$	$ns^2np_{1/2}$	$^2P_{1/2}$	59821	59276	68868
EE	$\rightarrow$	ns^2np^2	$^3P_1$	7819	7531	26342
	$\rightarrow$		$^3P_2$	10650	10307	28983
	$\rightarrow$		$^1D_2$	21458	20853	60956
	$\rightarrow$		$^1S_0$	29467	29259	67817
	$\rightarrow$	$ns^2np_{1/2}(n+1)s$	$^3P_0$	34960	34405	43111
	$\rightarrow$		$^3P_1$	35287	34711	43441

Table 9 Ionization potentials of group 14 elements (eV). Experimental data from the Handbook [50], calculated data: present work

	Experimental					Calculated	
	C	Si	Ge	Sn	Pb	Pb	Fl
IP1	11.260	8.152	7.900	7.344	7.417	7.484	8.539
IP2	24.383	16.346	15.935	14.632	15.032	15.012	16.871
IP3	47.888	33.493	34.224	30.503	31.937	31.941	35.739
IP4	64.494	45.142	45.713	40.735	42.322	42.372	46.272

orbitals, which becomes stronger in the superheavy elements. The higher IPs of Fl indicate it will be more inert and less metallic than lead.

More recently, extensive MCSCF calculations were carried out for the IPs discussed here [51]. The limited treatment of correlation led to substantial deviations from experimental values, between 0.4 and 1.4 eV for the four IPs of Pb, compared

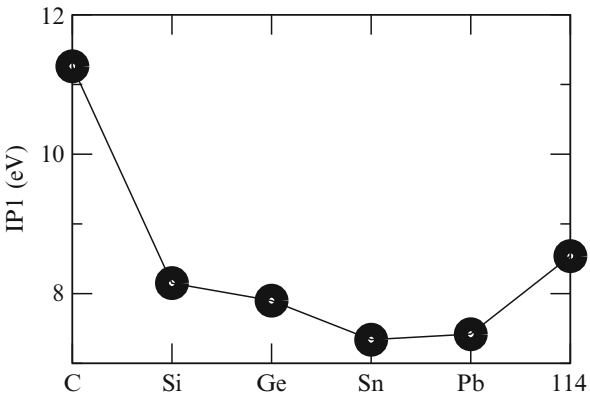


Fig. 2 First ionization potentials of group 14 elements

to the 0.02–0.07 eV shown in Table 9. The errors were extrapolated to obtain the corrected IPs of Fl. The corrected first IP was 8.28 eV, 0.26 eV below the earlier FSCC value [48], which is probably more accurate, due to the more extensive inclusion of correlation. Similar differences occur for the other IPs. Other FSCC calculations addressed the EA of Fl [52], found to be negative, and its adsorption behavior on inert surfaces [53]; the latter is important in setting up experiments studying the chemistry of this element.

Another atomic property of considerable interest is the dipole polarizability, calculated for Fl (and also for Cn) in [53]. This property was calculated at the CCSD(T) level for the group-14 elements and measured for Sn and Pb by Thierfelder et al. [54]. Good agreement with experimental values was obtained, although the latter have rather large error bounds. The calculations demonstrate the importance of relativity and electron correlation, as well as the nonadditivity of the two effects, in the case of Fl.

Element 118: Can a Rare Gas Have Positive Electron Affinity? How Important Is QED?

One of the most dramatic effects of relativity is the contraction and concomitant stabilization of *s* orbitals. An intriguing question is whether the 8*s* orbital of element 118, the next rare gas, would be stabilized sufficiently to give the atom a positive electron affinity. Using the neutral atom Dirac-Fock orbitals as a starting point raises a problem, since the 8*s* orbital has positive energy and tends to "escape" to the most diffuse basis functions. This can be avoided by calculating the unoccupied orbitals in an artificial field, obtained by assigning partial charges to some of the occupied shells. The nonphysical fields are then compensated by including an appropriate correction in the perturbation operator. A series of calculations with a variety of fields gave electron affinities differing by just a few wave numbers [55], from which an electron affinity of 0.056(10) eV was deduced. It should be emphasized that

correlated nonrelativistic or relativistic uncorrelated calculations yield no electron affinity. Another point to note is that the Rn atom does not show a bound state of the anion even at the relativistic CC level.

More recently, the issue of possible quantum electrodynamic (QED) effects on this quantity was raised. The impetus was a calculation of QED effects on the ionization potential of E119, which were estimated at 0.0173 eV [56], of the same order as the calculated EA of 118. Thus, QED effects could change the EA significantly, and their calculation has been undertaken [33]. An improved basis set with $36s32p24d22f10g7h6i$ uncontracted Gaussian-type orbitals was used, and all 119 electrons were correlated, leading to a better estimate of the electron affinity within the Dirac-Coulomb-Breit Hamiltonian, 0.064(2) eV [33]. Since the method for calculating the QED corrections [56] is based on the one-electron orbital picture, the $8s$ orbital of E118 was extracted from the correlated wave function by

$$\phi_{8s}(r) = \frac{\langle \Psi^{\mathrm{CC}}_{118} | \Psi^{\mathrm{CC}}_{118-} \rangle}{\left(\langle \Psi^{\mathrm{CC}}_{118} | \Psi^{\mathrm{CC}}_{118} \rangle \langle \Psi^{\mathrm{CC}}_{118-} | \Psi^{\mathrm{CC}}_{118-} \rangle \right)^{1/2}} \,. \tag{24}$$

The numerator of Eq. (24) is a configuration-space integral involving the product of Ψ^{*}_{118}, a 118-electron function, with the 119-electron function Ψ_{118-}; the integration is carried out over the coordinates of 118 electrons. It may be written explicitly as $\int \Psi^{\mathrm{CC}*}_{118}(1,\ldots,118)\ \Psi^{\mathrm{CC}}_{118-}(1,\ldots,119)\ \ d1\ldots d118$, with the result being a one-electron function. The normalization integrals in the denominator involve integration over all electrons, 118 for the first integral and 119 for the second. Using ϕ_{8s} and the total electron density, the self-energy and vacuum polarization terms were calculated, giving a total QED effect of 0.0059(5) eV, reducing the electron affinity by 9 % to 0.058(3) eV (for details see Ref. [33]). This is the largest relative QED effect found so far for neutral or weakly ionized species, confirming the importance of QED corrections for superheavy elements.

Beginning of the Eighth Period: Relativistic Effects Increase

Recent studies addressed the IPs, EAs, dipole polarizabilities, and other properties of elements 119 [57] and 120 [58]. As expected, stronger relativistic effects are observed. In general, properties in group-1 and group-2 elements change monotonically from row 2 to row 6 in the periodic table, due to the more diffuse s orbitals, reaching an extremum in Cs and Ba, respectively; the trend then breaks, with the properties going in the opposite direction for heavier elements. The changes observed going from row 6 to 7 increase for the eighth period elements. As an example, the dipole polarizabilities of group-1 elements are (in atomic units) 163 for Na, 291 for K, 319 for Rb, 401 for Cs, 311 for Fr, and 170 for E119, where the first four are experimental and the last two are calculated [57]. Other properties are

discussed in the original papers. These calculations made it possible to estimate the adsorption enthalpy of the two SHEs on Teflon, indicating that it is low enough to allow an easy transport of these atoms through Teflon capillaries to the chemistry set up.

Electron Affinities of E119 and Other Alkali Atoms: Accuracy at the 5 meV Level

Alkali atoms are conceptually simple one-valence-electron systems and have consequently attracted many experimental and theoretical studies. The electron affinities of the atoms up to Cs are known with great precision [44], but only a semiempirical value of 492 meV (with an uncertainty of 2.2 %) is available for Fr [59]. An interesting aspect of the alkali systems is the suggestion made some years ago [60–62] that Cs^- might have bound excited states, an idea refuted by more recent experimental [63] and theoretical [59] work. One of the first applications of the relativistic Fock-space coupled cluster (FSCC) method has been to transition energies of alkali atoms [32]. Excellent agreement with experiment was obtained for ionization potentials (an average error of 0.09 %) and excitation energies (0.2 % error), but electron affinities, particularly those of the heavier elements in the group, were less satisfactory, with errors of 4–9 % for K, Rb, and Cs. The newly developed intermediate Hamiltonian method allowed more extensive calculations with much larger P spaces, and the problem was therefore revisited [64].

The model spaces employed are believed to converge to within a few meV; they are listed in Table 10. The number of correlated electrons in the cation reference states varied from 8 in Na^+ up to 50 in $E119^+$; corresponding anions have two additional correlated electrons.

Calculated electron affinities are collected and compared with experiment in Table 11. The power of the intermediate Hamiltonian method is demonstrated by the excellent agreement with experiment. The Fock-space values start well for Na, but errors increase to 5 %, 7 %, and 9 % for the heavier K, Rb, and Cs,

Table 10 Correlated electrons in the reference cations and structure of model spaces

Atom	Correlated electrons		P_m orbitals	P_i orbitals
	Number	Orbitals		
Na	8	$2s2p$	$3–6s,3–5p,3–5d,4f$	$7–10s,6–9p,6–7d,5–7f$
K	16	$2s2p3s3p$	$4–7s,4–6p,3–5d,4f$	$8–11s,7–10p,6–7d,5–6f$
Rb	26	$3s3p3d4s4p$	$5–8s,5–7p,4–6d,4f$	$9–12s,8–11p,7–8d,5–6f$
Cs	26	$4s4p4d5s5p$	$6–9s,6–8p,5–7d,4f$	$10–13s,9–12p,8–9d,5–6f$
Fr	40	$4f5s5p5d6s6p$	$7–10s,7–9p,6–8d,5f$	$11–14s,10–13p,9–10d,6–7f$
E119	50	$4f5d6s6p6d7s7p$	$8–11s,8–10p,7–9d,5–6f$	$12–15s,11–13p,10–11d,7–8f$

Table 11 Electron affinities of the alkali atoms (meV). Fock-space (FS) and intermediate Hamiltonian (IH) compared with experimental [44] (semiempirical [59] for Fr) values

				% error	
Atom	Expt.	FS	IH	FS	IH
Na	547.9	549.1	549.9	0.3	0.4
K	501.5	525.4	506.8	4.8	1.1
Rb	485.9	519.3	490.8	6.9	1.0
Cs	471.6	516.0	474.6	9.4	0.6
Fr	492±2.2 %	542.2	491.3		
E119		717.1	662.5		

respectively. IHFSCC values, on the other hand, are all within 5 meV or 1 % of experiment. It should be noted that the FSCC function of the anions includes only one determinant in the P space, whereas the IHFSCC P space includes several thousand. The importance of these additional determinants (and of excitations from them to Q) increases with the size of the alkali anion.

The excellent agreement with known experimental values enables us to propose with confidence a reference value for the yet unmeasured electron affinity of Fr, 491±5 meV. This is in agreement with the recent semiempirical value [59] of 492±10 meV. It is interesting to note the increase of the electron affinity for the heaviest alkali atoms, due to relativistic stabilization of the s orbitals. While the EAs go down from Na to Cs, Fr exhibits a modest increase, and the superheavy element 119 is predicted to have a much larger affinity, larger in fact than any other alkali atom. This dramatic stabilization effect on the 8s orbital has been observed for element 118, which is predicted to be the first rare gas with positive electron affinity (see above).

Eka-actinium (E121) vs. Its Homologs La and Ac: When Is the Breit Term Important?

Element 121 and its lighter homologs were studied by an early application of the FSCC approach [30]. The IPs and EEs of the two light atoms are all within a few hundredths of an eV of experiment (when available). The calculations for E121, shown in Table 12, are expected to have similar accuracy. The interesting feature exhibited by the latter is the relativistic stabilization of the $8p_{1/2}$ orbital. Table 12 compares Ac and E121 levels, showing that the ground state of the latter, $8s^2 8p_{1/2}$, differs from the $ns^2(n-1)d_{3/2}$ ground states of La (n = 6), Ac (n = 7), and the lighter Sc (n=4) and Y (n = 5). Another finding involves the effect of the Breit term on transition energies. It is rather small (0.01–0.02 eV) for most transitions, but goes up to 0.1 eV for transitions involving f electrons. Similar behavior was found in other heavy and superheavy atoms.

Table 12 Ionization potential (IP) and excitation energies (EE) of elements Ac and eka-Ac (E121) and their cations (eV) by the Dirac-Coulomb (DC) and Dirac-Coulomb-Breit (DCB) FSCCSD method

	Final state		DC	DCB	Exp.[a]		Final state		DC	DCB
Ac^{2+}, ground state $7s\ ^2S_{1/2}$						$E121^{2+}$, ground state $8s\ ^2S_{1/2}$				
IP	[Rn]	$^1S_0$	17.518	17.512		IP	[E118]	$^1S_0$	18.67	18.65
EE	$6d$	$^2D_{3/2}$	0.157	0.145	0.099	EE	$7d$	$^2D_{3/2}$	2.883	2.859
		$^2D_{5/2}$	0.588	0.569	0.521			$^2D_{5/2}$	3.507	3.478
							$8p$	$^2P_{1/2}$	4.186	4.196
								$^2P_{3/2}$	6.831	6.819
							$6f$	$^2F_{5/2}$	4.557	4.471
								$^2F_{7/2}$	5.050	4.960
Ac^{+}, ground state $7s^2\ ^1S_0$						$E121^{+}$, ground state $8s^2\ ^1S_0$				
IP	$7s$	$^2S_{1/2}$	11.91	11.90	12.1	IP	$8s$	$^2S_{1/2}$	12.67	12.66
EE	$6d\,7s$	$^3D_1$	0.634	0.623	0.588	EE	$7d\,8s$	$^3D_1$	2.764	2.744
		$^3D_2$	0.702	0.690	0.653			$^3D_2$	2.850	2.829
		$^3D_3$	0.976	0.960	0.921			$^3D_3$	3.209	3.185
		$^1D_2$	1.194	1.176	1.127			$^1D_2$	3.800	3.775
							$8s\,8p$	$^3P_0$	3.001	3.010
								$^3P_1$	3.331	3.338
								$^3P_2$	5.293	5.284
							$5f\,8s$	$^3F_2$	5.149	5.082
								$^3F_3$	5.314	5.238
								$^3F_4$	5.635	5.558
Ac, ground state $7s^2 6d_{3/2}\ ^2D_{3/2}$						E121, ground state $8s^2 8p_{1/2}\ ^2P_{1/2}$				
IP	$7s^2$	$^1S_0$	5.31	5.32	5.17[b]	IP	$8s^2$	$^1S_0$	4.458	4.447
EE	$7s^2 6d_{5/2}$	$^2D_{5/2}$	0.274	0.290	0.277	EE	$8s^2 7d_{3/2}$	$^2D_{3/2}$	0.412	0.389
	$7s^2 7p_{1/2}$	$^2P_{1/2}$	0.969	0.984			$8s^2 7d_{5/2}$	$^2D_{5/2}$	0.738	0.714
	$7s^2 7p_{3/2}$	$^2P_{3/2}$	1.573	1.583			$8s^2 8p_{3/2}$	$^2P_{3/2}$	1.436	1.424

[a]Ref. [40]
[b]Ref. [50]

Thorium and Eka-thorium (E122): Different Level Structure

The discussion of Rg above describes an example where the ground-state configuration of a superheavy element differs from that of the lighter atoms in the group; another example of a superheavy element not following its lighter homologs is given by E122.

The Fock-space and intermediate Hamiltonian coupled cluster methods were applied to the ground and excited levels of the second actinide element, thorium, and its heavy homolog, eka-thorium (E122) [65]. The two Fock-space schemes described above, starting with the M^{4+} ion [Eq. (22)] and M^{2+} ion [Eq. (23)], are

Table 13 Valence orbitals in Fock-space and intermediate Hamiltonian calculations

	Th	E122
	Scheme (22)	
FS[a]	$7s, 6d, 5f$	$8s, 7p, 7d, 6f$
IH: P_m	7–8s,6–7p,6–7d,5–6f	8–9s,7–8p,7–8d,6–7f
IH: P	7–12s,6–11p,6–10d,5–9f,5–7g, 6h, 7i	8–14s,7–12p,7–11d,6–10f,5–7g, 6h, 7i
	Scheme (23)	
FS[a]	$6p, 6d, 5f$	$7p, 7d, 6f$
IH: P_m	$6p, 6d, 5f$	$9s, 7p, 7d, 6f$
IH: P	8–13s,6–10p,6–10d,5–8f,5–6g, 6h, 7i	9–13s,7–11p,7–11d,6–9f,5–6g, 6h

[a]Some determinants had to be moved from P to Q to achieve convergence of the FS calculations. These include all the p^2, pd, and pf determinants in scheme (22) for E122 and the p^2, f^2, and pf determinants in scheme (23) for both elements

used; here M denotes the Ac or E121 atom. It should be noted that the experimental ground state of Th^{2+} is $6d^2$ rather than $7s^2$ [66]; this is also the ground state determined by scheme (22). The closed-shell $7s^2$ state is, however, used as reference in the sequence (23). The calculated ground state of $E122^{2+}$ is $8s^2$.

The structure of the model space P in the Fock-space method and of P_m and P in the intermediate Hamiltonian approach is shown in Table 13. All determinants constructed from the orbitals listed in the table constitute the relevant space. P_m is a subspace of P in the IHFSCC approach. Convergence difficulties of the FSCC formalism in sector 2 made it necessary to use an incomplete model space [67, 68], moving certain determinants from P to Q. The IH calculations employ much larger P spaces, which are always complete (i.e., include *all* combinations of relevant orbitals). Orbital selection was determined primarily on the basis of orbital energies.

The ionization potentials and lower excitation energies of Th and its monocation are reported in Table 14. Many more energies, including those of the di- and trication, were shown in the original publication [65]. Very good agreement with experiment [66] is obtained: the mean absolute error of the 51 Fock-space energies at all ionization levels is 0.062 eV. The intermediate Hamiltonian approach reduces it to 0.051 eV. This level of accuracy is obtained in spite of the complicated interactions between different electronic configurations, which lead to a rather dense spectrum.

The ionization potentials and low excitation energies calculated for E122 are shown in Table 15. More values may be found in [65]. Intermediate Hamiltonian values for E122 and its monocation were calculated by the Dirac-Coulomb and Dirac-Coulomb-Breit schemes in order to explore the effect of the Breit interaction (2). Overall, the Breit term contribution is small (0.01–0.04 eV) for transitions not involving f electrons, but increases to 0.07–0.1 eV when f orbitals are involved in the excitations, as observed above (section "Eka-actinium (E121) vs. Its Homologs La and Ac: When Is the Breit Term Important?"). The ground state is predicted to be $8s^2 8p 7d$, in agreement with early

Table 14 Fock-space and intermediate Hamiltonian transition energies of thorium compared with experiment [66] (eV)

Transition		expt.	FS	IH
Th. Ground state $6d^2 7s^2\ {}^3F_2$				
$\rightarrow \mathrm{Th}^+\ 6d7s^2$	$^2D_{3/2}$[a]	6.537	6.497	6.521
$\rightarrow\ 6d^2 7s^2$	$^3P_0$	0.317	0.486	0.450
	$^3F_3$	0.355	0.345	0.369
	$^3P_2$	0.457	0.592	0.538
	$^3P_1$	0.479	0.680	0.669
	$^3F_4$	0.615	0.623	0.624
	$^1D_2$	0.902	1.062	1.008
$\rightarrow\ 5f6d7s^2$	$^3H_4$	0.966	1.048	1.062
$\rightarrow\ 6d^2 7s^2$	$^1G_4$	1.005	1.155	1.070
$\rightarrow\ 5f6d7s^2$	$^3F_2$	1.021	1.079	1.087
	$^1G_4$	1.290	1.361	1.371
	$^3G_3$	1.304	1.464	1.462
Th^+. Reference state $6d^2 7s\ {}^2D_{3/2}$[a]				
$\rightarrow\ 5f_{5/2}7s^2$	$^2F_{5/2}$	0.326	0.217	0.234
$\rightarrow\ 6d_{5/2}7s^2$	$^2D_{5/2}$		0.512	0.510
$\rightarrow\ 5f_{7/2}7s^2$	$^2F_{7/2}$	0.808	0.713	0.734

[a]The ground state of Th^+ is $6d^2 7s\ {}^4F_{3/2}$, which cannot be reached in the Fock-space scheme used here. The IP of Th and excitations of Th^+ are therefore shown for the $6d7s^2\ {}^2D_{3/2}$ state, which lies 0.23 eV above the ground state

Table 15 Fock-space and intermediate Hamiltonian transition energies of eka-thorium (element 122) (eV)

		FS	IH	
Config.	J	DCB	DC	DCB
E122. Ground state $8s^2 7d_{3/2}8p_{1/2}\ J=2$				
$\rightarrow \mathrm{E122}^+\ 7d_{3/2}$	3/2	5.651	5.613	5.595
$\rightarrow\ 8p_{1/2}^2$	0	0.185	0.157	0.162
$\rightarrow\ 7d_{3/2}^2$	2	0.348	0.385	0.353
$\rightarrow\ 7d_{3/2}8p_{1/2}$	1	0.662	0.651	0.636
$\rightarrow\ 7d_{3/2}7d_{5/2}$	3	0.860	0.891	0.856
$\rightarrow\ 7d_{5/2}8p_{1/2}$	2	0.875	0.872	0.862
$\rightarrow\ 7d_{5/2}8p_{1/2}$	3	0.955	0.954	0.940
$\rightarrow\ 7d_{3/2}7d_{5/2}$	4	1.012	1.028	0.988
	2	1.036	1.030	0.996
	1	1.154	1.144	1.113
$\mathrm{E122}^+$. Ground state $8s^2 7d_{3/2}$				
$\rightarrow \mathrm{E122}^{2+}\ 8s^2$	0	11.332	11.288	11.301
$\rightarrow\ 8s^2 6f_{5/2}$	5/2	0.262	0.342	0.261
$\rightarrow\ 8s^2 7d_{5/2}$	5/2	0.663	0.658	0.653
$\rightarrow\ 8s^2 8p_{1/2}$	1/2	0.696	0.644	0.681
$\rightarrow\ 8s^2 6f_{7/2}$	7/2	0.967	1.059	0.970

Dirac-Fock(-Slater) calculations [69, 70] but in contrast to the $8s^2 8p^2$ configuration obtained by density functional theory [71]. The separation of the levels is not large: the lowest $8p^2$ and $7d^2$ states appear just 0.16 and 0.35 eV, respectively, above the ground state. It should be noted that the ground state of thorium is $7s^2 6d^2$, and the increased relativistic stabilization of the p orbital of E122 changes the relative energy of the configurations. The first excited state of E122 is $8p^2$, whereas the corresponding Th level is quite high. Similar phenomena occur in the ions. The ground state of Th^+ is $6d^2 7s$, while the lowest level of $E122^+$ is predicted to be $7d\,8s^2$. Th^{2+} has a $6d5f\ ^3H_4$ ground state, with $6d^2\ ^3F_2$ less than 0.01 eV away; the accuracy of the current method is not sufficient to decide between these two states. The lowest level of $E122^{2+}$ is a closed-shell $8s^2$, with all the low excited states (up to 4 eV) having an $8snl$ configuration.

Quantum electrodynamic effects were not included here. Calculations of these effects for s electrons [56] gave estimates of about 0.04 eV for the ionization potential of the Tl^{2+} $6s$ electron (self-energy 0.05 eV, vacuum polarization −0.01 eV), 0.06 eV for the E111 $7s$ electron (self-energy 0.09 eV, vacuum polarization −0.03 eV), and similar shifts of 0.02–0.04 eV were calculated for the low-lying energies of atomic No [72]. Since the p electrons responsible for the transitions discussed in this work exhibit much weaker penetration into the nucleus, QED effects here are expected to be considerably smaller, at most 0.01–0.02 eV, within the error limits of the method ($\sim$ 0.05 eV) estimated by comparing calculated and experimental results for thorium (Table 14).

Summary

The no-virtual-pair Dirac-Coulomb-Breit Hamiltonian, correct to second order in the fine-structure constant α, provides the framework for four-component methods, the most accurate approximation applied in electronic structure calculations of heavy atomic and molecular systems, including SHEs. Electron correlation is taken into account by the powerful coupled cluster approach. This method is particularly suitable to SHEs, where the high density of states necessitates simultaneous treatment of large manifolds of levels. The intermediate Hamiltonian variant of FSCC is instrumental in avoiding intruder states, which destroy the convergence of the CC iterations, thereby allowing the use of extensive and converged model spaces. As demonstrated above, these methods are applicable to a variety of SHEs. Their accuracy was assessed by carrying out equivalent calculations for lighter homologs, up to 6th period and early 7th period elements, and by comparing these results with known experimental values. It was found that the atomic transition energies, including excitation energies and ionization potentials, are usually reproduced within a few hundredths of an eV; similar precision is expected for SHEs. Both the basis sets and model spaces must be carefully extended to convergence to achieve this benchmark.

The main advantages of IHFSCC include the simultaneous economical determination of large numbers of energy levels and the treatment of both dynamic and nondynamic electron correlation to high order. Its main shortcoming is the limitation to states obtained from a closed-shell configuration by adding and/or removing two electrons at most. Efforts have been made [25] to extend the method to sectors of the Fock space with a larger number of electrons or holes. Other work in progress includes the further development of the Hilbert-space and mixed-sector IHCC [26], as well as the double FSCC formalism mentioned in the introduction, which will include higher QED terms and allow a more precise treatment of SHEs, especially in the case of highly ionized species.

The spectroscopic and chemical behavior of the heaviest elements is often different from that expected by extrapolation of their lighter homologs. Perhaps the most interesting deviations from lighter element behavior involve changes in the ground-state electron configuration, which largely determines the chemistry. Examples include but are not limited to the following:

- Rg (E111) is predicted to have a $6d^9 7s^2$ ground state, unlike the (n-1)d^{10}ns^1 configurations of the lighter Cu, Ag, and Au.
- The ground state of element 121 will be $8s^2 8p^1_{1/2}$, compared with the $7s^2 6d^1_{3/2}$ ground configuration of the homolog Ac.
- Similarly, the ground state of element 122 will be $7d_{3/2} 8s^2 8p_{1/2}$ vs. the $6d^2_{3/2} 7s^2$ of Th.

These changes are caused by the relativistic stabilization of s and $p_{1/2}$ orbitals, which is also a major factor in other findings for the heavy elements:

- The IP of Lr, the last actinide, is one of the lowest in the periodic table, lower than that of Na (but higher than K).
- Element 118, the next rare gas, is predicted to bind an electron, with a positive electron affinity of 0.06 eV. An interesting finding is the 9 % reduction of the EA calculated within the DCB Hamiltonian by the Lamb shift, the largest relative QED effect shown to date in atomic transition energies.
- Cn (E112) chemistry will be affected by the increased HOMO-LUMO gap between the fully occupied 8s and unoccupied 8p and the larger IP compared to Hg, leading probably to a more inert element. This also applies to the cation of element 113, making the expected chemistry of this element different from that of Tl.
- The trends in atomic properties exhibited by the 7th-period atoms, which are often opposite to trends in the lighter elements, are even stronger in 8th-period atoms, as seen in calculations on elements 119 and 120.

Finally, it is important to treat relativity and electron correlation simultaneously and with similar precision. This is demonstrated by the ground state of Rf, where one candidate electron configuration is favored by relativity and another by correlation,

and only the careful simultaneous incorporation of both contributions leads to a decisive answer.

While experimental information on the chemistry of these elements is becoming available, this is not the case for spectroscopic data. Nevertheless, there is considerable progress. As shown above, such information has been obtained for Fm and Lr; the latter is the heaviest actinide, so we may expect exciting developments for transactinides in the near future.

References

1. Schädel M (2015) Philos Trans R Soc A 373:20140191
2. Sewtz M, Backe H, Dretzke A, Kube G, Lauth W, Schwamb P, Eberhardt K, Grüning C, Thörle P, Trautmann N, Kunz P, Lassen J, Passler G, Dong CZ, Fritzsche S, Haire RG (2003) Phys Rev Lett 90:163002
3. Rothe S, Andreyev AN, Antalic S, Borschevsky A, Capponi L, Cocolios TE, De Witte H, Eliav E, Fedorov DV, Fedosseev VN, Fink DA, Fritsche S, Ghys L, Huyse M, Imai N, Kaldor U, Kudryavtsev Yu, Köster U, Lane J, Lassen J, Liberati V, Lynch KM, Marsh BA, Nishio K, Pauwels D, Pershina V, Popescu L, Procter TJ, Radulov D, Raeder S, Rajabali MM, Rapisarda E, Rossel RE, Sandhu K, Seliverstov MD, Sjödin AM, van der Berg P, Van Duppen P, Venhart M, Wakabayashi Y, Wendt KDA (2013) Nat Commun 4:1835
4. Sato TK, Asai M, Borschevsky A, Stora T, Sato N, Kaneya Y, Tsukada K, Düllmann ChE, Eberhardt K, Eliav E, Ichikawa S, Kaldor U, Kratz JV, Miyashita S, Nagame Y, Ooe K, Osa A, Renisch D, Runke J, Schädel M, Thörle-Pospiech P, Toyoshima A, Trautmann N (2015) Nature 520:209
5. Borschevsky A, Eliav E, Vilkas MJ, Ishikawa Y, Kaldor U (2007) Eur Phys J D 45:115
6. Morss LR, Edelstein NM, Fuger J (eds) (2010) The chemistry of the actinide and transactinide elements, 4th edn. Springer, Dordrecht
7. Türler A, Pershina V (2013) Chem Rev 113:1237; Schädel M (2006) Angew Chem Int Ed 45:368; Schädel M (2012) Radiochim Acta 100:579; Radiochim Acta 99 (2011), special volume
8. Pyykkö P (2013) J Comput Chem 34:2667. http://rtam.csc.fi
9. Eliav E, Kaldor U (2010) In: Barysz M, Ishikawa Y (eds) Relativistic methods for chemists. Springer, London, pp 279ff.; also In: Pittner J, Charsky P, Paldus J (eds) Recent progress in coupled cluster methods: theory and applications. Springer, Dordrecht/Heidelberg/London/New York, pp. 113ff., 2010
10. Sucher J (1989) In: Johnson W, Mohr P, Sucher J (eds) Relativistic, quantum electrodynamic, and weak interaction effects in atoms. American Institute of Physics, New York, pp 28–57
11. Sucher J (1980) Phys Rev A 22:348; Phys Scr 36:271 (1987)
12. Lindgren I (1989) In: Kaldor U (ed) Many-body methods in quantum chemistry. Lecture notes in chemistry, vol 52. Springer, Heidelberg, pp 293–306; Nucl Instr Meth B 31:102 (1988)
13. Pyykkö P (2012) Chem Rev 112:371
14. Mittleman M (1971) Phys Rev A 4:893; Phys Rev A 6:2395–2401 (1972); Phys Rev A 24:1167 (1981)
15. Visscher L, Dyall KG (1997) At Data Nucl Data Tables 67:207
16. Kaldor U, Eliav E (1998) Adv Quantum Chem 31:313
17. Stanton RE, Havriliak S (1984) J Chem Phys 81:1910
18. Ishikawa Y, Binning RC, Sando KM (1983) Chem Phys Lett 101:111; Chem Phys Lett **105**, 189 (1984); Chem Phys Lett 117:444 (1985); Ishikawa Y, Baretty R, Binning RC (1985) Chem Phys Lett 121:130; Ishikawa Y, Quiney HM (1987) Intern J Quantum Chem Symp 21:523

19. Ilyabaev E (Eliav), Kaldor U, Chem Phys Lett 194:95 (1992); Phys Rev A 47:137 (1993); Eliav E, Kaldor U (1996) Chem Phys Lett 248:405; Eliav E, Kaldor U, HeßBA (1998) J Chem Phys 108:3409
20. Pal S, Mukherjee D (1989) Adv Quantum Chem 20:292
21. Kaldor U (1991) Theor Chim Acta 80:427
22. Kaldor U, Eliav E (1998) Adv Quantum Chem 31:313
23. Ishikawa Y, Binning RC, Sekino H (1989) Chem Phys Lett 160:206; Ishikawa Y (1990) Phys Rev A 42:1142; Chem Phys Lett 166:321 (1990); Ishikawa Y, Quiney HM (1993) Phys Rev A 47:1732; Ishikawa Y (1994) Koc K, Phys Rev A 50:4733
24. Salomonson S, Lindgren I, Mårtensson A-M (1980) Phys Scr 21:351; Lindgren I, Morrison J (1986) Atomic Many-Body Theory, 2nd edn. Springer, Berlin
25. Hughes SR, Kaldor U (1992) Chem Phys Lett 194:99; Chem Phys Lett 204:339 (1993); Phys Rev A 47:4705 (1993); J Chem Phys 99:6773 (1993); Int J Quantum Chem 55:127 (1995)
26. Landau A, Eliav E, Ishikawa Y, Kaldor U (2004) J Chem Phys 121:6634
27. Landau A, Eliav E, Kaldor U (1999) Chem Phys Lett 313:399; Landau A, Eliav E, Ishikawa Y, Kaldor U (2000) J Chem. Phys. 113:9905; Landau A, Eliav E, Kaldor U (2001) Adv Quantum Chem 39:172
28. Eliav E, Vilkas MJ, Ishikawa Y, Kaldor U (2005) J Chem Phys 122:224113
29. Eliav E, Kaldor U, Ishikawa Y (1994) Phys Rev A 49:1724
30. Eliav E, Shmulyian S, Kaldor U, Ishikawa Y (1998) J Chem Phys 109:3954
31. Autschbach J, Siekierski S, Seth M, Schwerdtfeger P, Schwarz WHE (2002) J Comput Chem 23:804
32. Eliav E, Kaldor U, Ishikawa Y (1994) Phys Rev A 50:1121
33. Goidenko IA, Labzowsky L, Eliav E, Kaldor U (2003) Pyykkö P, Phys Rev A 67:020101
34. Martin WC, Zabulas R, Hagan L (1978) Atomic energy levels – the rare earths elements. National standard reference data series, National Bureau of Standards, Washington, DC
35. Keller OL (1984) Radiochim Acta 37:169; Mann JB, quoted by Fricke B, Waber JT (1971) Actinides Rev 1:433; Glebov VA, Kasztura L, Nefedov VS, B.L Zhuikov (1989) Radiochim Acta 46:117; Johnson E, Fricke B, Keller OL, Nestor CW Jr, Tucker TC (1990) J Chem Phys 93:8041
36. Eliav E, Kaldor U, Ishikawa Y (1995) Phys Rev A 51:225
37. Eliav E, Kaldor U, Ishikawa Y (1995) Phys Rev Lett 74:1079
38. Schädel M (1995) Radiochim Acta 70/71:207
39. Hay PJ, Wadt WR, Kahn LR, Bobrowicz FW (1978) J Chem Phys 69:984; Pizlo A, Jansen G, Hess BA (1993) J Chem Phys 98:3945
40. Moore CE (1952, 1958) Atomic Energy Levels, National Bureau of Standards (U.S.), Circ No. 467, vol ii (1952), vol iii (1958). U.S. GPO, Washington, DC
41. Eliav E, Kaldor U, Schwerdtfeger P, Hess BA, Ishikawa Y (1994) Phys Rev Lett 73:3203
42. Eliav E, Kaldor U, Ishikawa Y (1995) Phys Rev A 52:2765
43. Eliav E, Kaldor U, Ishikawa Y, Seth M, Pyykkö P (1996) Phys Rev A 53:3926
44. Hotop H, Lineberger WC (1985) J Chem Phys Ref Data 4 539 (1975); ibid, 14:731
45. Arnau F, Mota F, Novoa JJ (1992) Chem Phys 166:77
46. Wijesundera WP (1997) Phys Rev A 55:1785
47. Carpenter DL, Covington AM, Thompson JS (2000) Phys Rev A 61:042501
48. Landau A, Eliav E, Ishikawa Y, Kaldor U (2001) J Chem Phys 114:2977
49. Seth M, Fægri K, Schwerdtfeger P (1998) Angew Chem Intl Ed 37:2493
50. Lide DR (ed) (1993) Handbook of Chemistry and Physics, 74th edn. CRC Press, Boca Raton
51. Yu YJ, Dong CZ, Li JG, Fricke B (2008) J Chem Phys 128:124316
52. Borschevsky A, Pershina V, Eliav E, Kaldor U (2009) Chem Phys Lett 480:51
53. Pershina V, Borschevsky A, Eliav E, Kaldor U (2008) J Chem Phys 128:024707
54. Thierfelder C, Assadollahzadeh B, Schwerdtfeger P, Schäfer S (2008) Phys Rev A 78:052506
55. Eliav E, Kaldor U, Ishikawa Y, Pyykkö P (1996) Phys Rev Lett 77:5350
56. Pyykkö P, Tokman M, Labzowsky L (1998) Phys Rev A 57:R689–R692; Labzowsky L, Goidenko I, Tokman M, Pyykkö P (1999) Phys Rev A 59:2707

57. Borschevsky A, Pershina V, Eliav E, Kaldor U (2013) J Chem Phys 138:124302
58. Borschevsky A, Pershina V, Eliav E, Kaldor U (2013) Phys Rev A 87:022502
59. Bahrim C, Thumm U (2000) Phys Rev A 61:022722
60. Fabrikant II (1982) Opt Spektrosk 53:223
61. Froese Fischer C, Chen D (1989) J Mol Struct 199:61
62. Greene CH (1990) Phys Rev A 42:1405
63. Scheer M, Thogersen J, Bilodeau RC, Brodie CA, Haugen HK, Andersen HH, Kristensen P, Andersen T (1998) Phys Rev Lett 80:684
64. Landau A, Eliav E, Ishikawa Y, Kaldor U (2001) J Chem Phys 115:2389
65. Eliav E, Landau A, Ishikawa Y, Kaldor U (2002) J Phys B 35:1693
66. Kramida A, Ralchenko Yu, Reader J, NIST ASD Team (2013). NIST Atomic Spectra Database (version 5.1). http://physics.nist.gov/asd [Wednesday, 20-Aug-2014 05:45:32 EDT]. National Institute of Standards and Technology, Gaithersburg, MD; VizieR astronomical database. http://vizier.u-strasbg.fr
67. Hose G, Kaldor U (1979) J Phys B 12:3827; Phys Scr 21:357 (1980)
68. Meissner L, Kucharski SA, Bartlett RJ (1989) J Chem Phys 91:6187; Meissner L, Bartlett RJ (1990) J Chem Phys 92:561
69. Fricke B, Greiner W, Waber JT (1971) Theor Chim Acta 21:231; Fricke B, McMinn J (1976) Naturwiss 63:162
70. Mann JB, Waber JT (1970) J Chem Phys 53:2397
71. Umemoto K, Saito S (1996) J Phys Soc Japan 65:3175
72. Fritzsche S (2005) Eur Phys J D 33:15

Relativistic Quantum Chemistry for Chemical Identification of the Superheavy Elements

27

Valeria Pershina

Contents

Introduction . . . 858
Relativistic and QED Effects on SHE . . . 860
Chemical Experiments . . . 862
Relativistic Quantum Chemical Methods and Approaches . . . 864
Atomic Properties of SHE and the Structure of the Periodic Table . . . 868
Gas-Phase Chemistry . . . 872
Aqueous Chemistry . . . 891
Summary . . . 893
References . . . 893

Abstract

Production and investigation of properties of superheavy elements (SHEs) belong to the most fundamental areas of physical science. They seek to probe the uppermost reaches of the periodic table of the elements where the nuclei are extremely unstable and relativistic effects on the electron shells are increasingly strong. Theoretical chemical research in this area is very important. Due to experimental restrictions, it is often the only source of useful chemical information. It enables one to predict the behavior of the heaviest elements in the sophisticated and demanding experiments with single atoms and to interpret their results. Spectacular developments in the relativistic quantum theory and computational algorithms in the last few decades allowed for accurate calculations of electronic structures and properties of SHE and their compounds. Results of those investigations, particularly those related to the experimental research, are overviewed in this chapter. The role of relativistic effects is elucidated.

V. Pershina (✉)
GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany
e-mail: V.Pershina@gsi.de

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6_35

Keywords
Predictions of experimental behavior • Quantum chemical calculations • Relativistic effects • Superheavy elements

Introduction

Superheavy elements (SHEs) are elements with $Z \geq 104$, also called transactinides. They are located in the periodic table after the actinide series ending with element 103, Lr. Nowadays, SHEs from $Z = 104$ through 118 are all known, and most of them have been named (Fig. 1) [1–5].

The SHEs are all man-made. The first members of the transactinides series, $Z = 104$, 105, and 106, were discovered in 1969 through 1974 in heavy-ion accelerators by bombardment of the heavy actinide (Cf) targets with light ions (C and O), so-called "hot-fusion" reactions. In the 1970s, a different type of fusion reactions was found and later used in the production of elements with Z from 106 through 113. These so-called "cold-fusion" reactions were based on targets in the vicinity of doubly magic ^{208}Pb (mainly Pb and Bi) and beams of the complementary medium-mass projectiles with $Z \geq 24$. The lifetime of the produced elements proved to be very short, for example, the half-life of ^{277}Cn is only 0.6 ms. The cross section was also found to decrease rapidly with increasing Z. It is, for example, only ~0.5 pb for ^{277}Cn [2]. It was, therefore, concluded that it would be very difficult to reach even heavier elements in this way. Thus, SHEs from $Z = 112$ through 118 were produced using "hot-fusion" reactions between ^{48}Ca ions and

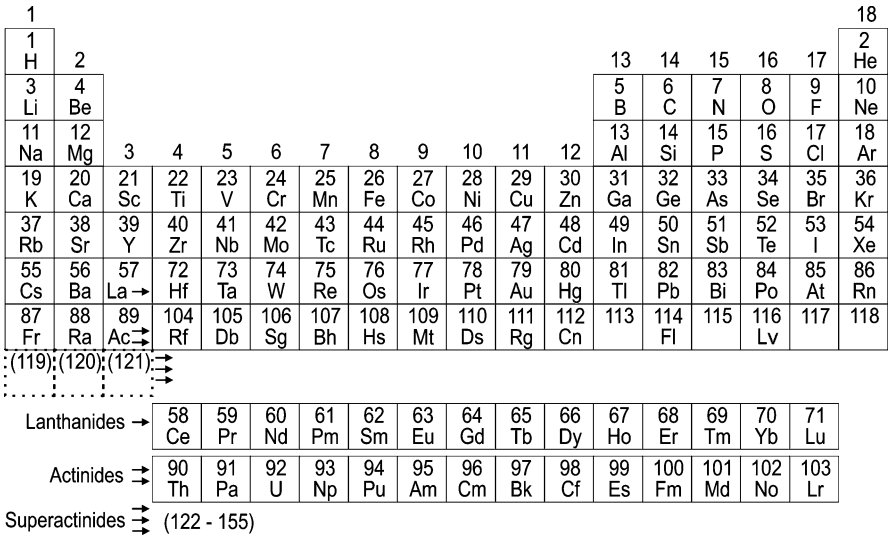


Fig. 1 Modern periodic table of the elements (Reproduced with permission from Ref. [5]. Copyright 2013 American Chemical Society)

^{238}U, ^{242,244}Pu, ^{243}Am, ^{245,248}Cm, ^{249}Bk, and ^{249}Cf targets [3]. These discoveries are of considerable interest for chemical studies, because the reported half-lives are much longer (many orders of magnitude) than those of the isotopes produced by "cold-fusion" reactions which lead to more neutron-deficient isotopes, e.g., $t_{1/2}(^{283}\mathrm{Cn}) = 3.8\,\mathrm{s}$ and $t_{1/2}(^{289}\mathrm{Fl}) = 2.1\,\mathrm{s}$.

For the positive identification of a new element and its placement in a proper position in the periodic table, its atomic number, Z, must be determined or deduced in some way. For the transactinides up to $Z = 113$, their atomic number has been identified first by "physical" techniques. One widely used technique is that of $\alpha\alpha$ – correlation of the element's α-decay to a known daughter and/or granddaughter nucleus. Positive identification becomes more difficult for species that decay predominantly by spontaneous fission. Also, neutron-rich isotopes of elements beyond $Z = 113$ decay into unknown products, so that their Z cannot be directly established. One of the indirect ways of determining Z in that area is by measuring X-ray spectra along the scheme established for element 115 [4].

But even when the atomic number can be positively assigned by α-decay chains, no knowledge is obtained about electronic configurations or chemical properties of these new elements from these physical methods. The elements are just placed in the periodic table in corresponding chemical groups or periods according to their Z. Thus, it is a matter of chemistry, both theoretical and experimental, to validate or contradict such a placement [5]. It is also essential to establish whether trends in properties observed in the chemical groups for the lighter elements is continued with SHE or whether deviations occur due to the increasingly important relativistic effects.

Due to the instability of isotopes of these elements and low production rates, experimental chemical research in this area is very demanding. Chemical experiments are usually designed so that the behavior of the unknown isotope is compared to that of lighter homologs in the chemical group in order to assess their similarity. Such experiments are restricted to measurements of only few properties – volatility and complex formation. Electronic ground-state configurations, lying in the basis of the periodicity, ionization potentials (IPs), electron affinities (EAs), and many other properties, cannot be measured for SHE. Even a chemical composition of SHE species is assumed in experimental studies by analogy in the behavior with that of their lighter homologs in the groups. Thus, in the area of the heaviest elements, chemical theory becomes extremely important and is often the only source of useful information. It also aims at predicting an outcome of sophisticated experiments with single atoms and interpreting their results. Finally, it is only the theory that can reveal relativistic effects influence on chemical properties of SHE: only by comparing the observed behavior with that predicted on the basis of relativistic *vs* nonrelativistic calculations can the importance and magnitude of relativistic effects be established.

In the past, predictions of chemical properties of the heaviest elements were made with the help of relativistic atomic calculations and extrapolations of the periodic trends [6,7]. Due to recent developments in the relativistic quantum theory, calculation algorithms, and computer techniques, very accurate calculations for

SHE and their compounds became possible. On their basis, reliable predictions of SHE properties and experimental behavior required for their chemical identification have been made. Examples of these studies are given in this chapter. Some latest reviews on this subject are those of [5, 8–11].

Relativistic and QED Effects on SHE

The relativistic mass increase for a particle (an electron) with velocity v is

$$m = m_0 \left[(1 - (v/c)^2\right]^{-1/2} \tag{1}$$

where m_0 is the mass at zero velocity (rest mass) and c is the speed of light. The Bohr model for a hydrogen-like species gives the following expressions for the velocity, energy, and orbital radius of an electron:

$$v = (2\pi e^2/nh)Z; \tag{2}$$

$$E = -(2\pi^2 e^4/n^2h^2)mZ^2; \tag{3}$$

$$r = Ze^2/mv^2, \tag{4}$$

where n is the principal quantum number, e is the charge of the electron, and h is Planck's constant. For SHE, m/m_0 is dramatically enlarging. It is, e.g., 1.79 for Fl and 1.95 for element 118. As a consequence, all the three relativistic effects on the valence AOs – contraction and stabilization of the s and $p_{1/2}$ AOs, expansion and destabilization of the $p_{3/2}$, d, f, and g AOs, and the spin-orbit (SO) splitting of AOs with $l > 0$ – are also very large for SHE. Figure 2 shows, e.g., the relativistic stabilization of the ns and $np_{1/2}$ AO and the SO splitting of the np AOs of group-14 elements, with the latter reaching 50 eV for element 164 [6].

For element 112, Cn, the 7s AO is 10 eV relativistically stabilized and 25 % contracted (Fig. 3). At this element, the relativistic contraction and stabilization of the 7s AOs reach their maximum in the seventh row of the periodic table [10].

In the 6d series, the relativistic destabilization and the SO splitting of the 6d AOs increase. Together with the stabilization of the 7s AO, this results in an inversion of the 7s and $6d_{5/2}$ energy levels at Cn, so that its first ionized electron is $(n-1)d_{5/2}$, but not the ns one as in Hg (Fig. 3). (The inversion of the 7s and $6d_{5/2}$ levels in the 7th row starts already at element 108, Hs.) The example of group-12 elements also shows that trends in the relativistic and nonrelativistic energies and R_{max}(ns) AOs (the same is valid for the $np_{1/2}$ AOs) are opposite with increasing Z in the groups, which results in opposite trends in relativistic and nonrelativistic properties of the elements defined by those AOs.

In the 7p series, the $7s^2$ pair is so stabilized that it becomes practically an inert core (Fig. 2). The 7p AO SO splitting is also very large: 4.7 eV for Fl and 11.8 eV for element 118 [12]. The relativistic stabilization and contraction of the 8s AO of

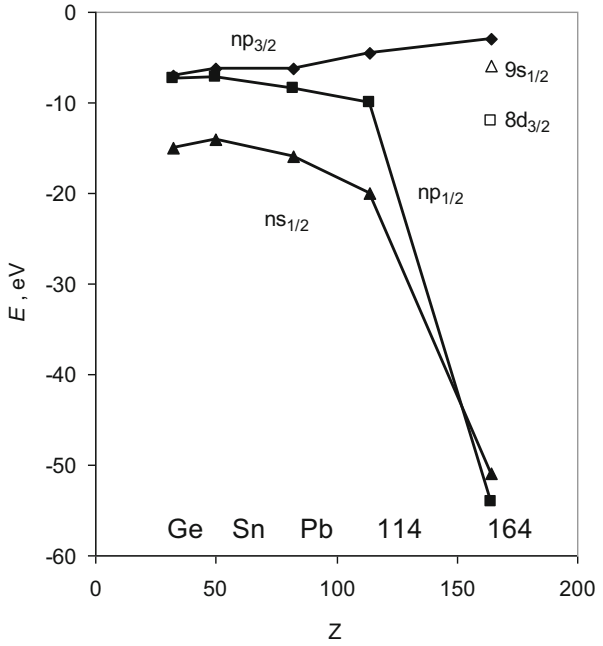


Fig. 2 Dirac-Slater eigenvalues of the valence electrons of group-14 elements in the sp^2p configuration (Redrawn from [6])

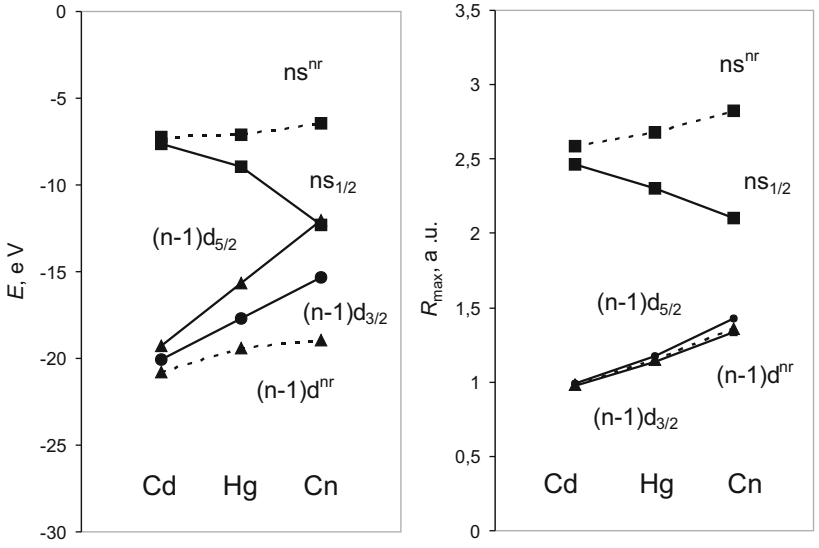


Fig. 3 Relativistic (*solid line*) and nonrelativistic (*dashed line*) energies and the maximum of the radial charge density, R_{max}, of the valence 7s and 6d AOs. The Dirac-Fock data are from [12] (Reprinted with permission from Ref. [11]. Copyright 2011 Oldenbourg Wissenschaftsverlag GmbH)

elements 119 and 120 are also enormous, so that they should behave like K and Ca, respectively. For the heavier elements, relativistic effects are even more pronounced and could lead to properties very much different from those of the lighter homologs [6]. Without relativistic effects the properties would, however, have been also very much different due to the diffuse valence s- and p-AOs and compact d, f, and g AOs [10].

Breit effects accounting for magnetic and retardation interactions on valence orbital energies and IP of the heaviest elements are small, for example, only 0.02 eV for element 121. They can, however, reach a few % for the fine-structure-level splitting in the 7p elements and are of the order of correlation effects there. In element 121, they can be as large as 0.1 eV for transition energies between states including f AOs [13].

Quantum electrodynamic (QED) effects, such as vacuum polarization and electron self-energy, are known to be very important for inner shells, for example, in accurate calculations of X-ray spectra. For the valence shells, the Breit and Lamb-shift terms were shown to behave similarly to the kinetic relativistic effects scaling as Z^2 [14]. For the group-11 and group-12 valence s-shells, the increase with Z is even larger. The nuclear volume effect grows faster with Z. Consequently, for SHE, its contribution to the orbital energy will be the second important one after the relativistic contribution. Thus, e.g., for element 118, QED effects on the binding energy of the 8s electron cause a 9 % reduction (0.006 eV) of EA [15]. QED corrections for some SHE are given in [16].

Chemical Experiments

Due to the short half-lives of SHE isotopes and low production rates, special techniques had to be developed that allow for measurements of macrochemical properties of these elements on the basis of single-atom events. Chemical separations in which a single atom rapidly participates in many identical chemical interactions to two-phase systems with fast kinetics that reach equilibrium quickly turned out to be appropriate. Thus, it is sufficient to combine results of many separate "one-atom-at-a-time" experiments or identical experiments with only one atom, in order to get statistically significant results. Two main types of experimental techniques – gas phase and liquid chemistry chromatography – are based on this principle.

- Gas-Phase Chemistry

Gas-phase chemistry deals with the study of volatility of the heaviest elements or their compounds [5, 8, 17, 18]. In macrochemistry, a measure of volatility of a substance is its sublimation enthalpy, ΔH_{sub}. In the gas-phase chromatography, a measure of volatility is the adsorption enthalpy, ΔH_{ads}, of a species on the surface of gold- or SiO_2-plated detectors located along the chromatography column. (Gold is chosen as it is free from oxide layers.) The obtained ΔH_{ads} are then used to

deduce ΔH_{sub} via a loose correlation between these quantities. Many assumptions are involved in this approach.

There are two kinds of such techniques. In the first one, *thermochromatography*, a longitudinal, negative temperature gradient (from about room temperature down to about −180 °C), is established along the chromatography column through which a gas stream is conducted. It contains volatile species of interest (atoms or molecules) that deposit on the surface of the column according to their volatilities. The deposition zones are registered by detectors, which are associated with specific deposition (adsorption) temperatures, T_{ads} (Fig. 4, left panel). The obtained T_{ads} are then used to deduce the adsorption enthalpy, ΔH_{ads}, using adsorption models and Monte Carlo simulations [17].

In another technique, *isothermal chromatography*, the entire column is kept at a constant temperature. Volatile species pass through the column undergoing numerous adsorption-desorption steps on the surface of the column, usually made of quartz. Their retention time in the column is indicative of their volatility at a given temperature. A series of temperatures is run, and the chemical yield of the species is studied as a function of the temperature (Fig. 4, right panel). A temperature, $T_{50\,\%}$, at which 50 % of the species pass through the column is taken as a measure of volatility in a comparative study. A Monte Carlo program is used to deduce ΔH_{ads} from the measured $T_{50\,\%}$ using adsorption models [17].

Both techniques were used to study volatility of compounds of Rf through Hs and of atoms of Cn and Fl, whose isotopes have half-life, $t_{1/2}$, of the order of at least one second [2, 5, 18].

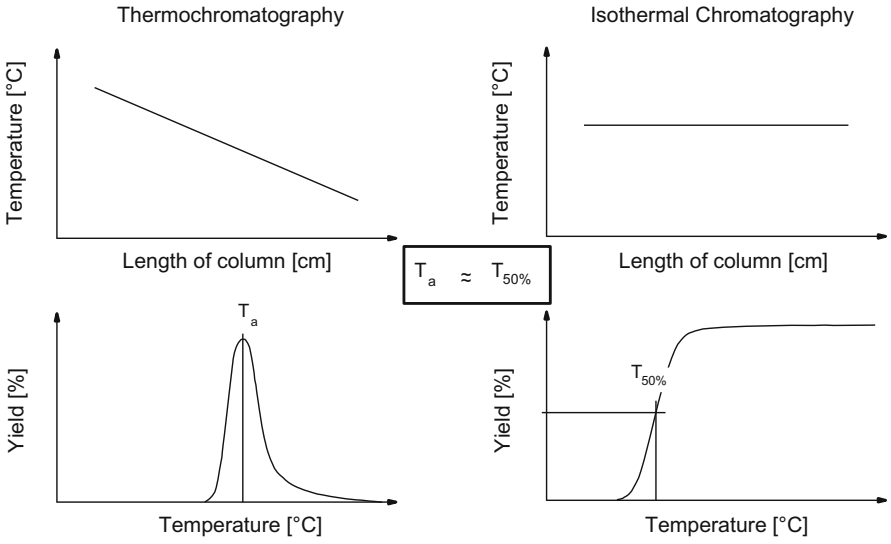


Fig. 4 *Upper panel*, temperature profiles employed in thermochromatography and isothermal chromatography, and *lower panel*, deposition peak and integral chromatogram resulting from thermochromatography and isothermal chromatography, respectively (Reprinted with permission from Ref. [18]. Copyright 2003 Kluwer Academic Publishers)

- Aqueous chemistry

To study complex formation of the heaviest elements and their homologs in aqueous solutions liquid-liquid or ion (cation, CIX, or anion, AIX) exchange chromatography fast separation techniques have been developed [19]. They allow for measurements of the distribution coefficient, K_d, between aqueous and organic phases:

$$K_d = \frac{K_{DM}[RB^+L^-]^p_{\text{org}}\beta_i\,[L^-]^{i-p}}{\sum\limits_0^N \beta_n\,[L^-]^n}, \tag{5}$$

where β_i is the complex formation constant. Obtained K_d (usually plots of K_d values *vs* acid concentration) are used to judge about stabilities of the formed complexes, β_i. In experiments with radioactive species, K_d is measured as a ratio of the activity of the studied species in the organic phase to that in the aqueous phase. It is closely related to the key observable, the retention time, t_r, in the chromatography column:

$$K_d = (t_r - t_0)\frac{V}{M}, \tag{6}$$

where t_0 is the column holdup time due to the free column volume, V is the flow rate of the mobile phase, and M is the mass of the ion exchanger.

Aqueous chemistry experiments that are more time-consuming than the gas-phase ones, mainly because of the time needed for the preparation of a sample suitable for α-spectroscopy, require isotopes with longer $t_{1/2}$, of the order of minutes. Therefore, only complex formation of Rf, Db, and Sg has been studied so far [19].

Relativistic Quantum Chemical Methods and Approaches

- Relativistic quantum chemical methods

For reliable predictions of SHE properties, quantum chemical methods should treat relativistic and correlation effects at the highest possible level of theory. A short description of the relativistic methods is given below. Comprehensive reviews can be found in [20, 21], as well as in this issue.

Wave-function-based (ab initio) methods. The most straightforward way to solve the Dirac many-electron equation

$$h_{\text{DCB}} = \sum_i h_D(i) + \sum_{i<j}\left(1/r_{ij} + B_{ij}\right), \tag{7}$$

where the one-electron Dirac operator is

$$h_D(i) = c\vec{\alpha}_i \vec{p}_i + c^2(\beta_i - 1) + V^n(i), \tag{8}$$

and $\vec{\alpha}$ and β are the four-dimensional Dirac matrices, V^n is the nuclear attraction operator, and the Breit term in the low photon frequency limit is

$$B_{ij} = -1/2\lfloor(\vec{\alpha}_i \vec{\alpha}_j) r_{ij}^{-1} + (\vec{\alpha}_i \vec{r}_{ij})(\vec{\alpha}_j \vec{r}_{ij}) r_{ij}^{-3}\rfloor, \tag{9}$$

is that without approximations. The operators of the Dirac equation (7) are 4×4 matrix operators, and the corresponding wave function is therefore a four-component (4c) vector (spinor). The V^n includes the effect of the finite nuclear size, while some finer effect, like QED, can be added to h_{DCB} perturbatively, although the self-energy QED term is more difficult to treat. The DCB Hamiltonian in this form contains all the effects through the second order in α, the fine-structure constant.

Since the relativistic many-body Hamiltonian cannot be expressed in closed potential form, which means it is unbound, projection one- and two-electron operators are used to solve this problem. The operator projects onto the space spanned by the positive-energy spectrum of the Dirac-Fock-Coulomb (DFC) operator. In this form, the "no-pair" Hamiltonian is restricted then to contributions from the positive-energy spectrum and puts Coulomb and Breit interactions on the same footing in the SCF calculation [22]. The proper way to go beyond the "no-pair" approximation has recently been discussed by Liu [23].

Because the Dirac equation is written for one electron, the real problem of ab initio methods for a many-electron system is an accurate treatment of electron correlation. The latter is of the order of magnitude of relativistic effects for binding energies and other properties. The DCB Hamiltonian (Eq. 7) accounts for these effects in the first order via the $V_{ij} = 1/r_{ij}$ term. Some higher orders of magnitude correlation effects are taken into account by the configuration interaction (CI) and many-body perturbation theory (MBPT) techniques, including the Møller-Plesset (MP) theory, or, presently, at the highest level of theory, coupled cluster with single-double and perturbative triple, CCSD(T), excitations, or Fock-space CC (FSCC) techniques [13, 24, 25].

The problems of electron correlation and proper basis sets make the usage of the 4c ab initio DF(C) methods rather limited in molecular calculations. These methods are still too computer time intensive and are not sufficiently economic to be applied to the heaviest elements in a routine manner, especially to the complex systems studied experimentally. Mostly small molecules, like hydrides or fluorides of SHE, were calculated with their use. The DC method is also implemented in the DIRAC program package [26].

Two-component (2c) methods. Due to the practical limitations of the 4c methods, the 2c ones are very popular in molecular calculations. In this approximation, the "positronic" and electronic solutions of the Dirac-Hartree-Fock (DHF) method are decoupled. This reduces the number of matrix elements in the Hamiltonian to inter-

actions solely among electrons (positive-energy states) and nuclei and, therefore, saves valuable computer time. Perhaps, the most applied method of decoupling the large and small components of the wave function is the Douglas-Kroll-Hess (DKH) approximation [27]. The 2c Hamiltonians often used in molecular applications are X2C [28–31] and BSS [32], also implemented in the DIRAC program package [26]. A comprehensive review on the X2C methods is that of [29].

Effective core potentials (ECPs) allow for more economic calculations within the DHF schemes by replacing inner core orbitals that do not take part in the bond formation by a special (effective core) potential. In this way, the number of basis functions and, therefore, two-electron integrals is drastically diminished. There are ECPs of two main types, as well as pseudo-potentials (PPs) and model potentials (MPs). Energy-adjusted PPs known as the Stuttgart ones [33, 34] and the shape-consistent relativistic ECPs (RECPs) [35, 36] are available for SHE. Generalized RECPs accounting for Breit effects have also been developed for some heaviest elements [37].

Density functional theory (DFT) is based on the knowledge of the ground-state electron density. Due to the high accuracy and efficiency, computational schemes based on the DFT methods are among the most popular in theoretical chemistry, especially for extended systems, such as large molecules, liquids, or solids [38]. Usually, self-consistent all-electron calculations are performed within the relativistic local density approximation (LDA). The general gradient approximation (GGA), also in the relativistic form (RGGA), is then included perturbatively in E^{ex}, the exchange correlation energy functional. The accuracy depends on the adequate knowledge of E^{ex}, whose exact form is, however, unknown. There is quite a number of such potentials and their choice is dependent on the system. Thus, PBE is usually favored by the physics community, PBE0, B3LYP, B88/P86, revPBE, etc., and by the chemistry community, while LDA is still used extensively for the solid state.

Spin-unrestricted, or spin-polarized (SP), 4c-DFT methods allowing for accurate calculations of open-shell systems are those of Anton et al. [39], the Beijing group (BDF) [40], and ReSpect [41]. The first two were extensively used for SHE. They differ by basis set techniques, though they give similar results. The 4c-DFT of Anton et al. [39] allows for treating explicitly very large systems such as clusters of up to more than 100 atoms and is, therefore, suitable for the treatment of adsorption phenomenon on surfaces of solids.

The 2c-DFT methods are a cheaper alternative to the 4c ones [29, 30, 42, 43]. Quasi-relativistic methods such as the spin-orbit zeroth-order regular approximation (SO ZORA) implemented in the Amsterdam DFT code (ADF) [44] and the DKH method [45] implemented in most program packages are also popular among theoretical chemists. Dispersion-corrected E^{ex} are available in ADF for $Z \leq 92$.

Periodic DFT codes are generally not developed for the SHE: they have neither basis sets nor suitable PP. Also, as a rule (with the exception of RFPLO [46] and ADF-BAND [44]), they are scalar relativistic (SR). In some cases, like, e.g., for calculations of solid Cn and Fl, PPs have been specially developed using the Stuttgart PP model [33]. The ADF-BAND [44] having basis sets for elements up to $Z = 120$ is presently an efficient tool to perform solid-state and adsorption calculations for SHE.

- Approaches used to predict experimentally measurable properties

For weak interactions, e.g., for predictions of the transport of SHE nuclides through Teflon or polyethylene capillaries from the accelerator to the chemistry set up or for the adsorption of volatile species on inert surfaces of a chromatography column, the usage of adsorption models proved to have advantages over direct DFT calculations of adatom-surface interactions. For example, for an atom or a symmetric molecule with a zero dipole moment adsorbed on a dielectric surface, the dispersion interaction energy can be calculated using the following equation coming from an adatom-slab adsorption model [47]:

$$E(x) = -\frac{3}{16}\left(\frac{\varepsilon - 1}{\varepsilon + 2}\right)\frac{\alpha_{\mathrm{mol}}}{\left(\frac{1}{IP_{\mathrm{slab}}} + \frac{1}{IP_{\mathrm{mol}}}\right)x^3}, \tag{10}$$

where ε is the dielectric constant of the adsorbent material and x is the adatom/molecule – surface distance (usually van der Waals radius). All the atomic or molecular properties of Eq. (10) can be accurately calculated using relativistic codes. Since the detailed structure of the surface of the column is, as a rule, unknown, in a comparative study, it is reasonable to predict ΔH_{ads} of a SHE species with respect to the measured ΔH_{ads} of a homolog. Thus, x can be deduced from the measured $-\Delta H_{\mathrm{ads}} \approx E_b$ (binding energy) of a lighter homolog using Eq. 10, while x of a species of interest can be estimated using a difference in their van der Waals radii.

In the case of adsorption of atoms on chemically reactive surfaces such as gold or quartz, calculations of E_b can be directly performed either via a cluster model (Fig. 5) using molecular codes or adatom-slab/supercell models using relativistic periodic codes (e.g., ADF-BAND [44]).

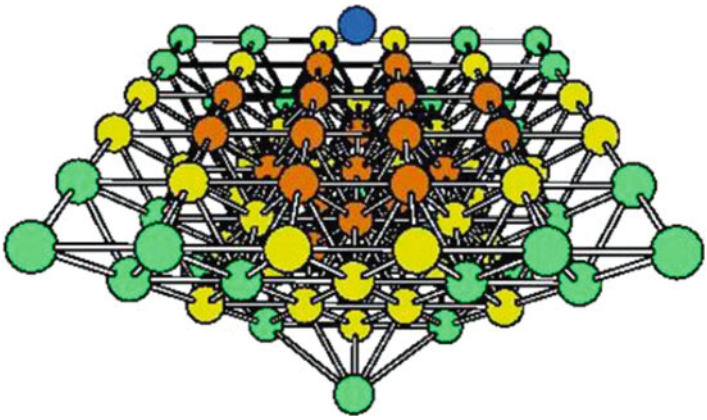


Fig. 5 A cluster model for the prediction of adsorption (Reprinted with permission from Ref. [9]. Copyright 2014 Springer Verlag)

Various correlations, e.g., between E_b in smaller systems and ΔH_{ads} or ΔH_{sub} of substances, turned also out to be useful in predictions of adsorption and sublimation properties of SHE (see below). Knowing then ΔH_{ads}, a relative yield of a volatile species at the end of a capillary or a chromatography column or its T_{ads} can be predicted using a model of mobile adsorption (see [5, 9] for reviews).

Atomic Properties of SHE and the Structure of the Periodic Table

Electronic configurations. Ground-state configurations of SHE from Z = 104 through Z = 172 were predicted in the past with the use of DS and DF methods [6, 7]. Later, MCDF calculations for neutral and ionized states of Rf through Hs [48, 49] and for grounds states of elements 119 through 164 [50–52] were performed. With the development of CC methods, results of the DC(B) FSCC calculations for Rf and Rg through Fl and for elements 118 through 122 became available (see [13] for a review).

All these calculations have shown that the relativistic stabilization of the 7s AO results in the availability of the $7s^2$ electron pair in the ground states of the 7th row elements, $7s^26d^q$ and $7s^27p^p$. This is in contrast to the 6th row, where Pt and Au have different ground states, $5d^96s$ and $5d^{10}6s$, respectively. For Rf, the MCDF calculations have given the $7s^27p6d$ configuration as the ground state [48]. More accurate DCB FSCC calculations have, however, corrected the MCDF result leading to the $7s^26d^2$ configuration as the ground [13]. A very high level of correlation with $l = 6$ was required to reach this accuracy. For elements 119 and 120, the 8s and $8s^2$ states, respectively, beyond the 118 core were found as most stable. Element 121 has an $8s^28p_{1/2}$ state in the difference to $Ac(7s^26d)$ due to the relativistic stabilization of the $8p_{1/2}$ AO. All these calculations generally agree on the ground states of the elements up to $Z = 122$. They, however, disagree at Z > 122 (Table 1).

Elements beyond the 7th row of the periodic table are characterized by mixing of states coming from partially filled $8p_{1/2,3/2}$, $7d_{3/2,5/2}$, $6f_{5/2,7/2}$, and $5g_{7/2,9/2}$, shells. The proximity of the valence SO bands makes the search for the correct ground state very difficult. The usual classification on the basis of a simple electronic configuration and the placement of these elements in a proper column in this part of the periodic table becomes, therefore, problematic. Thus, the filling of the electron shells and the structure of the periodic table beyond $Z = 122$ are still under discussion and debate (see, e.g., different versions of the periodic table in [6] and [53]). Accurate calculations, preferably using DBC CC + QED techniques,

Table 1 Ground states of element 121–124 (Z = 120 core +) and 143 (Z = 120 core + $8p_{1/2}^2$+)

Method	121	122	123	124 ...	143	Ref.
DF	8p	7d8p	6f7d8p	$6f^38p$	$5g^{17}6f^27d^2$	[6]
DF	8p	7d8p	6f7d8p	$6f^27d8p$	$5g^{18}7d^3$	[50]
MCDF	8p	7d8p	$6f^28p$	$6f^28p^2$	$5f^{17}6f^27d^2$	[51]
DCB FSCC	8p	7d8p	–	–	–	[13]

Table 2 Polarizabilities, α (in a.u.), and ionization potentials, IPs (in eV), of Hg and Cn

Method	Hg		Cn		Ref.
	α	IP	α	IP	
4c-BDF PBESIC	36.4	10.40	29.8	11.40	[54]
QR PP CCSD(T)	34.2	10.37	28.0	13.17	[54]
AR PP CCSD(T)	34.42	–	25.82	–	[55]
ECP CCSD(T)	28.48	10.39	28.68	11.675	[56]
DC CCSD(T)	34.15	10.445	27.64	11.97	[57,58]
Exp.	33.919	10.4375			[59]

are, therefore, highly needed in this area to resolve the contradictions. At the time being, advanced MCDF calculations including QED effects [52] confirmed earlier predictions from single-configuration DF calculations [6] about the end of the periodic table at $Z = 173$, when the energy of the 1s level becomes less than $-2mc^2$.

Ionization potentials. Various calculations – DF, DS, MCDF, and DC(B) CC – of IPs were performed for elements 104 through 166 (see [6–9, 13, 54–58]). The accuracy upon approximation is shown in Table 2 for Hg and Cn, as an example, with the DC CCSD results [57,58] being the most accurate.

The influence of relativistic effects on IP and other atomic properties of group-12 elements is shown in Fig. 6, as an example. The relativistic IPs are larger than the nonrelativistic ones, and they increase in group 12, with the maximum at Cn. Nonrelativistically, IP(Cn) would have been about that of Cd. Relativistic effects act similarly on elements 113 and 114, i.e., they cause an increase in IPs due to the relativistic stabilization and contraction of the $np_{1/2}$ AO. On the contrary, in groups 15 through 18, IPs decrease with atomic number due to the destabilization of the $np_{3/2}$ AO, so that DC FSCCSD IP(118) of 8.914 eV [60] is smaller than IP(Rn) of 10.799 eV [59].

Especially interesting are trends in properties of group-1 and group-2 elements, having the ns and ns^2 ground states, respectively. The trends in IP are reversed in groups 1 and 2 at Cs and Ba, respectively (Fig. 7) due to the trend reversal in the energies of the ns AOs. Thus, IP(119) should be about IP(K), and IP(120) should be about IP(Ca).

Electron affinities. EAs were calculated for a few of the heaviest elements. No bound anion was found for Cn by the DCB FSCC calculations [58]. Fl was shown to have no EA at the DC FSCC level of theory either [61]. On the contrary, element 118 has a positive EA of 0.058 eV, according to the DCB FSCC + QED calculations [15]. This is a result of the 8s AO relativistic stabilization.

EAs of group-1 and group-2 elements, like their IPs, show a reversal of the trends from the 6th row on, however, group 1 behaves differently from group 2. Thus, in group 1, the trend to a decrease in EA at the lighter elements is changed at Cs to an *increase* toward element 119 (Fig. 7), while in group 2, the trend to an increase at the lighter elements is changed at Ba to a *decrease* toward element 120. These opposite

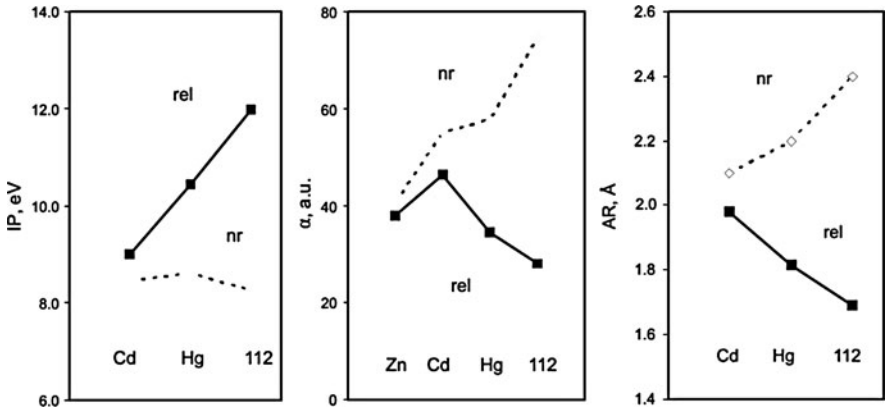


Fig. 6 Relativistic (*solid lines*) and nonrelativistic (*dashed lines*) ionization potentials (IPs), atomic radii (AR), and polarizabilities, α, of group-12 elements (Reprinted with permission from Ref. [47]. Copyright 2005 Elsevier)

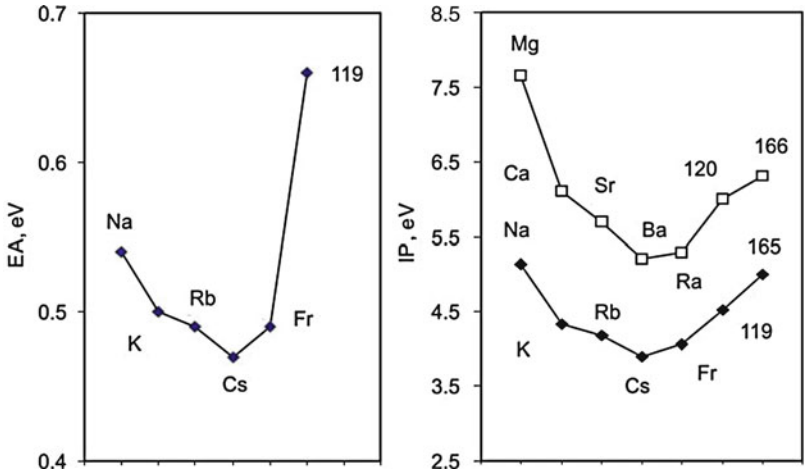


Fig. 7 Electron affinities, EAs, and ionization potentials, IPs, of alkali and alkaline-earth elements. The data for Na through Fr and Mg through Ra are experimental

trends in EA in groups 1 and 2 are due to the opposite trends in the energies of *different* AOs responsible for the electron acceptance process – the ns AOs in the former case and the $np_{1/2}$ + (n−1)d AOs in the latter. It was also shown that inclusion of the triple excitations (T) in the CC procedure for the electron correlation is crucial in stabilizing the 120 anion: the RCCSD(T) result is 21 meV, while the DHF result (without correlation) is much worse, −121 meV (Table 3) [62]. The EA of element 121 is the highest in group 3 due to the relativistic stabilization of the $8p_{1/2}$ AO [13].

Table 3 Electron affinity of group-2 elements calculated within different approximations [62]

Elem.	State	DHF	RCCSD	RCCSD(T)	Exp.
Ba	$6s^2 6p$	−0.143	0.070	0.138	0.144
Ra	$7s^2 7p$	−0.099	0.042	0.082	> 50
120	$8s^2 8p$	−0.121	−0.002	0.021	–

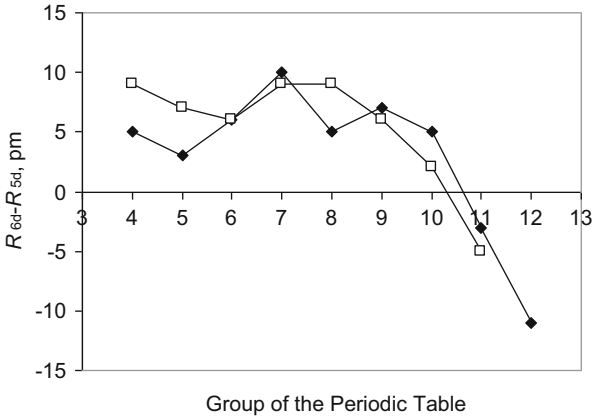


Fig. 8 The difference in the lengths of the single (*filled rhomboids*) and triple (*open squares*) bonds between the 6d and 5d metals (Reproduced with permission from Ref. [53]. Copyright 2011 Royal Society of Chemistry)

Atomic/ionic/covalent radii. Atomic (AR) and ionic (IR) radii of SHE were predicted using DS/DF and MCDF calculations of $R_{\max}$ of the charge density of outer AOs [6,48,49]. A set of atomic single- and triple-bond covalent radii (CR) for most of the elements of the periodic table including the heaviest ones till Z = 118 and 112, respectively, is also suggested in [63]. They were obtained from calculated bond lengths, R, in simple compounds. All the results show that the CR of the group 4 – 8 6d elements – are about 0.5–0.8 Å larger than those of the 5d elements. An important finding of these works is a decrease in the $R_{6d} - R_{5d}$ difference starting from group 9, reaching negative values in groups 11 and 12, as a result of the increasing 7s AO contribution (Fig. 8). This is called a "transactinide break" [63]. The relativistic AR contraction of Cn is also shown in Fig. 6.

In groups 13 and 14, AR also decrease with Z due to the relativistic contraction of the $7p_{1/2}$ AOs, while in groups 15 through 18, they increase with Z due to the relativistic expansion of the $np_{3/2}$ AOs.

Polarizabilities. Static dipole polarizabilities, α, were calculated at various levels of the relativistic quantum theory for Cn through element 119 (see [9] for a review). Table 2 shows results of various calculations for Hg and Cn [54–57]. According to the calculations, α(Cn) should be the smallest in group 12 due to the relativistic contraction of the outer 7s AO (Fig. 6). Correlation effects were shown to decrease α in Hg, Cn, and Pb and to increase it in Fl. Thus, one can see that exclusively due to

relativistic effects, Cn should be chemically rather inert, much more than the lighter homologs in group 12, while element 118 should be chemically most reactive in group 18, having a DC CCSD(T) α of 46.33 a.u., larger than α(Rn) of 35.04 a.u. [60].

The calculated atomic properties were used to predict the transport of single atoms from the accelerator to the chemistry setup though Teflon capillaries. Using Eq. (10), $-\Delta H_{\rm ads}$ of elements 113, 114, 118, and 120 were obtained as 14, 10.4, 10.8, and 35.4 kJ/mol, respectively, ensuring their safe delivery [11, 57, 60, 62]. The relative yield of SHE at the end of the transport system has also been given using the $\Delta H_{\rm ads}$. Thus, e.g., for the $^{300}120$ isotope with $t_{1/2} = 1$ s and $-\Delta H_{\rm ads}(120) = 35.4$ kJ/mol, the relative yield is calculated as 90 % and 60 % for an open Teflon column or a capillary with an inner diameter of 2 mm and a length of 1 m and 10 m, respectively, and a gas volume $Q = 1$ l/mol at room temperature [62]. Thus, for this element, the limiting factor for the delivery is not its volatility, but the short half-life.

Gas-Phase Chemistry

Groups 4–8, common features. Elements at the beginning of the 6d series were shown to form volatile halides, oxyhalides, and oxides by analogy with their lighter homologs in groups 4–8. With the aim to predict their stability and volatility, calculations for the following species – MF_4, MCl_4, $MOCl_2$, MBr_4 (M = Zr, Hf, and Rf), MCl_5, MBr_5, $MOCl_3$ (M = Nb, Ta, and Db), MCl_6, MO_3, $MOCl_4$, MO_2Cl_2, $M(CO)_6$ (M = Mo, W, and Sg), MO_3Cl (M = Tc, Re, and Bh), MO_4 (M = Ru, Os, and Hs), and MX (M = Rf – Cn, X = H, N, B, and C) – were performed with the use of relativistic DFT and CC methods. (A full list of calculated properties is given in Tables 8 and 9 of [9]). The calculations confirmed that compounds of Rf through Hs are homologs of the lighter congeners in the chemical groups and that bonding is due to the participation of the 6d and 7s AOs. An increase in covalence (a decrease in effective metal charges, Q_M, and an increase in overlap population, OP), as well as in the stability of the maximum oxidation state in the groups has been established. It was shown to be a relativistic effect (Fig. 9) [9].

The atomization energies, D_e, of the 6d-element molecules turned out to be smaller than D_e of the lighter homologs due to the larger SO effects and a smaller ionic contribution. Thus, e.g., SP 4c-DFT $D_e(RfCl_4)$ is 19.53 eV, smaller than $D_e(ZrCl_4)$ of 21.7 eV and $D_e(HfCl_4)$ of 21.1 eV [9]. Electron correlation is proven to contribute to more than 50 % to bonding in the SHE systems (see Table 5 below). Electron correlation and relativistic effects were shown to be nonadditive.

Group 4. In the difference to expectations, an unusual trend in volatility has been observed for group-4 halides, $ZrCl_4 \approx RfCl_4 > HfCl_4$ and $ZrBr_4 \approx RfBr_4 > HfBr_4$, using isothermal gas-phase chromatography with a quartz column (Fig. 10) (see [5] for a review).

To interpret this unusual behavior and to prove stability of the halide and oxyhalide of Rf, 2c- and 4c-DFT calculations were performed for MCl_4 and $MOCl_2$

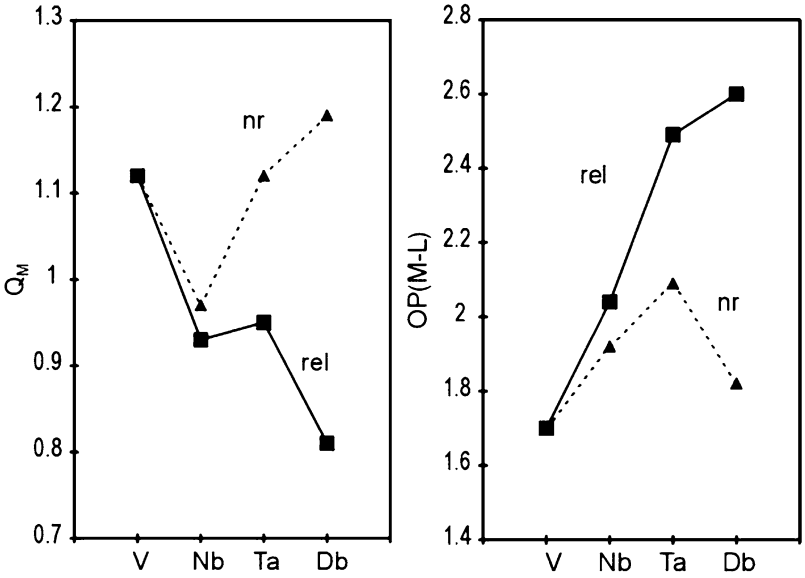


Fig. 9 Relativistic (*solid lines*) and nonrelativistic (*dashed lines*) effective charges, Q_M, and overlap populations, OP, for MCl_5 (M = V, Nb, Ta, and Db). L denotes the ligand (Reprinted with permission from Ref. [9]. Copyright 2014 Springer Verlag)

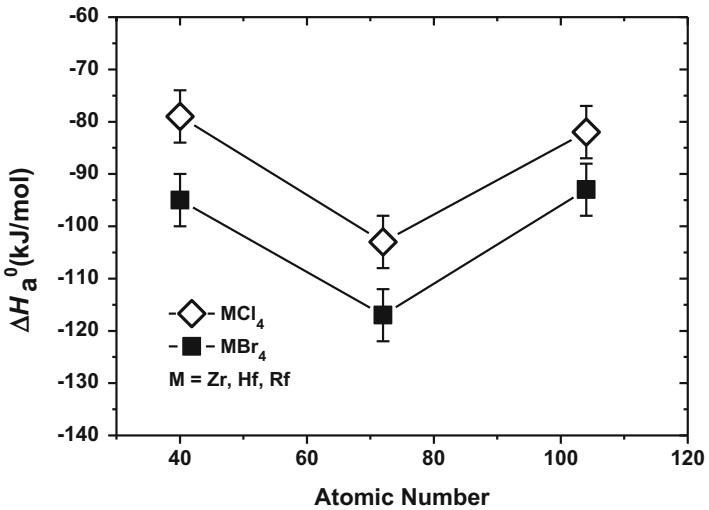


Fig. 10 Adsorption enthalpies of group-4 tetrachlorides and tetrabromides (Reproduced with permission from Ref. [5]. Copyright 2013 American Chemical Society)

Table 4 The X2C (B88/P86) calculated properties of MCl_4 (M = Zr, Hf, and Rf): equilibrium bond lengths, R_e (in Å); atomization energies, D_e (in eV); ionization potentials, IPs (in eV); polarizabilities, α (in a.u.); as well as adsorption enthalpies, ΔH°_{ads} (in kJ/mol), on a quartz surface and sublimation enthalpies, ΔH°_{sub} (in kJ/mol), of bulk

Molecule	R_e	D_e	IP	α	$\Delta H^{\circ a}_{ads}$	$\Delta H^{\circ b}_{ads}$	$\Delta H^{\circ c}_{sub}$
$ZrCl_4$	2.336	20.34	11.00	103.6	106.5	97	110.5
$HfCl_4$	2.316	20.80	11.00	99.3	103	103	104.7
$RfCl_4$	2.370	19.40	10.96	101.2	102.7	87	(104.2)[a]

[a]Theory [64, 65]
[b]Experiment (see [5] for references)
[c][59]

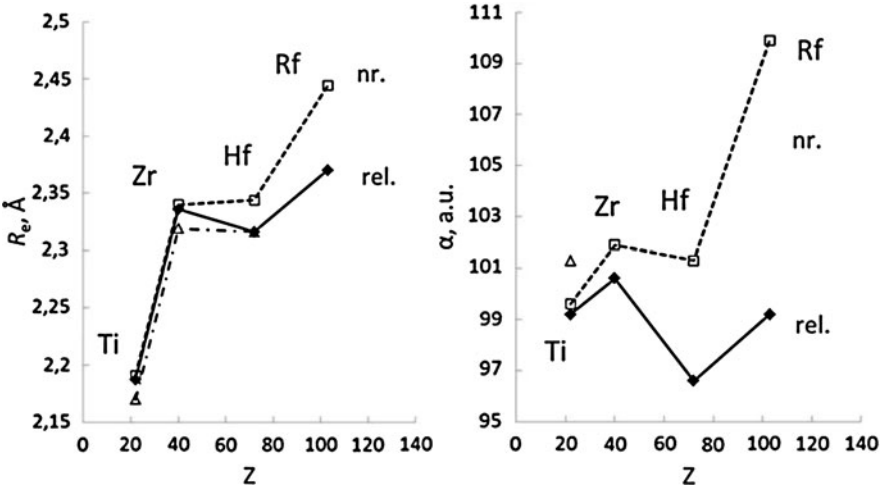


Fig. 11 Relativistic (*solid lines*) and nonrelativistic (*dashed lines*) bond lengths, R_e, and polarizabilities, α, of MCl_4 (M = Ti, Zr, Hf, and Rf). Experimental values are shown with *open triangles* connected by *dashed-dotted lines* (Reproduced with permission from Ref. [64]. Copyright 2014 American Institute of Physics)

(M = Ti, Zr, Hf, and Rf) [64, 65]. Results are shown in Table 4 for MCl_4. The trend in the formation of the tetrachlorides from the oxychlorides, formed in the oxygen atmosphere ahead of the pure halides, was found to be Zr < Hf < Rf. This means that if the pure chlorides of Zr and Hf exist under experimental conditions, $RfCl_4$ should also be stable.

Trends in R_e and α of MCl_4 in the group, as well as the action of relativistic effects on them, are shown in Fig. 11. Relativistic effects are found to be responsible for the bond contraction in $HfCl_4$ and for even a larger bond contraction in $RfCl_4$, so that $\alpha(RfCl_4)$ is almost equal to $\alpha(TiCl_4)$.

For the long-range interaction of the MCl_4 (M = Zr, Hf, and Rf) molecules with a quartz surface, ΔH_{ads} were predicted with the use of the calculated molecular properties (Table 4) and Eq. (10). The obtained enthalpies (Table 4) show that volatility of the chlorides should increase smoothly in group 4: Zr < Hf < Rf. Thus,

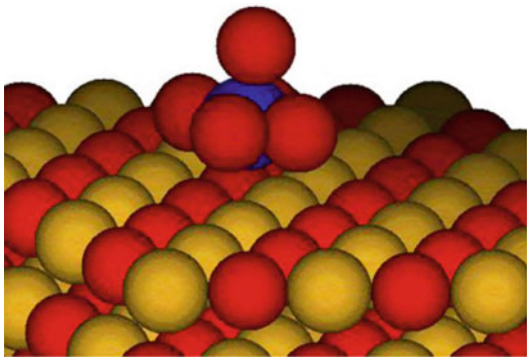


Fig. 12 The ML_6^- (M = Nb, Ta, and Db) complex formed on the KL (L = Cl, Br) surface (Reproduced with permission from Ref. [66]. Copyright 2012 American Institute of Physics)

$RfCl_4$ should be the most volatile species due to the relativistically decreased α. (Nonrelativistically, $RfCl_4$ would have been less volatile than its homologs.) This trend is also in line with $\Delta H_{sub}(ZrCl_4) > \Delta H_{sub}(HfCl_4)$ [59]. For the formation of the MCl_6^{2-} and $MOCl_4^{2-}$ complexes on a chlorinated quartz surface, the calculated energies of the complex formation reactions have given the following smooth trends in volatility: Zr > Hf > Rf and Zr < Hf < Rf, respectively [65]. The trend observed experimentally is, however, reversed, Zr < Hf > Rf, which cannot find its theoretical explanation from the molecular calculations and the assumed adsorption scenarios.

Group 5. Similarly to group 4, unusual reversed trends in volatility were also observed for group-5 halides: $NbCl_5 \approx DbCl_5 > TaCl_5$, $NbBr_5 > TaBr_5 > DbBr_5$, and $NbBr_5 \approx DbBr_5 > TaBr_5$ (see [5] for references). To find a reason for that, calculations for ML_5, MOL_3 (L = Cl and Br), and complexes ML_6 and MCl_5Br (M = Nb, Ta, and Db) that can be formed on a halogenated quartz surface (Fig. 12) were performed using the 4c-DFT method [66].

Results of the calculations reveal a smooth change in properties of the compounds, as well as in their E_{ads} on different surfaces, similarly to the group-4 (oxy)halides, but not the reversed trend observed experimentally.

The difference between the theoretical [64–66] and experimental [5] trends in volatility of group-4 and group-5 halides has not yet found its explanation. Detailed calculations of the interaction of the molecule with the adsorbent are not an easy task as modifications of the surface by the reactive agents occur and the structure of real surfaces is very hard to integrate into the model calculations. This work may, therefore, be prolonged, provided additional studies of surfaces and processes follow from the experimental side, as well as theoretical methods for molecular adsorption receive further development.

Group 6. In group 6, the most stable gas-phase oxychlorides are MoO_2Cl_2 and WO_2Cl_2. Stability and volatility of SgO_2Cl_2 were, therefore, to be investigated using isothermal phase-phase chromatography technique with a quartz column [67]. The 4c Dirac-Slater discrete variation (DS-DV) [68] and RECP [69] calculations for MO_2Cl_2 (M = Mo, W, and Sg) have found SgO_2Cl_2 to be also stable. However, its

Table 5 Correlation and SO effects on the electronic density distribution (Q_M and OP); dipole moments, μ (in D); and atomization energies, D_e (in eV), of MO_2Cl_2 (M = W and Sg)

	Molecule	RECP		DFT
		HF(AREP)[a]	SO-CCSD(T)[b]	DS-DV[c]
Q_M	WO_2Cl_2	2.18	1.71	1.08
	SgO_2Cl_2	1.94	1.52	0.97
OP	WO_2Cl_2	2.14	2.03	2.23
	SgO_2Cl_2	2.72	2.55	2.34
μ	WO_2Cl_2	1.80	1.51	1.35
	SgO_2Cl_2	2.64	2.39	1.83
D_e	WO_2Cl_2	11.7	22.2	23.8
	SgO_2Cl_2	14.2	21.0(22.5[d])	21.8

[a]Average relativistic, i.e., without SO and correlation [69]
[b]With SO effects and correlation [69]
[c]Fully relativistic [68]
[d]Without SO effect

D_e is somewhat smaller than $D_e(WO_2Cl_2)$ due to the SO effects (see Table 5). The following trend in volatility was predicted from the DFT calculated molecular properties (Table 5): $MoO_2Cl_2 > WO_2Cl_2 > SgO_2Cl_2$ [68]. The reason for this trend is increasing dipole moments in this row of molecules causing their stronger interaction with the surface. The experiments [67] have, indeed, demonstrated a decrease in volatility of the group-6 oxychlorides (see [5]), in good agreement with the predictions [68]. Thus, no unexpected behavior was observed in group 6 for Sg.

A search for a new class of volatile species suitable for gas-phase chromatography studies resulted in an idea to synthesize carbonyl complexes of the heaviest elements. Carbonyls proved to be also ideal for the transport of short-lived isotopes of the heaviest elements. Group-6 carbonyls including those of Sg were to be studied first using thermochromatography with a quartz column having a negative temperature gradient [70]. In turn, electronic structures of $M(CO)_6$ (M = Cr, Mo, W, and Sg) were calculated using the CCSD RECPs and 4c-DFT methods [71,72], with the latter work being devoted to predictions of their volatility. Since the nature of the molecule-surface interaction should be dispersive, ΔH_{ads} of Sg and of Mo carbonyls were predicted using Eq. (10) and the calculated molecular properties (Table 6) [72] with respect to the measured $-\Delta H_{ads}$ of $W(CO)_6$ of 46.5 ± 2.5 kJ/mol [70]. Results are shown in Table 6 and Fig. 13.

The obtained ΔH_{ads} of the Sg carbonyl turned out to be almost equal to ΔH_{ads} of $W(CO)_6$ due to the cancelling effects between its larger α, though relativistically decreased, and its larger size (x in Eq. 10). Experimental results [70] confirmed the theoretical predictions [72] showing an almost equal volatility of the W and Sg carbonyls (Table 6). Thus, Sg again proved to be an ordinary member of the group-6 elements.

Group 7. Group-7 elements Tc and Re form stable and volatile oxychlorides, MO_3Cl. Volatility of the Bh oxychloride, which should be formed by analogy with the lighter homologs, was to be studied using gas-phase isothermal chromatography

Table 6 Ionization potentials, IPs (in eV); average dipole polarizabilities, $<\alpha>$ (in a.u.); interatomic distances, R_e (in Å); and adsorption enthalpies, $-\Delta H^{\circ}_{ads}$ (in kJ/mol), of $M(CO)_6$, where M = Mo, W, and Sg, on quartz

Molecule	Method	R_e(M-C)/R_e(C-O)	IP	α	$\Delta H^{\circ a}_{ads}$	$\Delta H^{\circ b}_{ads}$
$Mo(CO)_6$	4c-DFT	2.063/1.145	9.00	156.4	48.1 ± 2.5	50 ± 2
$W(CO)_6$	4c-DFT	2.058/1.148	8.93	151.5	46.5 ± 2.5[c]	49 ± 2
$Sg(CO)_6$	4c-DFT	2.123/1.154	8.63	159.4	46.2 ± 2.5	50 ± 4
	RECP CC	2.112/1.150[d]				

[a]Theory [72]
[b]Experiment [70]
[c]An experimental value taken as a reference point
[d]Theory [71]

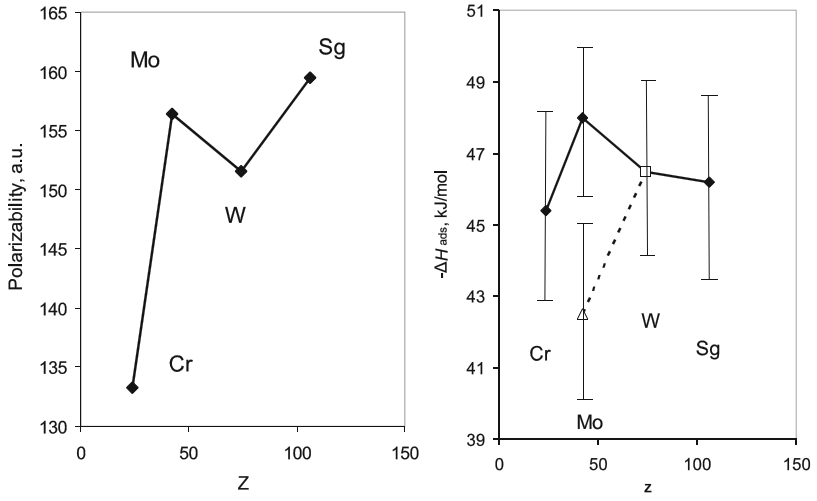


Fig. 13 Calculated (*solid lines*) polarizabilities and adsorption enthalpies of $M(CO)_6$ (M = Cr, Mo, W, and Sg) on quartz [72]. The points (*open symbols*) for Mo and W are results of two different experiments [70] (Reproduced with permission from Ref. [72]. Copyright 2013 American Institute of Physics)

with a quartz column [73]. To predict an outcome of these experiments, energy contributions to the total molecule-surface interaction energy $E(x)$ for molecules having nonzero dipole moments were determined for MO_3Cl (M = Tc and Bh) with respect to $E(x)$ of ReO_3Cl using a model of long-range interactions [74] (Table 7). The resulting $-\Delta H_{ads}(BhO_3Cl)$ of 78.5 kJ/mol and $-\Delta H_{ads}(TcO_3Cl)$ of 48.2 kJ/mol turned out to be in very good agreement with the experimental $-\Delta H_{ads}(BhO_3Cl)$ of 75 kJ/mol and $-\Delta H_{ads}(TcO_3Cl)$ of 51 kJ/mol establishing the following trend in volatility: $TcO_3Cl > ReO_3Cl > BhO_3Cl$ (see Fig. 14, where the BhO_3Cl yield curve is at the highest temperatures). Increasing dipole moments μ of $MOCl_3$ in the group were shown to be responsible for this decreasing trend.

Table 7 Contributions to the interaction energy $E(x)$ of the MO_3Cl molecules (M = Tc, Re, and Bh) with the chlorinated surface [$Q(Cl) = -0.4$] (From [74])

Molecule	μQe		αQe		$\alpha\alpha(Cl)$	
	$E10^{16}$	x^2, eV cm^2	$E10^{32}$	x^4, eV cm^3	$E10^{48}$	x^6, eV cm^6
TcO_3Cl	2.23		5.69		379.1	
ReO_3Cl	3.10		6.81		460.7	
BhO_3Cl	4.67		8.63		591.2	

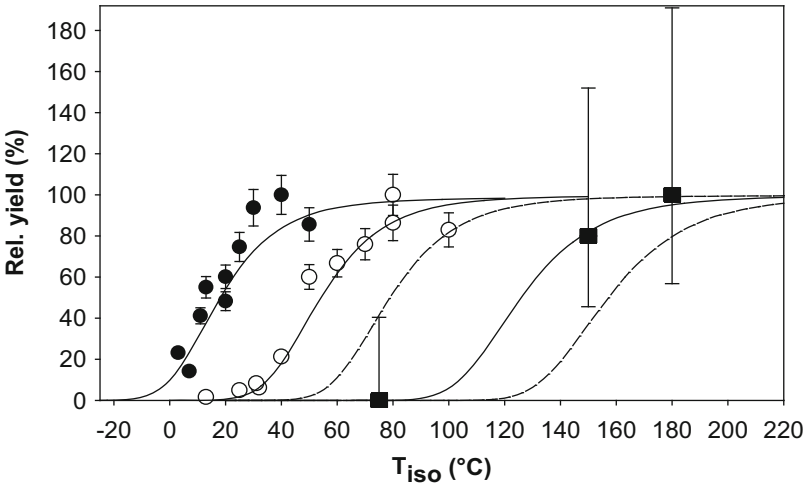


Fig. 14 The relative yields of TcO_3Cl (*filled black circles*), ReO_3Cl (*open circles*), and BhO_3Cl (*filled black squares*) as a function of the isothermal temperatures, T_{iso}, in isothermal gas-phase experiments (Reproduced with permission from Ref. [73]. Copyright 2000 Nature Publishing Group)

Group 8. Group-8 elements Ru and Os are known to form stable and volatile tetroxides, MO_4. Volatility of HsO_4 was, therefore, of interest. It was to be studied with respect to that of OsO_4 using gas-phase thermochromatography with a quartz column [75]. Accordingly, the 4c-DFT calculations were performed for MO_4 (M = Ru, Os, and Hs) [76]. Results are given in Table 8 and Fig. 15 showing excellent agreement of IP, α, and R_e with experimental data for RuO_4 and OsO_4. The high accuracy of the calculations was achieved by using extremely large basis sets.

The measured relatively low $-\Delta H_{ads}(OsO_4)$ on quartz was indicative of the van der Waals type of adsorption [75]. Taking this into account, $-\Delta H_{ads}(HsO_4)$ of only 6 kJ/mol larger than $-\Delta H_{ads}(OsO_4)$ was calculated using Eq. (10) and the computed molecular properties (Table 8) [76], in excellent agreement with the experimental $-\Delta H_{ads}(HsO_4)$ and the observed $T_{ads}(HsO_4) > T_{ads}(OsO_4)$ [75] (Fig. 16). The obtained $\Delta H_{ads}(MO_4)$ show a reversal of the trend in group 8, the same as α and IP do, which is stipulated by the trend reversal in the energies of the $(n-1)$d AOs (see Fig. 1b in [76]). This example shows that for very similar species, a particularly

Table 8 Ionization potentials, IPs (in eV); polarizabilities, α (in a.u.); bond lengths, R_e (in Å); vibrational frequencies, ν_e, of the M-O bond (in cm^{-1}); and adsorption enthalpies, $-\Delta H^\circ_{ads}$ (in kJ/mol), on quartz for MO_4 (M = Ru, Os, and Hs)

Property	Method[a]	RuO_4	OsO_4	HsO_4
IP	Calc.	12.21	12.35	12.29
	Exp.	12.19	12.35	–
α	Calc.	58.07	55.38	65.99
	Exp.	58.64	55.13	–
R_e	Calc.	1.712	1.719	1.779
	Exp.	1.706	1.711	–
ν_e	Calc.	851	900	989
	Exp.	880	965	–
ΔH_{ads}	Calc.	41.0 ± 1	39.0 ± 1	45.4 ± 1
	Exp.[b]		39.0 ± 1	46.0 ± 2

[a]Calculations are from [76]; see also references for experimental values there
[b]Ref. [75]

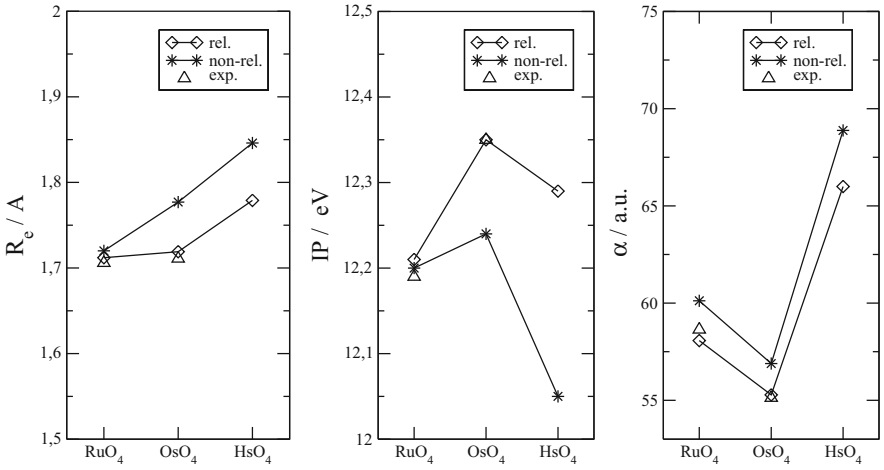


Fig. 15 Relativistic (rel.) and nonrelativistic (nonrel.) bond lengths, R_e; ionization potentials, IPs; and polarizabilities, α, of MO_4 (M = Ru, Os, and Hs) (Reproduced with permission from Ref. [76]. Copyright 2008 American Physical Society)

high accuracy of the calculations is required in order to "detect" tiny differences in their molecular and adsorption properties. A linear extrapolation of ΔH_{ads} in group 8 proved to be unreliable.

Relativistic effects were shown to have no influence on the trends in the molecular properties and $\Delta H_{ads}(MO_4)$ in group 8 (Fig. 15), because of identical trends in the energies and R_{max} of the charge density of the relativistic and nonrelativistic $(n-1)$d AOs responsible for bonding.

Group 9, 10, and 11. Mt and Ds have received little attention so far, because very short half-lives of their isotopes are not suitable for experimental investigations.

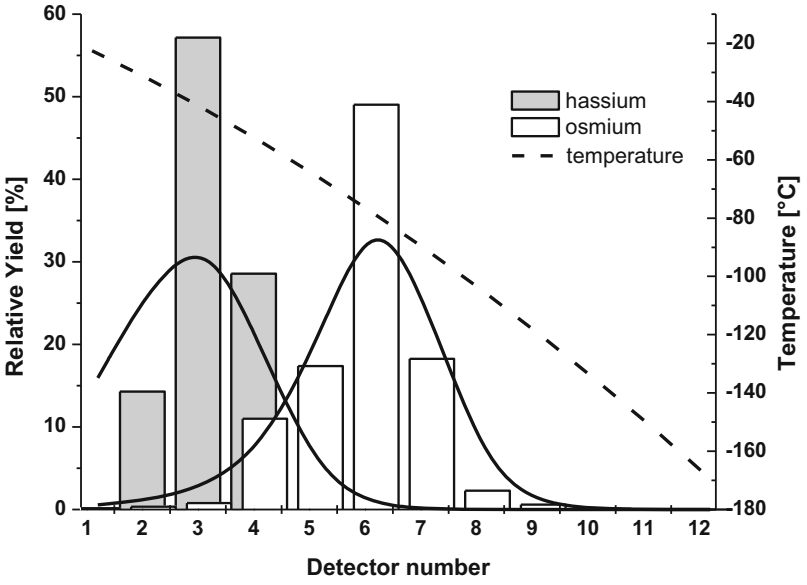


Fig. 16 Observed adsorption behavior of OsO_4 and HsO_4 in the gas-phase thermochromatography experiments (Reproduced with permission from Ref. [75]. Copyright 2002 Nature Publishing Group)

A few theoretical studies deal with DsF_6, DsX (X = H, C, and CO) of DsH_3 (see [9] for references).

Chemistry of Rg was, on the contrary, of much theoretical interest: unusual properties of its compounds were anticipated due to the strongest relativistic stabilization of the 7s AO in group 11. Particularly, the electronic structure of RgH, a sort of a test molecule like AuH, was of interest in probing the accuracy of various methods in predicting the trend from R_e(AuH) to R_e(RgH) (see Table 13 of [9]). The 4c-DFT and DHF CCSD(T) results were proven to be most accurate. The SR effects were shown to double D_e(RgH), though the SO effects on the Rg atom diminish it by 0.7 eV. The trend to an increase from D_e(AgH) to D_e(AuH) turns, therefore, out to be reversed from D_e(AuH) to D_e(RgH).

Group 12. Group-12 elements have a $(n-1)d^{10}ns^2$ closed-shell ground state. With increasing relativistic effects in this group, the elements become more inert. Thus, Hg is known to be a liquid. For Cn, the maximum of the relativistic stabilization of the 7s AO in the group (Fig. 2), as well as in the 7th row of the periodic table, was a reason to believe it to be noble gas-like: In 1975, Pitzer suggested that the very high valence-state excitation energy $6d^{10}7s^2 \rightarrow 6d^{10}7s7p_{1/2}$ of 8.6 eV will not be compensated by the energy gain of the chemical bond formation [77].

The idea of Cn having noble gas properties was tested by gas-phase thermochromatography experiments allowing for comparison of its volatility with that of Hg,

a lighter homologous d element, and Rn, an inert gas [78]. Gold was chosen as a surface of detectors of a chromatography column: it strongly adsorbs Hg, while very weakly Rn. The questions to the electronic structure theory were, therefore: Is Cn metallic in the solid state, or is it more like a solid noble gas? How volatile and reactive toward gold is the Cn atom in comparison with Hg and Rn?

Homonuclear dimers. Bonding in the solid state, i.e., cohesive energy, E_{coh}, of an element can be described in the first approximation by bonding in its homonuclear dimer, M_2. $D_e(Cn_2)$ had, therefore, to be accurately calculated. Moreover, Hg_2 and Cn_2 were of interest for theory in probing the accuracy of various methods in treating closed-shell interactions. Accordingly, the electronic structures of Hg_2 and Cn_2 were calculated using a variety of methods, such as X2C CCSD(T) [79], 4c-BDF, ECP CCSD(T), QR PP CCSD(T) [54], 4c-BDF [80], and 4c-DFT [81]. Some results are shown in Table 9 and Fig. 17. A more complete table is given in [9].

As expected, the DFT underestimates $D_e(M_2)$ (M = Hg and Cn) [54, 81], but nicely reproduces the trend to an increase in it of 0.04 eV from Hg_2 to Cn_2, also shown by the CCSD(T) calculations [54, 79] (Fig. 17). As the nature of this interaction is predominantly dispersive, such an increase is due to a smaller M-M distance caused by the smaller R_{max} of the 7s(Cn) AO in comparison with R_{max} of the 6s(Hg) AO. This is a first indication that the cohesive energy, $E_{\mathrm{coh}} \approx \Delta H_{\mathrm{sub}}$, of bulk Cn should be larger than E_{coh}(Hg).

Solid state. LDA DFT (nonrelativistic, scalar relativistic, SR, and 4c relativistic) band structure calculations were performed on the Cn solid state [83]. The results have shown that Cn prefers the *hcp* structure (as that of Zn and Cd) in the difference to Hg (*fcc*). Thus, it should differ from its lighter homolog Hg on a structural level and resemble the solid-state noble gases. A cohesive energy of 1.13 eV was obtained for Cn at the SR level of theory, which is larger than that of Hg (0.64 eV) and is an order of magnitude larger than those of the solid noble gases. It was concluded that Cn is not a metal, but rather a semiconductor with a band gap of at least 0.2 eV. In this sense, Cn resembles the group-12 metals more closely than it does the noble gases. This result is consistent with the larger $D_e(Cn_2)$ with respect to $D_e(Hg_2)$.

Table 9 Bond lengths, R_e (in Å), and dissociation energies, D_e (in eV), of Hg_2, Cn_2, and Fl_2

	Hg_2		Cn_2		Fl_2		
Method	R_e	D_e	R_e	D_e	R_e	D_e	Ref.
X2C CCSD(T)	3.744	0.050	3.461	0.084	3.547	0.117	[79]
4c-BDF (PBE)	3.904	0.025	3.363	0.075	3.490	0.12	[54,80]
RECP CCSD(T)	3.73	0.049	3.320	0.095	3.730	0.07	[54,80]
QR PP CCSD(T)	3.769	0.044	3.386	0.097	–	–	[54]
4c-DFT (B88/P86)	3.63	0.010	3.450	0.050	3.490	0.13	[81,82]
Exp.[a]	3.63	0.043	–	–	–	–	

[a]See the original publications for references

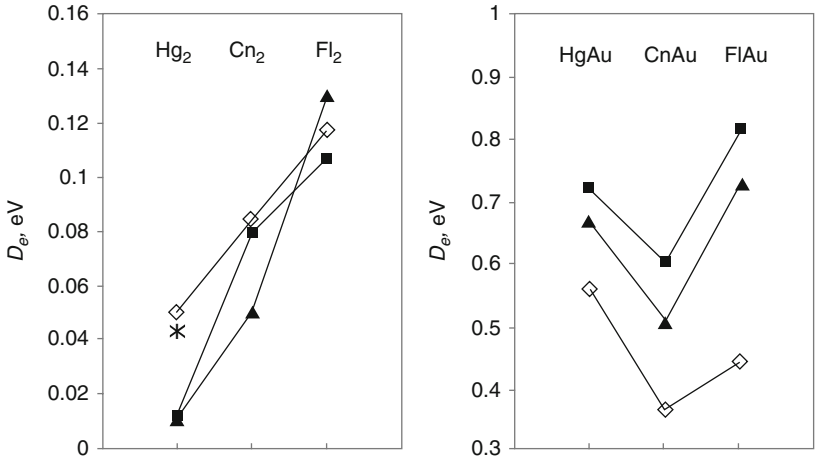


Fig. 17 Dissociation energies of M_2 and MAu, where M = Hg, Cn, and Fl: the (2c) BSS CCSD(T) values are *open rhomboids*, the (2c) BSS DFT (B88/P86) data are *filled squares* [79], and the 4c-DFT (B88/P86) data are *filled triangles* [81, 82]. The experimental $D_e(Hg_2)$ is shown with a star

Thus, the case of group-12 elements shows that the relativistic calculations are indispensable in order to predict the right trend in volatility, Cn < Hg, or ΔH_{sub}(Cn) > ΔH_{sub}(Hg), while an extrapolation of ΔH_{sub} in group 12 gives 22.2 kJ/mol for Cn, which is the lowest in group 12.

Interaction with metals. In order to predict volatility of Hg and Cn as adsorption enthalpy on gold, $\Delta H_{ads}{}^{Au}$(M) (and other noble metals), measured in the gas-phase experiments [78], 4c-DFT calculations were performed for MM′ (M = Hg, Cn; M′ = Ag, Au, Pt, Pd, and Cu) [84] and M-Au_n systems, where Au_n (n = 1 through 120) are clusters simulating a gold surface [85, 86]. It was shown that Cn interacts rather well with the noble metals. In CnAu, the σ-bond formation takes place between the doubly occupied 7s(Cn) AO and the singly occupied 6s(Au) AO (Fig. 18). D_e(CnAu) is, however, smaller than D_e(HgAu) due to the more stabilized 7s(Cn) AO than 6s(Hg) AO. Among different metals M, bonding of Cn with Pd should be the strongest, while with Ag the weakest [84].

Since the structure of the real gold surface is unknown, two types of ideal surfaces were considered: Au(100) and Au(111) [85, 86]. For the Au(111) surface that is more realistic, the convergence in E_b(M-Au_n) (M = Hg and Cn) with the cluster size was reached for n = 95 for the top, n = 94 for the bridge, and n = 120 for the hollow-1 and n = 107 for the hollow-2 positions. For Hg, the bridge position was found to be preferential for both types of surfaces, while for Cn, the hollow 2. E_b(Cn-Au_n) of 0.46 eV on the Au(111) surface was given then as a final prediction [86] (Table 10 and Fig. 19).

The measurements of volatility of Cn have shown that Cn adsorbs in the chromatography column on gold at T_{ads} around −20 °C, which is much lower than

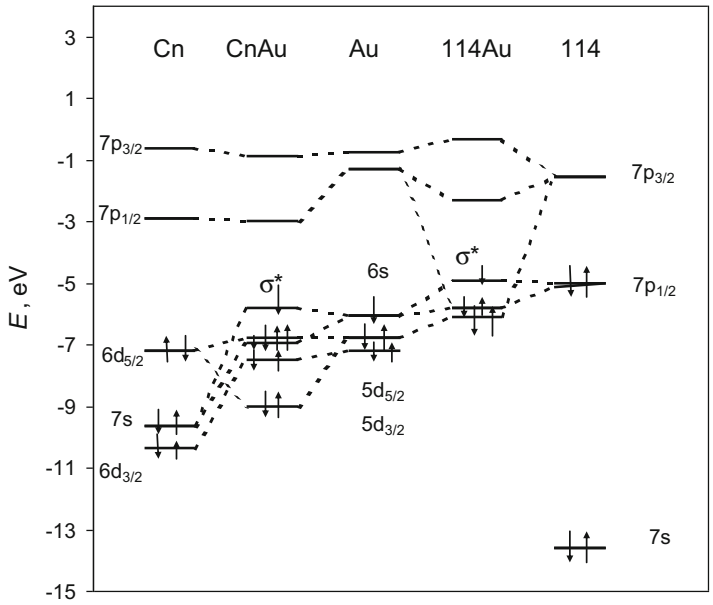


Fig. 18 Chemical bond formation in CnAu and FlAu (Reproduced with permission from Ref. [11]. Copyright 2011 Oldenbourg Wissenschaftsverlag GmbH)

Table 10 The Cn-Au$_n$ and Fl-Au$_n$ binding energies and $-\Delta H^\circ_{ads}$(M) on gold (M = Cn and Fl) (in eV)[a]

Method	n/surface	Cn	Fl	Ref.
4c-DFT	1	0.51	0.73	[86]
2c-DFT	1	0.47	0.72	[87]
SO DFT	3	0.47	0.77	[87]
2c-DFT	26^{br}/Au(100)	0.33	0.55	[86]
2c-DFT	37^{holl-2}/Au(111)	–	0.49	[88]
4c-DFT	95^{top}/Au(111)	0.30	0.47	[86]
4c-DFT	94^{br}/Au(111)	0.42	**0.71**	[86]
4c-DFT	107^{holl-2}/Au(111)	**0.46**	0.59	[86]
Exp.	$-\Delta H^\circ_{ads}$(M)/Au[b]	$0.54^{+0.2}_{-0.04}$	$0.34^{+0.5}_{-0.1}$	[78, 89]
			> 0.50	[90]

[a]Most stable states are in bold
[b]The structure of the gold surface is unknown

T_{ads} of Hg (Hg adsorbs right at the beginning of the column with 35 °C), but higher than T_{ads}(Rn) of −180 °C [78]. The obtained $-\Delta H_{ads}{}^{Au}$(Cn) $= 0.54^{+0.2}_{-0.01}$ eV (52^{+20}_{-4} kJ/mol) turned out to be in good agreement with the theoretical value [86]. Such a relatively high $-\Delta H_{ads}{}^{Au}$(Cn) was evident of the metal-metal bond formation, a behavior typical of group-12 lighter homologs.

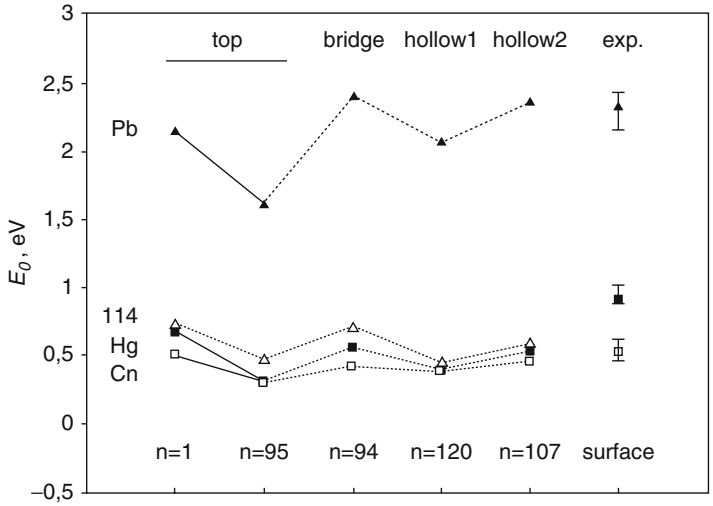


Fig. 19 The 4c-DFT calculated binding energies of Pb, Hg, Cn, and Fl with Au_n simulating a Au(111) surface and experimental $-\Delta H_{ads}$ of Pb, Hg, and Cn on gold (Reproduced with permission from Ref. [86]. Copyright 2009 American Institute of Physics)

Works on the RECP and 2c-DFT calculations for Hg and Cn interacting with different gold clusters arrived at the same conclusion: E_b(Cn-Au_n) is about 0.2 eV smaller than E_b(Hg-Au_n) (Table 10) [87]. (The data for FlAu_n are also given in Table 10 [88–90], but discussed below.) In [47], the influence of relativistic effects on E_b(M-Au_n) (M = Hg and Cn) was investigated. Relativistic effects were shown to increase E_b(M-Au_n) of both the Hg- and Cn-containing species and to be responsible for E_b(Hg-Au_n) > E_b(Cn-Au_n).

In summary, both the theoretical and experimental studies show that Cn forms a rather strong bond, however weaker than that of Hg, with metals. Thus, it behaves like a d element upon adsorption, but not like an inert gas. In this way, its position in group 12 of the periodic table has been confirmed. The obtained - $\Delta H_{ads}{}^{Au}$(Cn) < $-\Delta H_{ads}{}^{Au}$(Hg) *do not correlate* with ΔH_{sub}(Cn) > ΔH_{sub}(Hg) due to the different types of bonding in these two cases. The example of group-12 elements shows how important are theoretical calculations in order to distinguish between the nature of the studied processes and trends in the group.

Some other compounds of Cn, hydrides, and fluorides were also considered theoretically. A review of those works can be found elsewhere [5, 8–11].

Groups 13 through 18. In the 7p elements, the $7s^2$ electrons are bound more tightly than the $6s^2$ ones in the 6p homologs, so that they do not take part in the chemical bond formation. Also, a large SO splitting of the 7p shell into the *nlj* subshells should result in differences in the chemical bonding in comparison with the homologs having (almost) a complete *nl* shell. Thus, these elements are expected

to be more volatile than their lighter homologs, as their ΔH_{sub} obtained via linear extrapolations in the groups indicate [91].

Experimentally, volatility of the 7p elements at the beginning of the series, i.e., those with sufficiently long half-lives, is supposed to be studied using gas-phase chromatography techniques with gold- or SiO_2-plated detectors. First results for adsorption on gold with the use of thermochromatography have been published for ^{289}Fl ($t_{1/2} = 2.1$ s) and ^{288}Fl ($t_{1/2} = 0.7$ s) [89, 90]. The next element to be studied is $^{284}113$ with $t_{1/2} \approx 0.9$ s.

Homonuclear dimers. Keeping again in mind that bonding in M_2 is related to bonding in the solid state, $D_e(M_2)$ were calculated for the entire series of the 6p and 7p elements using the 4c-DFT method [92, 93]. (A few dimers were also calculated in [80]). All the dimers of group-13 through group-17 7p elements were shown to be, indeed, weaker bound than their 6p homologs, with the difference in $D_e(M_2)$ between them decreasing with the group number and a final reversal of the trend in group 18 (Fig. 20). The reason for that is very large SO effects on the 7p AOs preventing them from making a full σ-bond: in $(113)_2$ and Fl_2, bonding is preferentially made by the $7p_{1/2}$ AOs, while in the heavier elements, mainly by the $7p_{3/2}$AOs.

An interesting observation was the shift of the $D_e(M_2)$ maximum in the seventh row toward group 16 with respect to group 15 in the sixth and upper rows (Fig. 20). This is also a strong manifestation of the relativistic effects. As a result of the stabilization of the $7p_{1/2}$ AOs, the system of the highest bonding-antibonding MOs in M_2 consists of only four MOs composed of the $7p_{3/2}$ AOs, so that the half-filled shell (with four electrons) falls on $(116)_2$. In contrast, the $6p_{1/2}$ and $6p_{3/2}$ AOs are not so well separated energetically from each other and form a set of six highest bonding-antibonding MOs, so that the half-filled shell (with six electrons) falls on Bi_2 (see Fig. 4 in [92]).

Flerovium. Due to the very large SO splitting of the 7p AOs (Fig. 2), Fl has a quasi-closed-shell ground state, $7s^27p_{1/2}^2$. The argument of Pitzer, similar to that used for Cn, which due to the $7p_{1/2}$(Fl) AO stabilization, the $7p_{1/2}^2 \rightarrow 7p^2$ promotion energy to the metal valence state will not be compensated by the metal bond formation, led to the conclusion that this element should be a relatively inert gas or a volatile liquid bound by van der Waals forces [77]. To test this assumption at the MO level of theory, electronic structures of Fl_2 (and Pb_2 for comparison) were calculated using DFT and CC methods [80, 92]. The calculations agree on the fact that bonding in Fl_2 is stronger than a typical van der Waals one. It is stronger than that in Cn_2, but much weaker than that in Pb_2 (Table 9 and Fig. 20). A Mulliken population analysis indicates that both the $7p_{1/2}$ and $7p_{3/2}$ AOs of Fl take part in the chemical bond formation. SO effects were shown to decrease D_e, but increase R_e in both Pb_2 and Fl_2 [80]. The DFT results [81, 82] perfectly reproduce the trend to an increase in D_e from Hg_2 to Cn_2 and further to Fl_2 in agreement with the CC calculations [79], where dispersion interactions are more fundamentally taken into account (Table 9 and Fig. 17). Also, the 4c-DFT and CCSD(T) $D_e(Fl_2)$ nicely agree with each other

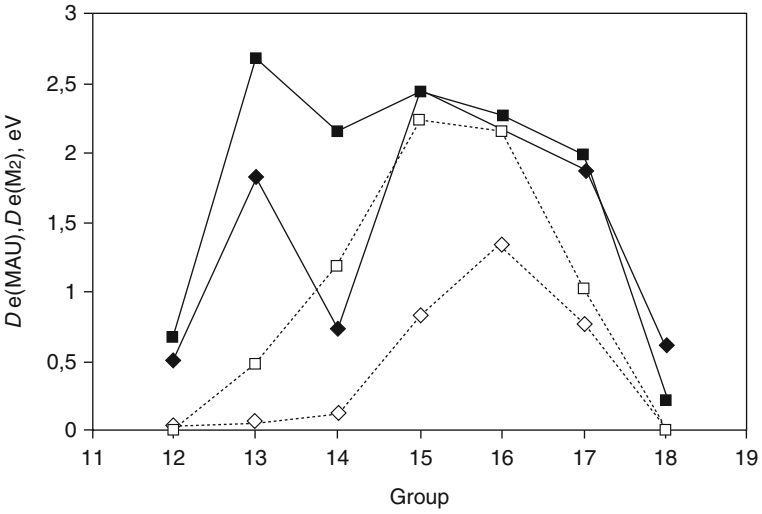


Fig. 20 Calculated dissociation energies of MAu and M_2 (M are elements Hg/Cn through Rn/118). *Filled* and *open squares* are D_e(MAu) and $D_e(M_2)$ of the 6p elements, respectively, while *filled* and *open rhomboids* are D_e(MAu) and $D_e(M_2)$ of the 7p elements, respectively (Reprinted with permission from Ref. [93]. Copyright 2010 American Institute of Physics)

confirming that the Fl-Fl interactions are of no van der Waals nature. This is a clear indication that the Fl-Fl bonding should be stronger than the Cn-Cn one, also in the bulk.

In the difference to the other 7p elements, the bond in $(118)_2$ should be stronger than that of the lighter group-18 homologs due to its largest $\alpha(118)$ in the group [92].

Sublimation enthalpies. The $D_e(M_2)$ of the np elements were shown [92] to nicely correlate with $\Delta H_{\rm sub} = \Delta H^{\circ}_f(g)$ of bulk of these elements [59]. Obtained on the basis of this correlation, $\Delta H_{\rm sub}$ of the 7p elements are smaller than those of the lighter homologs and agree rather well with $\Delta H_{\rm sub}$ predicted via a linear extrapolation in groups 13–17 [91]. Thus, elements 113 through 117 should, indeed, be more volatile than their lighter homologs.

In [94], $E_{\rm coh}$(Fl) of 0.5 eV was predicted from SR and SO DFT (PW91) periodic calculations. The resulting value is in good agreement with $\Delta H_{\rm sub}$ obtained from a correlation with $D_e(M_2)$ in group 14 [92]. SO effects were shown to lower $E_{\rm coh}$ and to lead to structural phase transitions for the solid Fl having the *hcp* structure in contrast to the *fcc* one for Pb.

Interaction with metals and adsorption on gold. In order to predict $\Delta H_{\rm ads}$ of the 7p elements on gold, 4c-DFT calculations for the MAu dimers were performed in [93]. Obtained D_e(MAu) are shown in Fig. 20 in comparison with $D_e(M_2)$. One can see that in group 13 and 14, D_e(MAu) of the 7p elements are smaller than D_e of the 6p homologs, which is explained by the relativistic stabilization of the $7p_{1/2}$ AOs. On

the contrary, in groups 15 through 17, they are about the same. This is in contrast to the trends in $D_e(M_2)$, where $D_e(Bi_2) >> D_e[(115)_2]$, $D_e(Po_2) >> D_e[(116)_2]$, and $D_e(At_2) > D_e[(117)_2]$. The relatively strong M-Au bonding of elements 115 through 117 with gold is explained by the relativistic destabilization of the $7p_{3/2}$(M) AO fitting nicely to the 6s(Au) AO, thus making – together with the $7p_{1/2}$(M) AO – a full σ-bond in the difference to M_2, where only the $7p_{3/2}$(M) AO are involved in bonding. In group 18, a reversal takes place, so that $D_e(118Au) > D_e(RnAu)$, in agreement with the trend in $D_e(M_2)$. This is due to the relativistically destabilized $7p_{3/2}(118)$ AO better fitting to the 5d and 6s AOs of Au than the $6p_{3/2}$(Rn) AOs.

The 4c-DFT calculations were also performed for group-14 intermetallic dimers MM', where M' are group-10 and group-11 elements [82]. The Fl-Pt bonding was found to be the strongest, while the Fl-Ag and Fl-Ni, the weakest.

Using D_e(MAu), $-\Delta H_{ads}$ of element 113 and Fl on gold of 159 and 92 kJ/mol, respectively, were estimated via a correlation between these quantities in groups 13 and 14 [93, 95]. In groups 15 through 17, no correlation is observed between these quantities, because D_e(M-Au) does not decrease linearly with Z, while D_e(M-M) does [93]. Thus, the case of the 7p elements with large SO effects shows that linear extrapolations of properties, such as $\Delta H_{ads}{}^{Au}$(M), from the lighter homologs in the groups might lead to erroneous predictions.

The 2c-DFT and 1c-CCSD(T) calculations were performed for MAu_n, where M = Tl and element 113 ($n = 1, 2, 3$, and 58) [96], with the gold clusters simulating Au(100) and Au(111) surfaces. The difference in E_b(M-Au_n) between Tl and element 113 was found to stay within $\pm$15 kJ/mol of 82 kJ/mol obtained for D_e(MAu) [93]. Thus, taking into account $-\Delta H_{ads}$(Tl) of 240 kJ/mol [97], $-\Delta H_{ads}$(113) can presently be given as 159 $\pm$15 kJ/mol.

Extended 4c-DFT calculations were also performed for M = Pb and Fl (along with Hg and Cn) interacting with large Au_n clusters ($n = 1$ through 120) simulating the Au(111) surface [86]. Both Pb and Fl were found to prefer the bridge adsorption position, where the convergence in E_b(M-Au_n) with the cluster size was reached for $n = 94$ (Table 10 and Fig. 19). The calculated E_b turned out to be in very good agreement with the experimental $-\Delta H_{ads}$ of Pb and Cn on gold [78] indicating that the Au(111) surface is obviously the proper one. Taking into account that Hg dissolves in gold, the trend in $-\Delta H_{ads}{}^{Au}$(M) was predicted as Cn < Fl < Hg $\ll$ Pb [86].

Particularly interesting is the comparison of Fl with Cn, both revealing relatively weak metal-surface interaction. Both the 4c-DFT [86, 93] and BSS CCSD(T) [79] calculations for the gold dimers, as well as for large metal-gold cluster systems [86–88], have shown Fl to be stronger bound with Au than Cn (Table 10 and Fig. 19). The reason for that is the following: in FlAu – even though both FlAu and CnAu are open-shell systems with one antibonding σ^* electron – electron density is donated from the lying higher in energy $7p_{1/2}$(Fl) AO to the 6s(Au) AO, while in CnAu, some excitation energy is needed to transfer some electron density from the closed $7s^2$(Cn) shell to the 6s(Au) AO (Fig. 18).

In contrast to the theoretical predictions, the first thermochromatography experiment with three events of Fl has shown that it adsorbs on the surface of chro-

matography column with gold detectors at temperatures about those of Rn and $-\Delta H_{ads}{}^{Au}$(Fl) of 0.34 $^{+0.54}_{-0.11}$ eV has been given on their basis [89], smaller than $-\Delta H_{ads}{}^{Au}$(Cn) of 0.54 $^{+0.02}_{-0.04}$ eV (Table 10). Such a small $-\Delta H_{ads}{}^{Au}$(Fl) was interpreted as an indication of "the formation of a weak physisorption bond between atomic Fl and a gold surface." According to these results, Fl should have been chemically more inert than Cn. The second similar experiment on volatility of Fl, conducted at a lower background of interfering products, registered two events of this element adsorbed on gold at room temperature [90]. Such a relatively high T_{ads} and estimated $-\Delta H_{ads}{}^{Au}$(Fl) > 48 kJ/mol support the theoretical conclusion that Fl should form a chemical bond with Au. Further experiments are underway to collect better statistics of the Fl events in order to have a more accurate ΔH_{ads}(Fl).

Other compounds. Other than elements 114 and 118, elements 113 and 115 through 117 will not be stable in the gas phase in the elemental state. Thus, element 113 should form hydroxides in the presence of traces of water in the experimental setup. Relativistic 4c-DFT calculations have shown the 113-OH bond to be rather strong (2.42 eV), but weaker than the Tl-OH one (3.68 eV) [95]. Element 113 may then adsorb on gold in the form of the hydroxide. $-\Delta H_{ads}{}^{Au}$(113OH) is expected to be lower than $-\Delta H_{ads}{}^{Au}$(113).

There are numerous relativistic – DCB, RECP, and 4c-DFT – calculations of other compounds of group-13 elements, such as MX and MX_3 (X = H, F, Cl, Br, and I), or (113)(117). A very large bond contraction is found in 113X due to the $7p_{1/2}$(113) AO contraction, as in none of the other compounds (see [9] as a review). The electronic structures of FlX (X = F, Cl, Br, I, O, O_2) were calculated using the 2c-RECP CCSD(T), 2c-DFT SO ZORA, and 4c-BDF methods [80]. In contrast to PbO_2 (D_e = 5.60 eV), FlO_2 (D_e = 1.64 eV) was predicted to be thermodynamically unstable with respect to the decomposition into the metal atom and O_2.

Chemistry of elements 115, 116, and 117 received little attention so far. Chemistry of element 118, on the contrary, was of much theoretical interest. It should be unusual due to the huge SO effects (11.8 eV) preventing the $7p^2_{1/2}$ pair from the participation in the chemical bond. Thus, a different geometry of $118F_4(T_d)$ than that of RnF_4 (D_{4h}) was predicted due to availability of only $7p^4_{3/2}$ electrons of element 118 for bonding [98]. (See also [7–11] for reviews.)

Summary. Predicted trends in volatility of group-4 through group-8, group-12, and group-14 species in comparison with observations are given in Table 11. They show largely the synergy between theory and experiment. Observed unusual trends in volatility of group-4 and group-5 halides, as well as of Fl with respect to Cn, belong to the few unresolved cases that may require further research efforts.

Elements 119 and 120. Elements 119 and 120 are the next elements awaiting discovery. Their properties should be defined by the 8s and $8s^2$ ground-state configurations, respectively. Volatility of atoms of elements 119 and 120 might be studied in the long term using an advanced (vacuum) chromatography technique designed for extremely short, presumably sub-millisecond, half-lives of their isotopes.

Table 11 Trends in volatility of group-4–8, group-12, and group-14 species

Group	Volatile species	Theoretically predicted	Ref.	Experimentally observed	Ref.[a]
4	ML_4 (L = Cl, Br)	Zr < Hf < Rf	[64, 65]	Hf < Rf ≈ Zr	See [5]
5	ML_5(L = Cl, Br)	Nb < Ta < Db	[66]	Ta < Db ≈ Nb	See [5]
	MOL_3 (L = Cl, Br)	Nb > Ta > Db	[66]	Nb > Ta > Db	See [5]
6	MO_2Cl_2	Mo > W > Sg	[68]	Mo > W > Sg	[67]
	$M(CO)_6$	Mo ≈ W ≈ Sg	[72]	Mo ≈ W ≈ Sg	[70]
7	MO_3Cl	Tc > Re > Bh	[74]	Tc > Re > Bh	[73]
8	MO_4	Ru < Os > Hs	[76]	Os > Hs	[75]
12	M	Hg < Cn	[86]	Fl > Cn	[78, 89]
14	M	Fl < Cn	[86]	Fl ≤ Cn	[90]

[a]For most experimental studies, see references in [5]

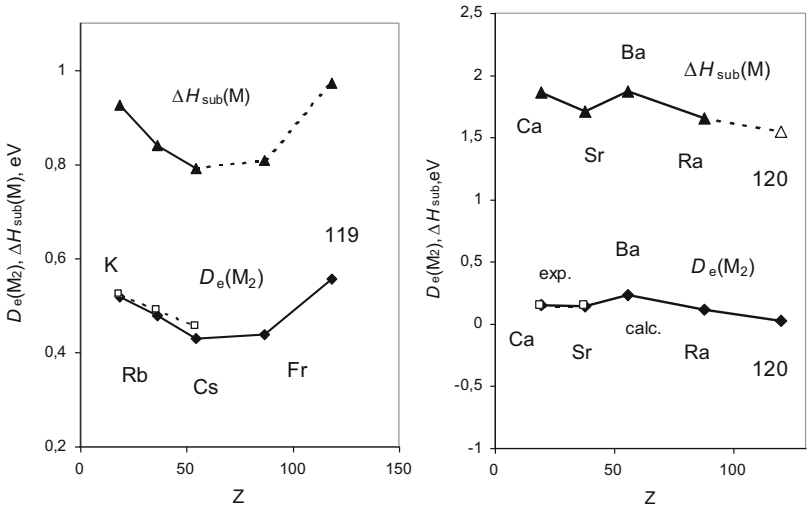


Fig. 21 Dissociation energies, D_e, of group-1 and group-2 M_2 (*filled rhomboids* are 4c-DFT calculations, *open squares* – experiment), as well as sublimation enthalpies, ΔH_{sub}(M) (*filled triangles* connected by solid lines are experiment, those connected by dashed lines – estimates) (Reproduced with permission from Refs. [99] and [100]. Copyright 2012 Elsevier and 2012 American Institute of Physics, respectively)

Homonuclear dimers and sublimation enthalpies. To estimate ΔH_{sub} of elements 119 and 120, the 4c-DFT calculations were performed for group-1 and group-2 homonuclear dimers [99, 100]. The obtained $D_e(M_2)$ show a reversal of trends at Cs in group 1 and at Ba in group 2, though in the opposite direction: an increase in group 1 from Cs on and a decrease in group 2 from Ba on (Fig. 21). The reason for these different trends is the different types of the M-M binding in these two cases: a covalent ns(M)-ns(M) one in group 1, while a van der Waals ns^2(M)-ns^2(M) one in group 2. Accordingly, $R_e(M_2)$ decreases in group 1 from Cs toward element 119, while it steadily increases in group 2.

A linear correlation between $D_e(M_2)$ and $\Delta H_{sub}(M)$ was established in these groups. On its basis, ΔH_{sub} of element 119 of 94 kJ/mol and of element 120 of 150 kJ/mol were obtained (Fig. 21) [99, 100]. According to these data, the element 119 metal should be as strongly bound as K, while the element 120 metal should be most weakly bound in group 2, though stronger than the element 119 one. A reversal of the M-M binding trends in these groups is a result of the E(ns) AO trend reversal from the 6th row on.

Intermetallic dimers and adsorption on noble metals. In order to predict $\Delta H_{ads}(M)$ of elements 119 and 120 on noble metals, electronic structures of MAu, where M are group-1 and group-2 elements, were calculated using the 4c-DFT method [99, 100]. The 2c-DFT calculations were also performed for the 120-Au_n clusters [101]. Elements 119 and 120 were shown to strongly interact with gold though weaker than the homologs. The D_e(MAu) values reveal a reversal of the increasing trend at Cs and Ba in group 1 and 2, respectively (Fig. 22), in accordance with the trend reversal in E(ns) AO. They were also shown to correlate with the semiempirical $-\Delta H_{ads}(M)$ [102] of K through Cs and Ca through Ba on gold and other metals, respectively. On the basis of these correlations, $-\Delta H_{ads}$ of element 119 of 106 kJ/mol and of element 120 of 353.8 kJ/mol on gold, as well as on other metals, were determined (Fig. 22). Finally, no proportionality was found between the $\Delta H_{sub}(M)$ and $-\Delta H_{ads}{}^{Au}(M)$ in group 1, because these quantities change in opposite directions with Z in these groups. In group 2, there is a correlation between $\Delta H_{sub}(M)$ and $-\Delta H_{ads}{}^{Au}(M)$.

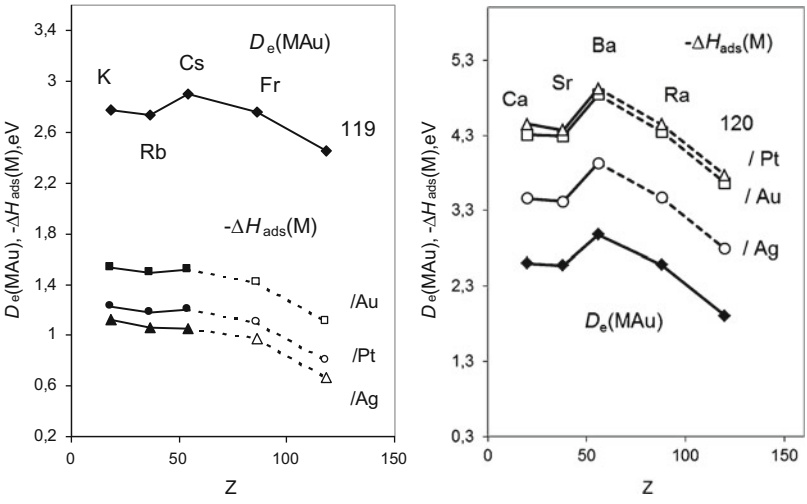


Fig. 22 The 4c-DFT dissociation energies, D_e(MAu), and adsorption enthalpies $-\Delta H_{ads}(M)$ on Au, Pt, and Ag of group-1 elements and group-2 elements (*filled symbols* are semiempirical calculations, while open ones are obtained via correlations with D_e(MAu) (Reprinted with permission from Refs. [99] and [100]. Copyright 2012 Elsevier and 2013 American Institute of Physics, respectively)

The predicted $\Delta H_{ads}{}^{Au}(M)$ mean, however, that very high temperatures must be applied to establish an equilibrium T_{ads}. This is presently not feasible, because the detectors cannot be heated above $\sim 35\,°C$.

Few other simple compounds of elements 119 and 120 were calculated, whose description can be found elsewhere [5, 8–11].

Elements with Z > 120. The chemistry of elements heavier than $Z = 120$ rests on a purely theoretical basis. Due to even stronger relativistic effects, as well as the existence of a plenty of open shells and their mixing, it will be much more different than anything known before. Very few molecular calculations exist in this SHE domain. The latest considerations of the chemistry of some of these elements can be found in [103, 104]. In this area, some unexpected oxidation states and coordination number like, e.g., $(E148)O_6$ or $(E158)X_8$ (X = halogen), could be reached.

Aqueous Chemistry

- Redox potentials and stability of oxidation states

Knowledge of redox potentials, E°, is very important for chemical separation of SHE. Early predictions of oxidation states and E° of the heaviest elements based on atomic DF and DS calculations and a Born-Haber cycle are summarized in [6]. Later, E° of Rf, Db, and Sg were estimated using results of atomic MCDF calculations of multiple IPs and known experimental redox potentials in group 4, 5, and 6 [48, 49]; see also [5, 8, 9] for reviews. The following trends were established for SHE: the stability of the maximum oxidation state increases within group 4 through 6 as a result of the proximity of the relativistically destabilized 6d AOs, while that of lower oxidation states decreases. Along the 7th period, the stability of the maximum oxidation state decreases: $E^\circ(Lr^{3+}/Lr^{2+}) < E^\circ(Rf^{IV}/Rf^{3+}) < E^\circ(Db^{V}/Db^{IV}) < E^\circ(Sg^{VI}/Sg^{V})$. A similar trend is observed for $E^\circ(M^{Zmax}/M)$. The increasing stability of the maximum oxidation state in group 4, 5, and 6 was shown to be a relativistic effect due to the proximity of the relativistic energy levels, while the 3+ state of Db and the 4+ state of Sg should be unstable. A summary of the redox potentials is given in Table 3.22 of [9].

- Complex formation and extraction by liquid chromatography

A number of theoretical works were devoted to predictions of complex formation, hydrolysis, and extraction behavior of Rf, Db, and Sg, in aqueous acid solutions (Eq. 5). Complex formation of Hs has also been studied as a specific case. (See [5, 8, 9] for the reviews.) The predictions were made with the use of a model that treats the formation energy of a compound as a sum of ionic and covalent constituents. The latter were obtained via 4c-DFT electronic structure calculations of the systems of interest [105, 106]. Those studies demonstrated that even though

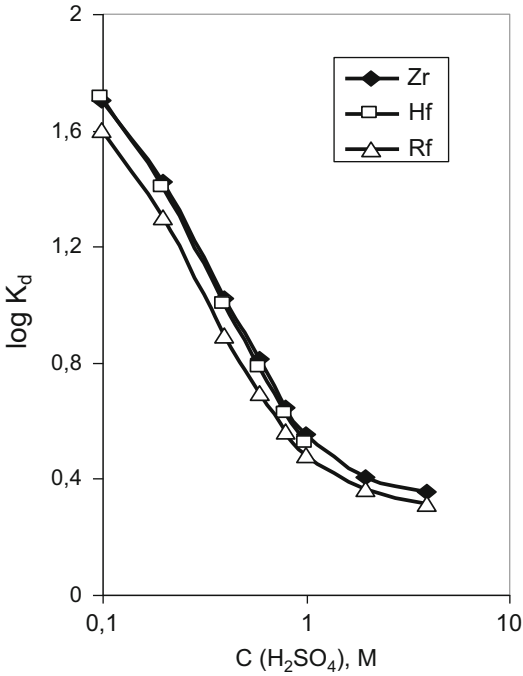


Fig. 23 Predicted $\log K_d$ for the extraction of Hf and Rf by amines with respect to the measured ones for Zr (Reprinted with permission from Ref. [107]. Copyright 2006 Oldenbourg Wissenschaftsverlag GmbH)

the heaviest elements are homologs of their lighter congeners in the chemical groups, trends are not necessarily continued with them (see, e.g., predictions of an unexpected trend in the distribution coefficients, K_d, of the group-5 complexes by extraction from the HCl solutions into amines [105], confirmed by experiments [19]). The calculations have also shown that the theory of hydrolysis based on the ratio of the cation size to its charge does not explain, e.g., the difference between Nb and Ta or Mo and W. Only by performing relativistic calculations of the real chemical equilibria can the complex formation or hydrolysis be correctly predicted.

As an example, predictions of $\log K_d$ for the extraction of Zr, Hf, and Rf from the H_2SO_4 solutions by amines are shown in Fig. 23. For that purpose, the formation energies of the $M(SO_4)_2(H_2O)_4$, $M(SO_4)_3(H_2O)_2^{2-}$, and $M(SO_4)_4^{4-}$ (M = Zr, Hf, and Rf) complexes were calculated using the 4c-DFT method [107]. According to the results, the following trend in the complex formation was predicted: Zr > Hf ≫ Rf. Aqueous chemistry extraction experiments [108] confirmed the theoretically predicted trend and have given the K_d(Rf) values close to the predicted ones.

A summary of the predicted trends in hydrolysis, complex formation, and extraction of the group-4 through group-6 elements including the heaviest is given in [5,8,9]. As one can see there, most of the predictions have been confirmed by the experiments, while some of them, e.g., predictions for Sg in HF solutions [109], are still awaiting the confirmation.

Summary

Theoretical studies on the electronic structures and properties of SHE have reached in the recent years remarkable achievements. With many of them linked to the experimental research, these investigations have contributed to better understanding of chemistry of these exotic elements, as well as the role and magnitude of relativistic effects. They helped predict the outcome of sophisticated and demanding experiments with single atoms and interpret their results.

Atomic structures were accurately predicted at the DCB CCFS level of theory for elements up to Z = 122. For the heavier elements, there is still some uncertainty in the ground states, so that the structure of the 8th row of the periodic table is still being discussed and debated.

Molecular calculations were performed for Rf through Z = 118 and a few heavier elements using a variety of relativistic methods, from DF CC ab initio to quasi-relativistic schemes. Most valuable information about properties of chemically interesting compounds (complex molecules, clusters, and solid state) was obtained with the use of the 4c-/2c-DFT and RECP/PP CC methods. They proved to be complimentary both conceptually and quantitatively, and their combination is the best way to investigate properties of the heaviest elements.

It was shown that elements Rf through 118 are homologs of their lighter congeners in the chemical groups, though their properties may differ due to very large relativistic effects. Straightforward extrapolations in the chemical groups can, therefore, result in erroneous predictions. For even heavier elements, properties should be even more unusual due to the mixing of many electronic states.

In the future, theoretical chemistry will still have a number of exciting tasks with respect to the new systems under investigation. Some further methodical developments in the relativistic quantum theory, like, e.g., inclusion of the QED effects on the SCF basis in molecular calculations, will be needed to achieve a required accuracy of the calculations.

Acknowledgements The author thanks her former collaborators J. Anton, T. Bastug, E. Eliav, U. Kaldor, and A. Borschevsky for the fruitful joint work. She also appreciates valuable discussions of the experimental results with her colleagues A. Yakushev, J.V. Kratz, Ch. E. Düllmann, and A. Türler.

References

1. Schädel M, Shaughnessy D (eds) (2014) The chemistry of the superheavy elements, 2nd edn. Springer, Heidelberg
2. Moody KJ (2014) Synthesis of superheavy elements. In: Schädel M, Shaughnessy D (eds) The chemistry of the superheavy elements, 2nd edn. Springer, Heidelberg, pp 1–82
3. Oganessian YuTs (2011) Synthesis of the heaviest elements in ^{48}Ca-induced reactions. Radiochim Acta 99:429–440
4. http://www.IUPAC.org: Discovery and assignment of elements with atomic numbers 113, 115, 117 and 118, 30 Dec 2015; Update 21 Jan 2016: technical reports available

5. Türler A, Pershina V (2013) Advances in the production and chemistry of the heaviest elements. Chem Rev 113:1237–1312
6. Fricke B (1975) Superheavy elements. A prediction of their chemical and physical properties. Struct Bond 21:89–144
7. Seaborg GT (1996) Evolution of the modern periodic table. J Chem Soc Dalton Trans 3899–3907
8. Hoffman DC, Lee DM, Pershina V (2006) Transactinide elements and future elements. In: Morss LR, Edelstein NM, Fuger J (ed) The chemistry of the actinide and transactinide elements, 3rd edn. Springer, Dordrecht, Ch 14, pp 1652–1752
9. Pershina V (2014) Theoretical chemistry of the heaviest elements. In: Schädel M, Shaughnessy D (eds) The chemistry of the superheavy elements, 2nd edn. Springer, Heidelberg, pp 135–240
10. Schwerdtfeger P, Seth M (1998) Relativistic effects on the superheavy elements. In: von Rague Schleyer P (ed) Encyclopedia on calculational chemistry, vol 4. Wiley, New York, pp 2480–2499
11. Pershina V (2011) Relativistic electronic structure studies on the heaviest elements. Radiochim Acta 99:459–476
12. Desclaux JP (1973) Relativistic Dirac-Fock expectation values for atoms with Z=1 to Z=120. At Data Nucl Data Tables 12:311–386
13. Eliav E, Kaldor U (2002) Four-component electronic structure methods. In: Schwerdtfeger P (ed) Relativistic electronic structure theory, Parts I and 2. Elsevier, Amsterdam, Part II, pp 279–350
14. Pyykkö P, Tokman M, Labzowsky LN (1998) Estimated valence-level Lamb shifts for group 1 and group 11 metal atoms. Phys Rev A 57:R689–692
15. Goidenko I, Labsowsky L, Eliav E, Kaldor U, Pyykkö P (2003) QED corrections to the binding energy of the eka-radon (Z=118) negative ion. Phys Rev A 67:020102(R)
16. Thierfelder C, Schwerdtfeger P (2010) Quantum electrodynamic corrections for the valence shell in heavy many-electron atoms. Phys Rev A 82:062503
17. Zvara I (2008) The inorganic radiochemistry of heavy elements. Springer, Berlin
18. Gäggeler HW, Türler A (2014) Gas-phase chemistry of superheavy elements. In: Schädel M, Shaughnessy D (eds) The chemistry of the superheavy elements, 2nd edn. Springer, Heidelberg, pp 415–484; Türler A, Gregorich KE (2003) Experimental techniques. *Ibid*: pp 261–308; also In: Schädel M (ed) The chemistry of superheavy elements, 1st edn. Kluwer Academic Publishers, Dordrecht, p 318
19. Kratz JV, Nagame J (2014) Liquid-phase chemistry of superheavy elements. In: Schädel M, Shaughnessy D (eds) The chemistry of the superheavy elements, 2nd edn. Springer, Heidelberg, pp 309–374
20. Schwerdtfeger P (ed) (2002) Relativistic electronic structure theory. Parts I and 2. Elsevier, Amsterdam
21. Barysz M, Ishikawa Y (ed) Relativistic methods for chemists. Springer, Dordrecht (2010)
22. Sucher J (1980) Foundation of the relativistic theory of many-electron atoms. Phys Rev A 22:348–362
23. Liu W (2014) Advances in relativistic molecular quantum mechanics. Phys Rep 537:59–89
24. Visscher L (2002) Post-Dirac-Fock methods. In: Schwerdtfeger P (ed) Relativistic electronic structure theory, Part I. Elsevier, Amsterdam, pp 291–331
25. Saue T (2002) Post Dirac-Fock-methods-properties. In: Schwerdtfeger P (ed) Relativistic electronic structure theory, Part I. Elsevier, Amsterdam, pp 332–397
26. DIRAC, a relativistic ab initio electronic structure program, written by Jensen AaHJ, Saue T, Visscher L et al. http://dirac.chem.sdu.dk
27. Douglas M, Kroll NM (1974) Quantum electrodynamical corrections to the fine structure of hellium. Ann Phys 82:89–155
28. Kutzelnig W, Liu W (2005) Quasirelativistic theory equivalent to fully relativistic theory. J Chem Phys 123:241102-4
29. Liu W (2010) Ideas of relativistic quantum theory. Mol Phys 108:1679–1706

30. Liu W, Peng D (2009) Exact two-component Hamiltonian revisitred. J Chem Phys 131:031104-4
31. Ilias M, Saue T (2007) An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation. J Chem Phys 126:064102-9
32. Barysz M (2010) Two-component relativistic theories. In: Barysz M, Ishikawa Y (ed) Relativistic methods for chemists. Springer, Dordrecht, pp 165–190
33. Dolg M (2002) Relativistic effective core potentials. In: Schwerdtfeger P (ed) Relativistic electronic structure theory, Part 1. Elsevier, Amsterdam, pp 793–862
34. Schwerdtfeger P (2011) The pseudopotential approximation in electronic structure theory. Chem Phys Chem 12:3143–3155
35. Lee YS (2002) Two-component relativistic effective core potential calculations for molecules. In: Schwerdtfeger P (ed) Relativistic electronic structure theory, Part II. Elsevier, Amsterdam, pp 352–416
36. Nash CS, Bursten BE, Ermler WC (1997) Ab initio relativistic effective potentials with spin-orbit operators. VII. Am through element 118. J Chem Phys 106:5153–5145
37. Mosyagin NS, Petrov AN, Titov AV, Tupitsyn II (2006) In: Julien JP (ed) Recent advances in the theory of chemical and physical systems. Springer, Dordrecht, pp 229–251
38. Koch W, Holthausen MC (eds) (2000) A chemist's guide to density functional theory. Wiley-VCH, Weinheim
39. Anton J, Fricke B, Engel E (2004) Noncollinear and collinear relativistic density-functional program for electric and magnetic properties of molecules. Phys Rev A 69:012505(10)
40. Liu W, Hong, G, Dai D, Li L, Dolg M (1997) The Beijing four-component density functional program package (BDF) and its application to EuO, EuS, YbO, and YbS. Theor Chem Acc 96:75–83
41. ReSpect, version 3.3.0. Relativistic Spectroscopy DFT program of authors Repisky M, Komorovsky S, Malkin VG, Malkina OL, Kaupp M, Ruud K, with contributions from Bast R, Ekstrom U, Knecht S, Malkin Ondik I, Malkin E. (2013). http://rel-qchem.sav.sk
42. Liu W, Peng D (2006) Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory. J Chem Phys 125:044102-10; Peng D, Liu W, Xiao Y, Cheng L (2007) Making one- and two-component density-functional methods fully equivalent based on the idea from atoms to molecules. J Chem Phys 127:104106-4
43. Van Wüllen C (2002) Relativistic density functional calculations on small molecules. In: Schwerdtfeger P (ed) Relativistic electronic structure theory, Part II. Elsevier, Amsterdam, pp 598–655
44. ADF, theoretical chemistry, Vrije Universiteit Amsterdam, The Netherlands. http://www.scm.com
45. Hess BA (1986) Relativistic electronic structure calculations employing a two-component no-pair formalism with external-field projection operators. Phys Rev A 33:3742–3748
46. Koepernik K, Eschrig H (1999) Full-potential nonorthogonal local-orbital minimum-basis band-structure scheme. Phys Rev B 59:1743–1757. http://www.FPLO.de
47. Pershina V, Bastug T (2005) Relativistic effects on experimentally studied gas-phase properties of the heaviest elements. Chem Phys 311:139–150
48. Johnson E, Fricke B, Keller OL Jr, Nestor CW Jr, Tucker TC (1990) Ionization potentials and radii of atoms and ions of element 104 (unnilquadium) and of hafnium (2+) derived from multiconfiguration Dirac-Fock calculations. J Chem Phys 93:8041–8050
49. Johnson E, Fricke B, Jacob T, Dong CZ, Fritzsche S, Pershina V (2002) Ionization potentials and radii of neutral and ionized species of elements 107 (bohrium) and 108 (hassium) from extended multiconfiguration Dirac-Fock calculations. J Phys Chem 116: 1862–1868
50. Nefedov VI, Yarzhemcky VG, Trzhaskovskaya MB (2004) Periodic low as applied to superheavy elements: specific features arising from relativistic effects. Russ J Inorg Chem 49:1871
51. Nefedov VI, Trzhaskovskaya MB, Yarzhemcky VG (2006) Electronic configurations and the periodic table for superheavy elements. Dokl Phys Chem 408:149

52. Indelicato P, Bieron J, Jönnson P (2011) Are MCDF calculations 101% correct in the superheavy element range? Theor Chem Acc 129:495–505
53. Pyykko P (2011) A suggested periodic table up to $Z \leq 172$, based on Dirac-Fock calculations on atoms and ions. Phys Chem Chem Phys 13:161–168
54. Liu W, Dolg M, Schwerdtfeger P. Benchmark relativistic all-electron density functional and ab initio pseudopotential study of group 12 dimers M_2 (M = Zn, Cd, Hg and eka-Hg). Unpublished
55. Seth M, Schwerdtfeger P, Dolg M (1997) The chemistry of the superheavy elements. I. Pseudopotentials for 111 and 112 and relativistic coupled cluster calculations for $(112)H^+$, $(112)F_2$, and $(112)F_4$. J Chem Phys 106:3623
56. Nash CS (2005) Atomic and molecular properties of elements 112, 114 and 118. J Phys Chem A 109:3493–3500
57. Pershina V, Borschevsky A, Eliav E, Kaldor U (2008) Prediction of the adsorption behavior of elements 112 and 114 on inert surfaces from ab initio Dirac-Coulomb atomic calculations. J Chem Phys 128:024707
58. Eliav E, Kaldor U, Ishikawa Y (1995) Transition energies in mercury and eka-mercury (element 112) by the relativistic coupled-cluster method. Phys Rev A 52:2765
59. Haynes WM (ed) (2012–2013) CRC handbook of chemistry and physics, 93rd edn. Taylor & Francis Ltd, London
60. Pershina V, Borschevsky A, Eliav E, Kaldor U (2008) Adsorption of inert gases including element 118 on noble metal and inert surfaces from ab initio Dirac–Coulomb atomic calculations. J Chem Phys 129:144106
61. Borschevsky A, Pershina V, Eliav E, Kaldor U (2009) Electron affinity of element 114, with comparison to Sn and Pb. Chem Phys Lett 480:49–51
62. Borschevsky A, Pershina V, Eliav, Kaldor U (2013) Ab initio predictions of atomic properties of element 120 and its lighter group-2 homologues. Phys Rev A 87:022502(8)
63. Pyykkö P, Atsumi M (2009) Molecular single-bond covalent radii for elements 1-118. Chem Eur J 15:186–197; Pyykkö P, Riedel S, Patzsche M (2005) Triple-bond covalent radii. *ibid* 11:3511–3620
64. Pershina V, Borchsevsky A, Ilias M (2014) Theoretical predictions of properties and volatility of chlorides and oxychlorides of group-4 elements. I. Electronic structure and properties of MCl_4 and $MOCl_2$ (M = Ti, Zr, Hf, and Rf). J Chem Phys 141:064314(8)
65. Pershina V, Borchsevsky A, Ilias M, Türler A (2014) Theoretical predictions of properties and volatility of chlorides and oxychlorides of group-4 elements. II. Adsorption of tetrachlorides and oxydichlorides of Zr, Hf, and Rf on neutral and modified surfaces. J Chem Phys 141:064315(7)
66. Pershina V, Anton J (2012) Theoretical predictions of properties and gas-phase chromatography behaviour of bromides of group-5 elements Nb, Ta and element 105, Db. J Chem Phys 136:034308(7)
67. Türler A, Brücle W, Dressler R, Eichler B, Eichler R, Gäggeler HW, Gärtner M, Glatz J-P, Gregorich KE, Hübener S, Jost DT, Lebedev VYa, Pershina V, Schädel M, Taut S, Timokhin N, Trautmann N, Vahle A, Yakushev AB (1999) First measurements of a thermochemical property of a seaborgium compound. Angew Chem Int Ed 38:2212–2213
68. Pershina V, Fricke B (1996) Group-6 dioxydichlorides MO_2Cl_2 (M = Cr, Mo, W, and element 106, Sg): the electronic structure and thermochemical stability. J Phys Chem 100:8748–8751
69. Han YK, Son SK, Choi YJ, Lee YS (1999) Structures and stabilities for halides and oxides of transactinide elements Rf, Db, and Sg calculated by relativistic effective core potential methods. J Phys Chem A 103:9109–9115
70. Even J, Yakushev A, Düllmann ChE, Haba H, Asai M, Sato TK, Brand H, Di Nitto A, Eichler R, Fan FL, Hartmann W, Huang M, Jäger E, Kaji D, Kanaya J, Kaneya Y, Khuyagbaatar J, Kindler B, Kratz JV, Krier J, Kudou Y, Kurz N, Lommel B, Miyashita S, Moritomo K, Morita K, Murakami M, Nagame Y, Nitsche H, Ooe K, Qin Z, Schädel M, Steiner J, Sumita T, Takeyama M, Tanaka K, Toyoshima A, Tsukada K, Türler A, Usoltsev I, Wakabayashi Y, Wang Y, Wiehl N, Yamaki S (2014) Synthesis and detection of a seaborgium carbonyl complex. Science 345:1491–1493

71. Nash CS, Bursten BE (1999) Prediction of the bond lengths, vibrational frequencies, and bond dissociation energy of octahedral seaborgium hexacarbonyl, $Sg(CO)_6$. J Am Chem Soc 121:10830–10831
72. Pershina V, Anton J (2013) Theoretical predictions of properties and gas-phase chromatography behaviour of carbonyl complexes of group-6 elements Cr, Mo, W, and element 106, Sg. J Chem Phys 138:174301(6)
73. Eichler R, Brüchle W, Dressler R, Düllman ChE, Eichler B, Gäggeler HW, Gregorich KE, Hoffman DC, Hübener S, Jost DT, Kirbach UW, Laue CA, Lavanchy VM, Nitsche H, Patin JB, Piguet D, Schädel M, Shaughnessy DA, Strellis DA, Taut S, Tobler L, Tsyganov YS, Türler A, Vahle A, Wilk PA, Yakushev AB (2000) Chemical characterization of bohrium (element 107). Nature 407:63–65
74. Pershina V, Bastug T (2000) The electronic structure and properties of group 7 oxychlorides, MO_3Cl, where M = Tc, Re, and element 107, Bh. J Chem Phys 113:1441–1446
75. Düllmann ChE, Brüchle W, Dressler R, Eberhardt K, Eichler B, Eichler R, Gäggeler HW, Ginter TN, Glaus F, Gregorich K, Hoffman DC, Jäger E, Jost DT, Kirbach UW, Lee DE, Nitsche H, Patin JB, Pershina V, Piguet D, Qin Z, Schädel M, Schausten B, Schimpf E, Schött HJ, Soverna S, Sudowe R, Thörle P, Timokhin SN, Trautmann N, Türler A, Vahle A, Wirth G, Yakushev AB, Zielinski PM (2002) Chemical investigation of hassium (element 108). Nature 418:859–862
76. Pershina V, Anton J, Jacob T (2008) Fully-relativistic DFT calculations of the electronic structures of MO_4 (M = Ru, Os, and element 108, Hs) and prediction of physisorption. Phys Rev A 78:032518(5)
77. Pitzer KS (1975) Are elements 112, 114, and 118 relatively inert gases? J Chem Phys 63:1032–1033
78. Eichler R, Aksenov NV, Belozerov AV, Bozhikov GA, Chepigin VI, Dmitriev SN, Dressler R, Gäggeler HW, Gorshkov VA, Haenssler F, Itkis MG, Laube A, Lebedev VYa, Malyshev ON, Oganessian YuTs, Petrushkin OV, Piguet D, Rasmussen P, Shishkin SV, Shutov SV, Svirikhin AI, Tereshatov EE, Vostokin GK, Wegrzecki M, Yeremin AV (2007) Chemical characterization of element 112. Nature. Letters 447:72–73
79. Borschevsky A, Pershina V, Eliav E, Kaldor U (2014) Relativistic couple cluster study of diatomic compounds of Hg, Cn, and Fl. J Chem Phys 141:084301-7
80. Liu W, van Wüllen Ch, Han YK, Choi YJ, Lee YS (2001) Spectroscopic constants of Pb and Eka-lead compounds: comparison of different approaches. Adv Quant Chem 39:325–355
81. Anton J, Fricke B, Schwerdtfeger P (2005) Non-collinear and collinear four-component relativistic molecular density functional calculations. Chem Phys 311:97–103
82. Pershina V, Anton J, Fricke B (2007) Intermetallic compounds of the heaviest elements and their homologs: the electronic structure and bonding of MM, where M = Ge, Sn, Pb, and element 114, and M = Ni, Pd, Pt, Cu, Ag, Au, Sn, Pb, and element 114. J Chem Phys 127:134310(9)
83. Gaston N, Opahle I, Gäggeler HW, Schwerdtfeger P (2007) Is eka-mercury (element 112) a group 12 metal? Angew Chem Int Ed 46:1663–1666
84. Pershina V, Bastug T, Fricke B, Jacob T, Varga S (2002) Intermetallic compounds of the heaviest elements: the electronic structure and bonding of dimers of element 112 and its homolog Hg. Chem Phys Lett 365:176–183
85. Sarpe-Tudoran C, Fricke B, Anton J, Pershina V (2007) Adsorption of superheavy elements on metal surfaces. J Chem Phys 126:174702(5)
86. Pershina V, Anton J, Jacob T (2009) Theoretical predictions of adsorption behavior of elements 112 and 114 and their homologs Hg and Pb. J Chem Phys 131:084713(8)
87. Zaitsevkii A, Titov A (2009) Relativistic pseudopotential model for superheavy elements: applications to chemistry of eka-Hg and eka-Pb. Russ Chem Rev 78:1173–1181
88. Zaitsevskii A, van Wüllen C, Rykova EA (2010) Two-component relativistic density functional modeling of the adsorption of element 114 (eka-led) on gold. Phys Chem Chem Phys 12:4152–4156
89. Eichler R, Aksenov NV, Albin YuV, Belozerov AV, Bozhikov GA, Chepigin VI, Dmitriev SN, Dressler R, Gäggeler HW, Gorshkov VA, Henderson RA, Johnsen AM, Kenneally JM,

Lebedev VYa, Malyshev ON, Moody KJ, Oganessian YuTs, Petrushkin OV, Piguet D, Popeko AG, Rasmussen P, Serov A, Shaughnessy DA, Shishkin SV, Shutov AV, Stoyer MA, Svirikhin AI, Tereshatov EE, Vostokin GK, Wegrzecki M, Wittwer PA, Yeremin AV (2010) Indication for a volatile element 114. Radiochim Acta 98:133–139

90. Yakushev A, Gates JM, Türler A, Schädel M, Düllmann ChE, Ackermann D, Andersson LL, Block M, Brüchle W, Dvorak J, Eberhardt K, Essel HG, Even J, Forsberg U, Gorshkov A, Graeger R, Gregorich KE, Hartmann W, Herzberg RD, Hessberger F, Hild D, Hübner A, Jäger E, Khuyagbaatar J, Kindler B, Kratz JV, Krier J, Kurz N, Lommel B, Niewisch J, Nitsche H, Omtvedt JP, Parr E, Qin Z, Rudolph D, Runke J, Schausten B, Schimpf E, Semchenkov A, Steiner J, Thörle-Pospiech P, Uusito J, Wegrzecki M, Wiehl N (2014) Superheavy element flerovium (element 114) is a volatile metal. Inorg Chem 53:1624–1629
91. Eichler B (1976) Das Flüchtigkeitsverhalten von Transactiniden im Bereich um Z = 114 (Voraussage). Kernenergie 19:307–311; Eichler B, Zvara I (1982) Evaluation of the enthalpy of adsorption from thermodynamic data. Radiochim Acta 30:233–238
92. Pershina V, Borschevsy A, Anton J, Jacob T (2010) Theoretical predictions of trends in spectroscopic properties of homonuclear dimers and volatility of the 7p elements. J Chem Phys 132:194314(11)
93. Pershina V, Borschevsy A, Anton J, Jacob T (2010) Theoretical predictions of trends in spectroscopic properties of gold containing dimers of the 6p and 7p elements and their adsorption on gold. J Chem Phys 133:104304(10)
94. Hermann A, Furthmüller J, Gäggeler HW, Schwerdtfeger P (2010) Spin-orbit effects in structural and electronic properties for the solid state of the group-14 elements from carbon to superheavy element 114. Phys Rev B 82:155116(8)
95. Pershina V, Anton J, Jacob T (2009) Electronic structures and properties of MAu and MOH, where M = Tl and element 113. Chem Phys Lett 480:157–160
96. Fox-Beyer BS, van Wüllen C (2012) Theoretical modelling of the adsorption of thallium and element 113 atoms on gold using two-component density functional methods with effective core potentials. Chem Phys 395:95–103
97. Serov A, Eichler R, Dressler R, Piguet D, Türler A, Vögele A, Wittwer D, Gäggeler HW (2013) Adsorption interaction of carrier-free thallium species with gold and quartz surfaces. Radiochim Acta 101:421–426
98. Nash CS, Bursten BE (1999) Spin-orbit coupling versus the VSEPR method: on the possibility of a nonplanar structure for the super-heavy noble gas tetrafluoride (118)F_4. Angew Chem Int Ed 38:151–153
99. Pershina V, Borschevsky A, Anton J (2012) Fully relativistic study of intermetallic dimers of group-1 elements K through element 119 and prediction of their adsorption on noble metal surfaces. Chem Phys 395:87–94
100. Pershina V, Borschevsky A, Anton J (2012) Theoretical predictions of properties of group-2 elements including element 120 and their adsorption on noble metal surfaces. J Chem Phys 136:134317(10); *ibid* (2013) 139:239901
101. Demidov Y, Zaitsevskii A, Eichler R (2014) First principles based modelling of the adsorption of atoms of element 120 on a gold surface. Phys Chem Chem Phys 16:2268–2270
102. Eichler B, Rossbach H (1983) Adsorption of volatile metals on metal surfaces and its application in nuclear chemistry. Radiochim Acta 33:121–125
103. Pyykkö P (2012) The physics behind chemistry and the periodic table. Chem Rev 112:371–384
104. Pyykkö P (2012) Predicting new, simple inorganic species by quantum-chemical calculations: some successes. Phys Chem Chem Phys 14:14734–14737
105. Pershina V (1998) Part II: hydrolysis and complex formation of Nb, Ta, Ha and Pa in HCl solutions. Radiochim Acta 80:75–84
106. Pershina V, Kratz JV (2001) Solution chemistry of element 106: theoretical predictions of hydrolysis of group 6 cations Mo, W, and Sg. Inorg Chem 40:776–780
107. Pershina V, Polakova D, Omtvedt JP (2006) Theoretical predictions of complex formation of group-4 elements Zr, Hf, and Rf in H_2SO_4 solutions. Radiochim Acta 94:407–414

108. Li ZJ, Toyoshima A, Tsukada K, Nagame Y (2010) Ion-exchange behavior of Zr and Hf as homologues of element 104, Rf, in H_2SO_4 and $H_2SO_4/HClO_4$ mixed solutions. Radiochim Acta 98:7–12
109. Pershina V (2004) Theoretical treatment of the complexation of element 106, Sg, in HF solutions. Radiochim Acta 92:455–462

Index

A

Ab initio calculations, 768, 786
  and crystal-field models, 787–791
  Dirac equation to c-SOCI, 768–769
  effective Hamiltonian theory, 771–772
  SOF electron correlation, 770–771
ab initio methods, 865
ab initio model potentials (AIMP), 470
Adsorption enthalpy, ΔH_{ads}, 862
Analytic continuation, 297
Asymptotic properties, 229–231
Atom-centered Gaussian basis functions, 84
Atomic and molecular electronic structure theory, 108
Atomic balance, 120
Atomic clock, 611–652
Atomic Hartree units, 109
Atomic mean-field integral (AMFI) approximation, 769
Atoms and molecules, 586, 607

B

Basis set expansion methods, 115–121
BBR shift (body radiation shift), 617, 628–630
Bilinear covariants, 18–19
Binuclear complexes, model Hamiltonians, 779
  giant-spin and block-spin Hamiltonians, 779–782
  multispin Hamiltonians, 782–785
Bloch equation, 318, 332
Block structures, 484–486
Born-Oppenheimer approximation, 260
Bound state quantum electrodynamics, Dirac equation, 133. *See also* Dirac equation
Breit exchange energy, 563
Breit interaction, 114, 323, 553
Brown-Ravenhall disease, 119, 125
'BSS' approach, 404

C

Calibration, 474
Central force, 37–40
Charge conjugation, 32–33, 348–352
Charge conservation
  quantum electrodynamics, 551
Clifford algebra, 14
Coalescence condition, 535, 536, 540, 542
  relativistic electron-electron, 502–521
  relativistic electron-nucleus, 499–502
Cohesive properties of solids
  relativistic corrections, 576
Complete active space self-consistent field (CASSCF), 770, 771
Complex formation, 859, 864, 875, 891, 892
Configuration interaction (CI), 590–593, 597, 607, 635–636
  method, 268
Constants of motion, 21
Contracted basis sets, 99–101
Contracted spin-orbit configuration interaction (c-SOCI), 768–769, 774, 784
Core-polarization potential (CPP), 451, 472–473
Correlation effects, 862, 864, 865, 871
Coulomb (C) interaction, 453
Coulomb potential, 38
Coupled cluster (CC), 635–638, 865, 872
  approach, 321
  method, 832, 835, 848
Covariance group, 7
Covariance of the equations, 7
Covariant, 14
Covariant-evolution operator, 324, 326

D

Density functional theory (DFT), 866, 867, 872, 876
  calculations, 452

W. Liu (ed.), *Handbook of Relativistic Quantum Chemistry*,
DOI 10.1007/978-3-642-40766-6

Diagonal local approximation to unitary transformation (DLU), 404
Diamagnetism, 672–677
Difference dedicated configuration interaction (DDCI), 771
Dirac-Coulomb
  breit and gaunt interaction, 113–114
  Hamiltonian, 109, 118
  nuclear volume and recoil corrections, 115
Dirac-Coulomb-Breit (DCB), 865, 868
  equation, 268–272
  Hamiltonian, 251, 252
Dirac derivation, 11–13
Dirac equation, 4, 9–13, 84, 90, 92, 244, 246, 768–769
  bilinear forms, 138
  Coulomb potential, 137, 141–142
  covariant form of, 15–16
  definition, 137
  external field, 25–27
  free fermion, 133
  4-momentum expression, 137
  Landé factor, 135–136
  operators mean-value for, 144–145
  predictions of, 146
  properties of, 143–144
  relativistic covariance of, 13–19
  spinors, 139–141
  tridimentional form, 135
Dirac-Fock-Coulomb (DFC), 865
Dirac-Fock-Slater or Dirac-Slater, 110
Dirac Hamiltonian, 84
Dirac-Hartree-Fock (DHF), 717
Dirac-Kohn-Sham (DKS), 717
  methods, 124
Dirac matrices, 550
Dirac γ matrices, 228–229
Dirac operator, 5
Dirac oscillator, 37
Dirac Sea, 110–112
Dirac spinor, 116
  Transformation properties of, 16–19
Dirac's relativistic theory, 396
DKH Hamiltonian, 400
Doppler effect, 615, 621
Double point group symmetry, 799, 811
Douglas–Kroll–Hess (DKH) theory, 396
Douglas-Kroll-Hess (DKHn) Hamiltonians, 769
Douglas–Kroll transformations, 399–402, 769
  Implementation of, 402–403

E
Effective core potential (ECP)
  approach, 450
  atomic, 456–457
Effective electric field, 582, 585, 588–590, 593–595, 606
Effective potential, 108, 110, 121, 125
Effective QED Hamiltonian, 267–284
Eigenvalue problem, 20–25
Electromagnetic interactions, 396
Electron affinities, 801–805
Electron correlation, 827, 831–834, 851
  effects, 831
  nondynamic, 827
  relativity and, 844, 852
Electron electric dipole moment, 581–607
Electron g-factor, 727
Electron paramagnetic resonance (EPR), 112, 726–761
Electron removal energy
  relativistic corrections
    atoms, 566
Electron self-energy, 111
Electron spin, 728, 729, 731–733, 740, 743
Energy-consistent approach, 468–469
Energy shift, 588, 589
Equation-of-motion coupled-cluster theory
  EEs, IPs or EAs, 801–805
  excited state and transition properties, 814–816
  implementation, 809
  time-reversal and spatial symmetry, 810–814
  transition and Excited state properties, 805–809
Exact decoupling, 397–398
  approaches, 403–405
Exactly solvable models, 34
Exact two-component (X2C), 383–392, 414, 418
Exchange-correlation energy functional
  relativistic density functional theory, 550, 553
Exchange-correlation energy relativistic corrections atoms, 563
Exchange-correlation magnetic field
  relativistic spin density functional theory, 555
Exchange energy functional no-pair relativistic density functional theory, 557
Exchange potential relativistic corrections atoms, 564

Excitation energy, 801–805
Existence theorem
relativistic density functional theory, 551
Explicit correlation, 535, 536
Extended no-pair projection, 535, 540
External energy functional
relativistic density functional theory, 553
External fields, 25–33

F
Fine structure constant, 612, 614, 629
Fine structure splitting, 818
Finite basis set disease, 125
Finite nucleus effects, 741
First-order DKH Hamiltonian, 401
Fock space, 323
Foldy–Wouthuysen transformations, 398–399
Four-component 'spin density, 679
Four-component theory, 688
Four gradient, 550
Fourier-Bessel, 63
Free Dirac operator, 20
Free-particle spinors, 22
Frequency standard, 612, 613, 620
Frozen-core approximation, 456
Frozen-core errors, 457, 458
Furry's theorem, 199

G
Galilei/Lorentz transformations, 8
Gamma matrices, 14
Gauge invariance, 27
Gauss-type, 62
Gaussian basis sets, 96–99
Gaussian nuclear charge model, 115
g-factor, 255–257
Gell-Mann–Low–Sucher method, 288, 289
General relativistic atomic structure (GRASP), 454
Generalized gradient approximation
relativistic extension, 575
Gordon decomposition, 32
Green's function, 156, 157, 160, 177, 185, 190, 199, 202, 207–211, 215, 218, 325
Green's operator, 327
Ground state density
relativistic density functional theory, 550
Ground state energy
relativistic corrections atoms, 564
relativistic density functional theory, 550, 553
relativistic LDA, GGA
closed-subshell atoms, 573
open-shell atoms, 574
Ground state four current
relativistic Kohn-Sham system, 552

H
Hamiltonian-like equation, 134
Harmonic oscillators, 34
Hartree energy functional
no-pair relativistic density functional theory, 553
relativistic density functional theory, 553
Heavy elements, 826–853
Hohenberg-Kohn logic, 121
Hohenberg-Kohn theorem, 122
Hybrid functionals, 123, 124
Hyperfine coupling (HFC), 728, 730, 735, 749–751

I
Infinitesimal rotation, 18
Instability, 613
Integral equation, relativistic optimized potential method, 560
magnetization-dependent, 562
spin-dependent, 563
Integrals of electron-nucleus attraction, 76
Interacting v-representability
relativistic density functional theory, 551
Internal symmetry, 490–493, 506–512
Ionization potential, 801–805
atoms, 568
relativistic corrections
atoms, 566
Isotropic coupling, 779

K
Källén-Sabry approximation, 168
Kinetic balance, 85–89, 116, 120
Kinetic energy functional
relativistic density functional theory, 553
Kohn-Sham eigenvalues
relativistic corrections
atoms, 565, 568

Kohn-Sham equations
relativistic density functional theory, 553
relativistic spin density functional theory
collinear, 556
non-collinear, 555
Kohn-Sham potential
relativistic density functional theory, 554
Kramers degeneracy, 812, 816
Kramers derivation, 10

L
Lamb shift, 111, 115, 246–253, 862
history, 246–249
modern explanation, 250–253
Landé factor, 135–136
Lentz vector, 141
Linewidth, 612, 613, 615
Linked-diagram theorem, 318
Local decoupling, 404
Local density approximation
relativistic exchange, 571
Lorentz-covariance, 7
Lorenz-invariant Dirac equation, 244
Lorentz transformation, 17, 150

M
Magic wavelengths, 643
Magnetic balance, 101–105, 678, 689
Magneto-structural correlations
ab initio calculations, 785–791
crystal-field models, 787–791
Matrix elements of potential energy, 75
Matter antimatter asymmetry, 585, 607
Mean-field approximation, 108
Mean-field spin-orbit Hamiltonian, 436
Mercury monohalides, 605–606
Metric tensor, 549
Micromotion, 615, 621, 622
Microwave clock, 613, 614, 650
Minimum coupling, 134
Minkowski indices, 549
Minkowski space, 4
Model core potential (MCP), 471
Model Hamiltonians, 772–773
binuclear complexes, 779–785
mononuclear complexes, 773–779
Modelpotentials (MPs), 469–470
ab initio modelpotentials, 470
model core potentials, 471
Molecular effective core potentials, 471
Mononuclear complexes, model Hamiltonians
$S = 1$ and $S = 3/2$ systems, 773–777
$S = 2$ and $S = 5/2$ systems, 777–779
Mulliken population analysis, 885
Multiconfiguration Dirac-Fock (MCDF), 268
Multi-configuration finite-difference Dirac-Hartree-Fock (MCDHF) , 454
Multiconfigurational self-consistent field (MCSCF), 770
Multiconfiguration Dirac-Fock (MCDF), 112–113, 219, 868, 869, 871, 891
method, 268
Multireference configuration interaction (MRCI), 771

N
Neutron, 53
Neutron number, N, 54
NMR shielding
theoretical description of, 661
Nonadiabatic dynamics, 415
Nonrelativistic, 71
basis sets, 99
limit and direct approach, 28
quantum chemistry, 84
theory, 109
No-pair approximation, 552
No-pair Hamiltonian, 383
Notation
relativistic density functional theory, 553
Nuclear attraction integrals, 75
Nuclear charge density distribution, 56, 58–65, 67
Fermi-type, 63
homogeneous (or uniform), 62
point-like, 61
Sum-of-Gaussians, 65
Nuclear charge form factor, 59
Nuclear electric quadrupole moment, 56
Nuclear magnetic dipole moment, 56
Nuclear magnetic resonance (NMR), 695–697
Nuclear magnetization distributions, 57–70
exponential, 68
Gauss-type, 69
modified exponential, 68
modified Gauss-type, 69
Nuclear mass number A, 55
Nuclear quadrupole coupling constants (NQCC), 56
Nuclear rms charge radius a, 59
Nuclear shielding
application, 681–682
Nuclear skin thickness t, 57
Nuclear spin, 55, 728, 730, 759
Nuclear spin-rotation tensor, 697–720
Hamiltonian definition, 694

nuclear shielding constants, 717
relativistic theory of, 696, 698, 715
Nucleon, 52–58
Nuclides, 54
chart of, 54

O

One-loop self-energy, 170–172
coordinate space, renormalization in, 176–177, 177
low-energy part, 173–175
regularization, 175–176
singular terms, evaluation of, 177–183
subtraction terms, evaluation of, 183–189
Open-shell systems, 797, 799, 801, 805
Optical clock, 613–617, 652

P

pair function, 320
Pauli corrections, 31
Pauli exclusion principle, 111
Pauli-Villars regularization, 175, 190
Periodic boundary conditions (PBCs), 767
Perturbation theory, 151
$\left\langle S^{(1)}_{\epsilon,1} \right\rangle_c$, 158–159
$\left\langle S^{(2)}_{\epsilon,1} \right\rangle_c$, 159, 170
adiabatic evolution operator, 151
Wick's theorem, contractions and propagators, 153 (*See also* Wick's theorem)
Photon exchange, 271
Poisson equation, 198
Position and velocity operators, Zitterbewegung, 22–25
Positrons, 32
Potential energy function, 58–66, 70–72
Predictions of experimental behaviour, 860
Primitive Gaussian basis sets, 96–99
Proton, 53
Proton number Z, 54
Pseudopotentials (PPs), 460–462
adjustment, 466–469
local form, 463
nonlocal form, 465–466
semilocal form, 463–465

Q

Quadrupole shift, 619, 622, 627–628
Quality factor, 613
Quantum chemical calculations, 864–868
Quantum electrodynamics (QED), 108, 111, 244–261, 287, 288, 348–363, 396
challenges, 258–261
continuum dissolution, 147
corrections, 268, 270–272
first order, 272–278
effects, 253–257
electromagnetic field, 150
electron field, 149–150
eQED Hamiltonian, algebraic derivation, 358–363
eQED Hamiltonian, diagrammatic derivation, 353–358
evolution operator, 148–149
perturbation theory, S-Matrix and energy shift, 150 (*see also* Perturbation theory)
representations, 147
self-energy screening, 200 (*see also* Self-energy screening)
two-photon, two-electron diagrams, 219–224
vacuum polarization, 189 (*see also* Vacuum polarization)
Quantum electrodynamics (QED), 287
corrections, 268, 270–272
first order, 272–278
Hamiltonian, systematic approach, 278–282
systematic approach, 278–282
Quantum field theory (QFT), 146
Quantum Monte Carlo (QMC) calculations, 452
Quasi-four-component (Q4C), 376–383

R

Radial behavior, 94–96
Radial function, atomic structure
relativistic, 72
series expansion for small r, 72
Radial functions in atomic structure, 71–73
Random phase approximation (RPA), 638–639
Redox potentials, 891
Relativistic configuration interaction (RCI), 219
Relativistic coupled cluster method, 600–606
Relativistic effective core potentials (RECP), 866, 875
Relativistic effects, 859, 860, 862, 865, 869, 872, 874, 879, 885, 891, 893
Relativistic Hamiltonian, 345–372, 484, 493

Relativistic Hartree-Fock, 109–113
matrix equations, 116–118
negative-energy eigenstates, 110–112
open-shell atoms, 112–113
Relativistic Kohn-Sham, 109
density functional theory basics, 121–123
equations, 552, 553
spin-density functional procedures, 123
Relativistic many-body perturbation (RMBPT) theory, 268
Relativistic mapping, 696, 697, 712–714
Relativistic molecular Hamiltonian, 663–666, 697, 698
Relativistic nuclear recoil theory, 115
Relativistic self-consistent field, 112, 124–125
Relativistic spin density, 556
functional theory, collinear, 556
Relativistic wave function, 482–495, 535, 536, 539
DCB Hamiltonian, 518
DCG Hamiltonian, 517–518
DC Hamiltonian, 517, 520
reduced Hamiltonian, 506–512
reduced two-electron problem, 502–506
Relativity, 826, 827, 839
dramatic effects of, 844
and electron correlation, 844, 852
interplay of, 836–838
Reproducibility, 613
Response function
relativistic Kohn-Sham system, 560
Restricted kinetic balance, 120
Roentgenium (Rg)
electronic states of, 455
orbital energies of, 458
radial expectation values $\mathrm{h}Ar\check{A}B$ of, 459
Rotational London orbitals (RLO), 697, 706–712

S

Scalar kinetic balance, 93
Schrödinger
Zitterbewegung, 24
equation, 8–9, 147, 148, 244, 260, 288
s-d transfer energy, 569
Second-order DKH Hamiltonian, 400
Secular motion, 621
Self energy, 271
contribution, 275–278
Self-energy screening, 200–202
Feynman diagrams, 201
global renormalisation, 216–219
high-energy part, 211–216
low-energy Green's function correction, 207–211
lower-order terms, 205
perturbative derivation, 202–204
wave function correction, 205–206
series expansion for small, 72
Shape-consistent approach, 467–468
$\vec{\Sigma}$-matrices, 555
Signal-to-noise ratio, 614
Simultaneous violation, of parity and time reversal symmetries, 582, 584, 607
Single-molecule magnet (SMM), 766–767
Space inversion and parity, 18
Spectroscopic constants
relativistic corrections, 576
Spin-angular functions, 91
Spin coupling, 90–94
Spin-dependent property, 437
Spin Hamiltonian, 726–735
Spinor basis sets, 116
Spin-orbit coupling(SOC), 727, 729, 731–733, 739, 745, 760, 766, 797
Spin-orbit free (SOF), 766, 770–771
Spin-orbit splitting, 96
Spinor space representations, 13
Spin separation
Dirac equation, 417
at two-component level, 417–423, 425
Spin-spin coupling (SSC), 766
Standard Model (SM), physics beyond, 582, 584
Stark shift, 616, 622, 624–626
Sublimation enthalpy, ΔH_{sub}, 862
Summation convention, 550
Superheavy elements (SHE), 858–893
Symmetry transformation, 7

T

ThO, 596–599
Time-reversal symmetry, 810–814
Tracy-Singh product, 486–490
Transactinide element, 859
Transition properties, 805–809, 814–816
Transverse exchange energy, 557
relativistic LDA, GGA
atoms, 572
Transverse interaction, 553
Two-component relativistic theory, of nuclear shielding, 682–688
2N-time Green's function, 289–294
Two-Time Greens Function (TTGF) method, 287–310, 324
analytical properties, 294–299

energy shift of single level, 299–304
quasi-degenerate level, 304–310

U
Uncontracted basis sets, 99

V
Vacuum polarization, 189, 271
contribution, 273–275
higher-order contributions, 199–200
one-potential Green function, 198–199
Uehling potential, 193–198
zero potential contribution, 190–192
Valence basis sets, 473
Variational equation
relativistic density functional theory, 551
Vector potential, 58, 66–70
Virtual electron-positron pairs, 24
Virtual pair effects, 355
Volatility, 862, 863

W
Wave functions, 769, 770
Wichmann-Kroll contribution, 274
Wick's theorem, 153, 154, 160, 165
electron field propagator, 154–157
photon field operators, 157
unperturbed states, 157

X
X2C, 403–404

Y
YbF, 593–596, 602–605

Z
Zeeman interaction, 730, 733–735, 743, 746, 747, 749, 752, 755–758
Zeeman shift, 620, 622–624
Zero-field splitting (ZFS), 766
ab initio calculations, 768–772
in magnetic properties, 767
magneto-structural correlations, 785 (*see also* Magneto-structural correlations)
model Hamiltonians and effective Hamiltonians, 772–773 (*see also* Model Hamiltonians)