

Exact decoupling of the relativistic Fock operator

Daoling Peng · Markus Reiher

Received: 18 June 2011 / Accepted: 31 August 2011 / Published online: 7 January 2012
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract It is generally acknowledged that the inclusion of relativistic effects is crucial for the theoretical description of heavy-element-containing molecules. Four-component Dirac-operator-based methods serve as the relativistic reference for molecules and highly accurate results can be obtained—provided that a suitable approximation for the electronic wave function is employed. However, four-component methods applied in a straightforward manner suffer from high computational cost and the presence of pathologic negative-energy solutions. To remove these drawbacks, a relativistic electron-only theory is desirable for which the relativistic Fock operator needs to be exactly decoupled. Recent developments in the field of relativistic two-component methods demonstrated that exact decoupling can be achieved following different strategies. The theoretical formalism of these exact-decoupling approaches is reviewed in this paper followed by a comparison of efficiency and results.

Keywords Relativistic electronic structure theory · Fock operator · Douglas–Kroll–Hess method · X2C method · Picture change error

1 Introduction

It is a well-established experimental fact that any mathematical description of electromagnetic phenomena

involving electrons and atomic nuclei has to obey the principles of special relativity [1]. As a consequence, a fundamental theory for chemistry should be a relativistically correct quantum-mechanical all-electron theory [2]. While a truly Lorentz-covariant many-electron theory is not available—although its basic principles have been cast in the theory of quantum electrodynamics [3]—it turned out that a semi-classical theory that quantizes the matter field only (first quantization) is sufficient if chemical accuracy for energies is sought, i.e., if relative energies shall be calculated with an accuracy of about 1 kJ/mol. For such a first-quantized theory, a relativistic many-electron Hamiltonian operator may be formulated as

$$H = \sum_i h_D(i) + \sum_{i < j} g(i, j) + V_{NN}, \quad (1)$$

where V_{NN} is the repulsion potential energy operator of the nuclei in a molecule, h_D Dirac's 4×4 one-electron operator, and $g(i, j)$ the two-electron interaction operator including the leading Coulomb term plus magnetic and retardation corrections as comprised by the Breit operator [2]. Hence, the electron–electron interaction in Eq. 1 is approximate. An additional approximation invoked in almost all practical cases is the Born–Oppenheimer approximation, i.e., the assumption of clamped nuclei. For the purist we should note that we refrained from embracing the one-electron Dirac Hamiltonian by positive-energy projectors (see below) for the sake of brevity.

Because of the four-dimensional matrix structure of the Dirac Hamiltonian h_D , the resulting orbital-based electronic structure methods are called four-component methods named after the number of functions that constitute a one-electron state. These one-electron states commonly referred to as orbitals are known as spinors in relativistic theory. A four-component orbital is called a 4-spinor, a

Published as part of the special collection of articles celebrating the 50th anniversary of Theoretical Chemistry Accounts/Theoretica Chimica Acta.

D. Peng · M. Reiher (✉)
Laboratorium für Physikalische Chemie, ETH Zürich,
Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland
e-mail: markus.reiher@phys.chem.ethz.ch

two-component orbital, which may be an eigenvector of a two-dimensional one-electron Hamiltonian, is called a 2-spinor. Unfortunately, jargon occasionally mixes this notation for operators with that for the corresponding orbitals and one speaks of four-component operators although four-dimensional operators are meant.

As the Dirac Hamiltonian includes a kinetic energy term associated with the electron's spin momentum in a natural way, spin-orbit interactions are consistently described. However, this can already be achieved to arbitrary accuracy by a purely two-dimensional Hamiltonian [4]. The four-dimensional structure is, however, the origin of negative-energy states which are interpreted as positronic states in quantum electrodynamics, but which require ad hoc assumptions in a first-quantized relativistic theory to assure stability of matter (Dirac's hole theory)—and, consequently, lead to conceptual problems when applied in chemistry.

From the point of view of numerical results, quantum chemistry based on standard Schrödinger quantum mechanics, the so-called non-relativistic approach, may yield numerical results that do not deviate significantly from a relativistic description. It is then said that the quantity studied is not affected by so-called relativistic effects. For example, most aspects that are studied in the context of organic molecules hardly show relativistic effects. But if we aim for a quantum mechanical theory valid for all chemistry, i.e., for molecules and molecular aggregates that may contain any atom from the periodic table of the elements, a “fully relativistic” four- or two-component approach is mandatory.

Here, it is important to distinguish between methods that are “quasi-relativistic” and thus do not completely reproduce four-component reference results for the same choice of electronic wave function approximation and methods that yield in principle the same result (and in practice results that agree with four-component results to the desired chemical accuracy). While this work is devoted to the latter kind, we should note that the former type of methods is usually split into scalar relativistic (also called one-component) approaches, whose computational cost are basically equivalent to that of the non-relativistic approach, and spin-orbit interactions including (two-component) approaches. Since the inclusion of spin in the Hamiltonian requires invoking the spin operator and hence the two-dimensional Pauli spin matrices, spin-orbit coupling including methods are always two-component and spin itself is no longer a good quantum number. As the scalar-relativistic approaches describe only kinematic relativistic effects, they change only the one-electron kinetic energy operators in the many-electron Hamiltonian of Eq. 1, but cannot account for any effects that are due to electron spin—like the spin-orbit interaction.

The four-component approach for the optimization of orbitals is computationally more demanding than the non-relativistic one by some, not very large constant factor (because of the matrix structure of the one-electron Hamiltonian in the former case). The four components of the orbital (spinor) also affect the scaling behavior of the four-index transformation for the application of subsequent correlation methods like configuration interaction or coupled cluster, but not the correlation methods themselves [5]. Hence, in *ab initio* correlation calculations that do not optimize the orbitals, four-component methods have basically the same computational cost as two-component methods. However, the four-component optimization of orbitals always automatically involves the optimization of the negative-energy solutions, which require positive-energy projectors if a variational procedure shall be applied to the otherwise unbounded Hamiltonian. One may consider the negative-energy states as pathologic as they cause interpretative problems in semi-classical relativistic theory employed here, “solved” only by Dirac's hole theory. As a consequence, actual calculations require the continuous update of projectors [6]. From an algorithmic point of view, these projectors are usually implicitly defined and clearly depend on the external potential. As a consequence, free-particle projectors proposed by Sucher [7] are not very appropriate as Heully et al. [8] pointed out. However, for the bound solutions, we are interested in, Talman [9, 10] showed that a minimax principle holds, which can be implemented in variational procedures also in the many-electron case [11]. Still, negative-energy states may cause numerical problems because of the choice of the one-particle basis set (usually kinetically balanced atom-centered Gauss-type functions), in which the negative-energy (continuum) solutions are to be represented.

The negative-energy solutions can be considered as superfluous for an electron-only theory that aims at the positive-energy solutions. However, they can be exactly removed by a decoupling of the negative- and positive-energy states. For this, two pathways have been followed in the past decades, namely the so-called elimination of the small component and the unitary transformation approach. The resulting Hamiltonians feature two-dimensional one-electron operators, which allow one to set up two-component methods.

The development of these two-dimensional operators was neither straightforward nor without difficulties. For instance, some featured energy-dependent operators or operators that are difficult to calculate. The early two-component approaches were called “quasi-relativistic” since some terms were discarded or approximated in the desire to obtain efficient methods to implement in computer programs. This has led to very efficient quasi-relativistic methods of which the zeroth-order regular

approximation (ZORA) [12–14] and the second-order Douglas–Kroll–Hess approximation (DKH2) [15] are the most prominent examples. Only within the past decade, it has become clear that one may actually formulate “exact” two-component methods which approach four-component reference results to (almost) arbitrary degree of accuracy. For reviews of these developments see Refs. [16–24].

This work attempts a presentation of the “exact” two-component approaches from the perspective of the one-electron equation that determines the orbitals (spinors) in a quantum chemical calculation. This point of view is usually not taken and instead only the Dirac Hamiltonian in an external electrostatic field is considered, which allows one to omit the discussion of how to deal with the electron–electron interaction. It is, however, the treatment of these two-electron terms (often accompanied by additional atom-based (local) approximations) that introduces approximations which are the reason why we put “exact” in quotation marks. It must be stressed that only a derivation of the various “exact” two-component methods from the point of view of the four-component Fock equation allows us to highlight all approximations made within a particular approach and to relate the different approaches to one another. In addition to this review of the formal aspects of “exact” two-component methods, we present a detailed comparison of them on the basis of new numerical results.

2 The Fock one-electron equation for the definition of orbitals

By contrast to standard presentations of the subject, we start from the relativistic (four-dimensional) Fock operator rather than from the Dirac Hamiltonian. This choice is made to clearly highlight the differences of “exact” two-component approaches and the corresponding four-component one. Most quantum chemical approaches approximate the many-electron wave function by a direct product of one-particle states. These one-particle states are obtained as eigenvectors of an effective one-electron Hamiltonian, well known as the Fock operator f .

The Fock operator comprises an operator for the kinetic energy of an electron and its interaction with external electromagnetic potentials (including those produced by the atomic nuclei) as well as operators that describe the interaction of the electron with other electrons. Especially, the latter ones are not easy to approximate without compromising the resulting accuracy of numerical results. The particular choice for such two-electron operators defines the electronic structure method under consideration, be it Hartree–Fock (HF), multi-configuration self-consistent field (MCSCF) or Kohn–Sham density functional theory (DFT).

All these methods rely on an effective one-electron eigenvalue equation

$$f\psi_i = \epsilon_i\psi_i \quad (2)$$

with orbital energy ϵ_i and orbital ψ_i . For a four-component method, the Fock operator reads in Gaussian units

$$f = h_D + q_e V_{\text{eff}} - q_e \boldsymbol{\alpha} \cdot \mathbf{A}_{\text{eff}}. \quad (3)$$

Here, the electromagnetic scalar and vector potentials, V_{eff} and \mathbf{A}_{eff} , respectively, are sums of all electromagnetic potentials that couple to the electron’s charge $q_e = -e$. Note that the potential energy operator $q_e V_{\text{eff}} - q_e \boldsymbol{\alpha} \cdot \mathbf{A}_{\text{eff}}$ is to be understood as a four-dimensional matrix operator.

The field-free Dirac Hamiltonian is given by

$$h_D = c\boldsymbol{\alpha} \cdot \mathbf{p} + (\beta - 1)m_e c^2, \quad (4)$$

where $\boldsymbol{\alpha}$ and $\beta = \text{diag}(1, 1, -1, -1)$ are the Dirac matrices and c is the speed of light. The vector $\boldsymbol{\alpha}$ contains three four-dimensional matrices whose off-diagonal two-dimensional entries are the Pauli spin matrices σ_i ,

$$\alpha_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \forall i \in \{x, y, z\}. \quad (5)$$

The first term on the right hand side of Eq. 4 is the relativistic kinetic energy operator and the second term is the rest energy operator of the electron, $\beta m_e c^2$ (m_e is its rest mass), shifted by its rest energy $-m_e c^2$ in order to have the same zero-energy reference level as in Schrödinger quantum mechanics.

The effective electromagnetic potentials in Eq. 3 depend on how the electron–electron interaction is described. This question has two facets as it refers to the choice of the interaction operator, i.e., of $g(i, j)$ in Eq. 1, and to the approximation of the *many-electron* wave function (or density). If we neglect all magnetic and retardation effects in $g(i, j)$, we obtain only terms in V_{eff} but not in \mathbf{A}_{eff} , while those effects, cast, e.g., as Breit operators, enter \mathbf{A}_{eff} (but can be written as a contribution to the scalar potential V_{eff} ; see the detailed discussion in chapter 8.1 of Ref. [2]). The wave-function approximation then affects the expressions for the electromagnetic potentials in Eq. 3 as the electromagnetic interaction of an electron with another electron depends on the other electron’s 4-current, i.e., on its density and current density to be calculated from the wave function.

We may write the effective electrostatic potential in general as

$$V_{\text{eff}} = V_{\text{ext}} \mathbf{1}_4 + V_{\text{ee}}, \quad (6)$$

where V_{ee} represents the electron–electron interaction whose explicit form depends on the wave function approximation. The external potential V_{ext} created by the

atomic nuclei and any other external electrostatic field is taken as a diagonal matrix operator (i.e., we do not consider magnetic and retardation effects for the electron–nucleus interaction, which is a common approximation in quantum chemistry as it does not compromise chemical accuracy) However, for the sake of brevity, we suppress the four-dimensional unit matrix $\mathbf{1}_4$ in the following.

In the independent-particle model of four-component Hartree–Fock theory, i.e., in Dirac–Hartree–Fock (DHF) theory we obtain

$$V_{\text{eff}}^{\text{DHF}} = V_{\text{ext}} + J - K, \quad (7)$$

where J denotes the classical Coulomb potential

$$J(\mathbf{r}) = q_e \sum_{j=1}^{\text{occ}} \int \frac{\psi_j^\dagger(\mathbf{r}') \cdot \psi_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' = q_e \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (8)$$

and K denotes the exchange potential defined as

$$K\psi_i(\mathbf{r}) = q_e \sum_{j=1}^{\text{occ}} \int \frac{\psi_j^\dagger(\mathbf{r}') \cdot \psi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \psi_j(\mathbf{r}) \quad (9)$$

(K produces off-diagonal (so-called odd) contributions of the scalar electron–electron Coulomb potential). In the MCSCF case, the two-electron terms have three spinor indices for a given Fock equation [2, 25, 26]. The four-component Kohn–Sham DFT case is similar to DHF theory and the effective electrostatic potential reads

$$V_{\text{eff}}^{\text{DFT}} = V_{\text{ext}} + J + V_{\text{XC}}, \quad (10)$$

where all exchange and electron-correlation terms are then obtained as functional derivatives of a properly approximated exchange–correlation energy contribution E_{xc} ,

$$V_{\text{XC}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}. \quad (11)$$

If magnetic interactions among the electrons shall be considered in DHF theory or relativistic DFT, the effective vector potential must be considered

$$\mathbf{A}_{\text{eff}}(\mathbf{r}) = \mathbf{A}_{\text{ext}}(\mathbf{r}) + \mathbf{A}_{\text{G}}(\mathbf{r}) + \mathbf{A}_{\text{XC}}(\mathbf{r}), \quad (12)$$

where \mathbf{A}_{ext} is an external vector potential induced by an external magnetic field. \mathbf{A}_{G} denotes the (unretarded) magnetic interaction of electrons

$$\mathbf{A}_{\text{G}}(\mathbf{r}) = \frac{q_e}{c} \int \frac{\mathbf{j}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (13)$$

which yields the Gaunt interaction and depends on the total current density \mathbf{j} of the electrons in the system. In DHF theory \mathbf{A}_{XC} would contain “magnetic” exchange integrals, while it is the relativistic current-density functional derivative in relativistic DFT,

$$\mathbf{A}_{\text{XC}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\mathbf{j}(\mathbf{r})]}{\delta \mathbf{j}(\mathbf{r})}. \quad (14)$$

In the relativistic realm, the exchange–correlation energy depends on both the electron density and the current density, therefore the exchange–correlation energy functional $E_{\text{xc}} = E_{\text{xc}}[\rho(\mathbf{r}), \mathbf{j}(\mathbf{r})]$ also appears in the expression for the effective vector potential.

Retardation effects that account for the transmission of electromagnetic fields between the electrons with the finite speed of light introduce additional terms that we here neglect as it has turned out that such terms—approximated, for instance, in the Breit interaction—lead to negligibly small corrections for chemical applications. Note that these terms depend on the choice of gauge for the electromagnetic fields. Also the magnetic Gaunt interaction has only very small contributions to electronic energy differences and is thus usually omitted in calculations. Note that these electronic contributions to the effective vector potential are usually written as contributions to the effective electrostatic potential (see chapter 8.1 of Ref. [2]).

For the decoupling of the negative-energy states, it will make a difference whether the effective potentials contribute to the block-diagonal or to the off-diagonal entries of the Fock operator. Therefore, it is advantageous to focus first on the electrostatic effective potential and to consider the magnetic contributions at a later stage.

The relativistic Fock operator in the absence of external magnetic field, $\mathbf{A}_{\text{eff}} = \mathbf{0}$, is

$$f = h_D + q_e V_{\text{eff}}. \quad (15)$$

If only a (block)-diagonal electrostatic potential is taken into account, which is an approximation often made in the derivation of two-component methods, the explicit matrix form of the relativistic Fock operator will then read

$$f = \begin{pmatrix} q_e V_{\text{diag}} & c\boldsymbol{\sigma} \cdot \mathbf{p} \\ c\boldsymbol{\sigma} \cdot \mathbf{p} & q_e V_{\text{diag}} - 2m_e c^2 \end{pmatrix}, \quad (16)$$

where $\boldsymbol{\sigma}$ denotes the vector of Pauli spin matrices. For example, the one-electron Dirac operator and the Fock operator in four-component DFT have a diagonal electrostatic potential, $V_{\text{diag}} = V_{\text{ext}}$ and $V_{\text{diag}} = V_{\text{ext}} + J + V_{\text{XC}}$, respectively. Recall that a diagonal term in the 2×2 superstructure representation of the relativistic Fock operator is actually a block-diagonal 2×2 matrix contribution.

The 2×2 superstructure of the Fock operator is a general feature and allows one to also write the corresponding 4-spinor in split notation, $\psi = (\varphi, \chi)$, with φ and χ as its upper and lower 2-spinor components. The components φ and χ are often called large and small components, respectively, because of their relative size for small external potentials (i.e., for small nuclear charge numbers Z). Since “relative size” is somewhat vaguely defined, we should

emphasize that it refers to a comparison of the absolute values of both components over the whole spatial domain, which shows only few exceptions (e.g., positions where the large components have nodes). One may use the square root of the integral of the large and of the small component *densities* over the whole spatial domain to define “size”. However, this classification is not valid for very heavy elements (e.g., for nuclear charge numbers approaching $Z = 100$ and higher Z 's) and therefore we consistently use the former notation of ‘upper’ and ‘lower’ to distinguish the components of the 4-spinor.

3 Theory of exact decoupling

The relativistic Fock equation in Eq. 2 has both positive-energy (so-called electronic) and negative-energy (so-called positronic) solutions

$$f\psi_i^{(+)} = \epsilon_i^{(+)}\psi_i^{(+)} \quad \text{and} \quad f\psi_i^{(-)} = \epsilon_i^{(-)}\psi_i^{(-)}. \quad (17)$$

The negative-energy solutions $\psi_i^{(-)}$ are not needed in molecular calculations which consider electrons and nuclei only. In DHF and relativistic DFT calculations, they are always kept unoccupied and are never used to construct the electron density and current density from which the effective potentials are calculated. Since HF, MCSCF and DFT approaches all utilize the variational principle, the negative-energy states, whose number is finite in molecular calculations only because of the choice of a finite basis set for the representation of the components of the spinor, cause problems for the variational stability. For this reason it is desirable to derive a Fock operator for electrons only.

3.1 Two principal options

In principle, there are two conceptually different approaches to obtain an electron-only Fock operator. One is the projection approach and the other one the block-diagonalization by unitary transformation. In the former one, a projection operator is introduced [7] to remove the negative-energy states

$$P^{(+)} = \sum_i |\psi_i^{(+)}\rangle\langle\psi_i^{(+)}|, \quad (18)$$

(the sum runs over all positive-energy states). The projected Fock operator $f^{(+)}$,

$$f^{(+)} = P^{(+)}fP^{(+)}, \quad (19)$$

possesses then only the positive-energy spectrum. But this four-component projected Fock operator is not useful in practice because the projection operator requires all information of the positive-energy states which are actually the result we aim for. One may introduce an atomic

approximation to the projection operator to overcome this problem, which is efficient, but would be a first approximation in an exact-decoupling scheme.

The second option is to block diagonalize the four-component Fock operator by a unitary transformation

$$f^{\text{bd}} = UfU^\dagger = \begin{pmatrix} f^{(2+)} & 0 \\ 0 & f^{(2-)} \end{pmatrix}, \quad (20)$$

where $f^{(2+)}$ and $f^{(2-)}$ possess only the electronic and positronic solutions, respectively:

$$f^{(2+)}\phi_i^{(+)} = \epsilon_i^{(+)}\phi_i^{(+)} \quad \text{and} \quad f^{(2-)}\phi_i^{(-)} = \epsilon_i^{(-)}\phi_i^{(-)}. \quad (21)$$

Obviously, one can now simply discard the $f^{(2-)}$ operator and employ the two-component electron-only Fock operator $f^{(2+)}$ to perform a relativistic molecular calculation. The computational applicability of the decoupling transformation depends crucially on whether one can derive an efficient and computationally feasible algorithm to compute the unitary transformation U as well as the corresponding Fock operator $f^{(2+)}$. We will discuss the possible options for their calculation in the following section.

3.2 The straightforward solution

The Fock equation, Eq. 2, can be rearranged to derive an operator X that relates the large (upper (U)) component φ and the small (lower (L)) component χ of a 4-spinor,

$$\chi^{(+)} = X\varphi^{(+)}. \quad (22)$$

Heully et al. [27] gave a closed form expression of the exact decoupling transformation in terms of this key operator X ,

$$U_X = \begin{pmatrix} \frac{1}{\sqrt{1+X^\dagger X}} & \frac{1}{\sqrt{1+X^\dagger X}}X^\dagger \\ -\frac{1}{\sqrt{1+XX^\dagger}}X & \frac{1}{\sqrt{1+XX^\dagger}} \end{pmatrix}, \quad (23)$$

whose derivation from the Fock equation is straightforward [2]. This transformation has been called a Foldy–Wouthuysen transformation [28], although the X -operator has not been introduced in the original paper by Foldy and Wouthuysen, that had a different focus to which we come back later. Note that the operator X is an electron-only operator which connects the large and small components of a positive-energy spinor. No analytical energy-independent closed form of the X -operator was discovered until now. However, the energy-dependent form of X can be easily derived from the Fock equation with the Fock operator of Eq. 16 [2],

$$X = \frac{c\boldsymbol{\sigma} \cdot \mathbf{p}}{2m_e c^2 - q_e V_{\text{diag}} + \epsilon_i^{(+)}}. \quad (24)$$

A more general expression can be derived for the general potential V_{eff} ,

$$X = \frac{c\boldsymbol{\sigma} \cdot \mathbf{p} + q_e V_{\text{eff}}^{LU}}{2m_e c^2 - q_e V_{\text{eff}}^{LL} + \epsilon_i^{(+)}}. \quad (25)$$

This energy-dependent expression of X then depends on which electronic spinor it acts on and does not lead to eigenvalue equations, which makes it useless for actual calculations.

3.3 The sequential solution

We can decompose the overall transformation U into a sequence of unitary transformations

$$U = \cdots U_3 U_2 U_1 U_0. \quad (26)$$

which is beneficial if the individual U_i are easier to obtain than an expression for the total U . Foldy and Wouthuysen [29, 30] were the first to attempt a derivation of such unitary matrices. Their main result was to give a closed form expression for U^{fpFW} that (block)-diagonalizes the field-free Dirac Hamiltonian,

$$\begin{aligned} U^{\text{fpFW}} \begin{pmatrix} 0 & c\boldsymbol{\sigma} \cdot \mathbf{p} \\ c\boldsymbol{\sigma} \cdot \mathbf{p} & -2m_e c^2 \end{pmatrix} U^{\text{fpFW},\dagger} \\ = \begin{pmatrix} E_p - m_e c^2 & 0 \\ 0 & -E_p - m_e c^2 \end{pmatrix}, \end{aligned} \quad (27)$$

The scalar relativistic energy E_p is given by $E_p = \sqrt{\mathbf{p}^2 c^2 + m^2 c^4}$, which is occasionally also abbreviated as E_0 because it is the zeroth-order term of the exactly decoupled operators [without subtraction of the rest energy]. However, any attempt to achieve a sequential decoupling with operators ordered according to the formal expansion parameter $1/c$ failed because of an ill-defined series expansion [4] that yields operators at most to be used in perturbation theory.

As we aim for an exact variationally stable procedure, we must consider the only other formal expansion parameter, namely the potential, which yields a convergent series expansion of a variational one-electron Hamiltonian [4]. The idea for an expansion in terms of the external electrostatic potential was first proposed by Douglas and Kroll [31], but found no application in electronic structure theory until rediscovered by Hess who also turned its low-order approximation into a practical method [15, 32]. The first unitary transformation in this Douglas–Kroll–Hess (DKH) transformation protocol must necessarily [4] be the above-mentioned closed-form Foldy–Wouthuysen transformation, $U_0 = U^{\text{fpFW}}$, if the off-diagonal terms to be eliminated in the Fock operator are given by $c\boldsymbol{\sigma} \cdot \mathbf{p}$ as in Eq. 16 and the new expansion parameter, the potential, shall be introduced to first order for the stepwise elimination up to infinite order by subsequent unitary transformations $U_1, U_2, \dots, U_\infty$.

Just as a historical side remark, we should mention that Nakajima and Hirao [33] were the first to present third-order DKH results based on Hess' original work on second-order DKH, followed by fourth- and fifth-order results by our group [34] and then sixth-order results by van Wüllen [35]. The final step in this direction was then our implementation of the first infinite-order DKH protocol [36] that allowed the explicit symbolic derivation and evaluation of the DKH one-electron Hamiltonian through all orders in the external potential, which has been made available in standard quantum chemistry programs [37, 38]. Its efficiency was significantly increased by Peng and Hirao [39] (see also below).

In addition to the analytical insights into the Douglas–Kroll–Hess approach presented in Ref. [4], we also presented in Ref. [34] a crucial reformulation called the *generalized* DKH transformation. This was necessary as different authors employed different parametrizations of the unitary transformation matrices in terms of an off-diagonal anti-Hermitian operator W to be chosen such that the lowest-order off-diagonal term in the one-electron Hamiltonian is stepwise eliminated. And it was not at all clear whether these different parametrizations yield the same results for finite DKH orders. The generalized DKH transformation clarified these matters by definition of the most general parametrization of a unitary transformation given by a Taylor series expansion,

$$U_k \rightarrow U_k(W_k) = \sum_{i=0}^{\infty} a_{k,i} W_k^i, \quad (28)$$

where the expansion coefficient $a_{k,i}$ are different for different parameterization schemes. These coefficients must be chosen such that the unitarity of the transformation is guaranteed.

The infinite-order DKH transformation,

$$U_{\text{DKH}} = \prod_{k=\infty}^0 U_k, \quad (29)$$

can exactly decouple the relativistic Fock operator of Eq. 16 (we denote the exact sequential DKH decoupling transformation simply as “DKH”, while we add a number to it if decoupling is achieved only to a given order in the electrostatic potential). If a sequential decoupling scheme is applied to the block-diagonalization of the relativistic Fock operator, then Douglas–Kroll–Hess theory is a unique approach and no other option to produce closed-form expressions (ordered by increasing powers of the potential) exists [4]. For exact block-diagonalization of the operator in Eq. 16, the full electrostatic potential V_{diag} must be considered as an expansion parameter. Standard DKH implementations take, however, only the electron–nucleus potential V_{ext} because this then requires only a modification

of the one-electron part of the Fock operator and leaves the electron–electron term V_{ee} untouched. This approximation yields an efficient scheme but, of course, introduces inaccuracies [40–43]. Of course, the exchange interaction in DHF theory as well as vector potential (i.e., magnetic) contributions make the DKH scheme even more involved due to the occurrence of off-diagonal potential terms in the Fock operator. However, to properly discuss the inclusion of magnetic fields in exact decoupling methods is beyond the scope of this work and we may refer the reader to the overview in Ref. [44] instead.

3.4 The two-step solution

The most severe problem of the DKH expansion approach when viewed as an exact-decoupling method is that it introduces too many unitary transformations and thus too many operators, which may make its implementation very complicated (this is no problem for the efficient standard low-order approximations like DKH2 or DKH4, but becomes a true challenge for high orders). Therefore, we should step back and reconsider the case if only one transformation is introduced after the free-particle Foldy–Wouthuysen transformation U_0 ,

$$U_{\text{BSS}} = U'_1 U_0 \quad \text{with} \quad U_0 = U^{\text{fpFW}}. \quad (30)$$

This idea was first proposed by Barysz et al. [45] and is therefore known as the BSS approach although it has later been called by Barysz and Sadlej the infinite-order two-component (IOTC) method [46]. Their transformation U'_1 is a U_X -like operator

$$U'_1 = \begin{pmatrix} \frac{1}{\sqrt{1+R^\dagger R}} & \frac{1}{\sqrt{1+R^\dagger R}} R^\dagger \\ -\frac{1}{\sqrt{1+RR^\dagger}} R & \frac{1}{\sqrt{1+RR^\dagger}} \end{pmatrix}, \quad (31)$$

where, instead of the original operator X of the (one-step) U_X transformation of Eq. 23, the operator R connects the upper and lower components of the U_0 -transformed spinor obtained as an eigenstate of the U_0 -transformed Fock operator

$$U_0^\dagger F U_0^\dagger = \begin{pmatrix} F_{UU} & F_{UL} \\ F_{LU} & F_{LL} \end{pmatrix}. \quad (32)$$

Here, we do not write explicit expressions as this is not important for our following discussion. Barysz and co-workers obviously had to face the same problem discussed above for the U_X transformation. I.e., the expression for U'_1 contains either unknown operators or impractical energy-dependent operators and no analytical form was obtained. However, the operator R can be obtained a solution of the following equation

$$F_{LL}R = RF_{UU} + RF_{UL}R - F_{LU}, \quad (33)$$

which is easy to derive from the free-particle Foldy–Wouthuysen-transformed Fock equation. Eq. 33 can be iteratively solved with a proper initial guess. Obviously, the iterative equation and iterative schemes for R are not unique. There exist many ways to set up and solve the iterative equations—a simple alternative to Eq. 33 would be

$$RF_{UU} = F_{LL}R - RF_{UL}R + F_{LU}, \quad (34)$$

but they may result in divergent numerical solutions (for a detailed discussion see Ref. [47]).

There are more difficulties associated with this iterative approach. First, if the iterations are carried out explicitly, the analytic expression of R becomes very complicated already after a few iterations (therefore, the iterative equations are always solved with matrix representations of the corresponding operators). Second, the solutions of the iterative equations are not unique, as it may produce an operator connecting the upper and lower components of any solution of the original Fock operator, which need not necessarily be an electronic solution. This means that even if the iteration converges, it may converge to an unwanted solution. However, by properly organizing the iterative solution, the correct operator R can be obtained and the Fock operator is then exactly decoupled by the BSS transformation.

3.5 The one-step solution

It should be clear that if the U_X -like transformation U'_1 can be obtained by an iterative solution, the original transformation U_X also has an iterative solution. The only difference is that the iterative equations are slightly different and may have different convergence behavior. The U_X transformation is indeed obtained within matrix representation via the so-called eXact-2-Component (X2C) approach [47–53]. However, an important characteristic of the X2C approach is that it invokes a non-iterative construction of the key operator X in U_X . The drawbacks of the iterative construction method discussed above do not exist in the X2C approach. In this non-iterative construction method, the matrix operator X is obtained from the electronic eigenvectors of the relativistic Fock–Roothaan equation

$$\begin{pmatrix} F_{UU} & F_{UL} \\ F_{LU} & F_{LL} \end{pmatrix} \begin{pmatrix} C_U^{(+)} \\ C_L^{(+)} \end{pmatrix} = \epsilon^{(+)} \begin{pmatrix} C_U^{(+)} \\ C_L^{(+)} \end{pmatrix}, \quad (35)$$

where the components of the 4-spinor are expanded in a one-electron basis set (in the above equation we kept the typical 2×2 superstructure of the Fock operator). $C_U^{(+)}$ and $C_L^{(+)}$ are the coefficients of the basis set expansions of upper (U) and lower (L) components of the 4-spinor eigenvectors, respectively. The diagonal matrix $\epsilon^{(+)}$

contains all positive-energy eigenvalues. Once the coefficients $C_U^{(+)}$ and $C_L^{(+)}$ have been obtained, the matrix X is simply obtained by

$$X = C_L^{(+)} \left(C_U^{(+)} \right)^{-1}. \quad (36)$$

This “trick” was called “Douglas–Kroll the easy way” by Jensen [48], who introduced it in a talk at the REHE2005 conference and which was then the starting point for the extensive development by Kutzelnigg and Liu [47, 49, 50]. Since the X matrix is directly evaluated from the electronic solutions of the Fock operator, the four-component Fock operator must first be diagonalized. *But this already solves the problem and no additional unitary transformation is required.*

Hence, a two-component electron-only Fock operator is then actually no longer needed to obtain electronic solutions. However, there are two points that must be considered. First, the exact-decoupling approach not only separates the electronic solutions from the positronic solutions but also constructs the electronic *two-component* spinors instead of four-component ones. This is clearly an advantage, but it will also reduce the effort for the four-index transformation if post-DHF correlation methods shall be applied. These issues have been discussed in detail in Ref. [54]. Second, as relativistic many-electron calculations require an external-field no-pair projection one may view this projection to be accomplished by the exact-decoupling approach. However, if actually a four-component calculation must be carried out before the two-dimensional operator can be evaluated (as in the X2C case), the “projection by two-component approach” is no valid advantage as the four-component variational solution for the 4-spinors already required (implicit) projection to the electronic solutions (in iterative protocols like the self-consistent field algorithm, the projectors may even be optimized implicitly when solving for the electronic 4-spinors; see the discussion in the Introduction and references given there).

Formally, the X2C decoupling transformation is just the U_X transformation

$$U_{X2C} = U_X, \quad (37)$$

although U_X is not obtained by Eq. 23 but through Eq. 36.

4 Algorithmic aspects of “exact” decoupling methods

Complementary to the principles of exact-decoupling methods discussed above, their actual implementation poses additional challenges that we shall discuss in this section. The most decisive insight which connects all decoupling methods is the construction of a so-called kinetically balanced basis set that ensures the correct

non-relativistic limit when the speed of light approaches infinity (in actual electronic structure calculations, the speed of light is then set to a sufficiently high value, e.g., to 1,000,000 Hartree atomic units).

In four-component calculations, basis functions for the small component must be chosen carefully. One cannot simply employ the same basis set for the large and small components. This would lead to variationally unstable results as already observed in the first attempt by Kim [5, 55] and to a wrong non-relativistic limit [56]. The correct non-relativistic limit is obtained if the kinetic-balance (KB) condition [28, 57–60] which relates the basis sets for large and small components,

$$\varphi \rightarrow \sum_{\mu} C_{U,\mu} \varphi_{\mu}, \quad \varphi_{\mu} \in \{\lambda_k\} \quad (38)$$

$$\chi \rightarrow \sum_{\mu} C_{L,\mu} \chi_{\mu}, \quad \chi_{\mu} \in \{\boldsymbol{\sigma} \cdot \mathbf{p} \lambda_k\}, \quad (39)$$

is obeyed. Here, $\{\lambda_k\}$ represents the space spanned by the set of basis functions λ_k . The KB condition is a natural requirement as it has its origin in the off-diagonal terms of the 2×2 superstructure of the Fock operator in Eq. 16. Strictly speaking, the requirement in Eq. 39 is a restricted KB condition. The most rigorous KB condition is defined by Eq. 22, but this equation is not useful as no closed-form solution for X exists and the energy-dependent expressions of Eqs. 24 and 25 are, of course, totally impractical for actual calculations. However, any choice of basis set which guarantees the correct non-relativistic limit is sufficient for exact-decoupling approaches. Hence, the one defined in Eq. 39, which fulfills

$$\chi_{\mu}^S = X_{\text{KB}} \chi_{\mu}^L, \quad \text{with} \quad X_{\text{KB}} = \frac{1}{2c} \boldsymbol{\sigma} \cdot \mathbf{p}, \quad (40)$$

is appropriate. In the case of numerical instabilities associated with Eq. 40, it can be advisable to choose the small-component basis functions normalized.

Equivalent to the KB condition is the idea of transferring the restriction on the small component’s basis functions to the relativistic Fock operator of Eq. 15 by

$$\begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\sigma} \cdot \mathbf{p} \end{pmatrix} f \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\sigma} \cdot \mathbf{p} \end{pmatrix} = \begin{pmatrix} q_e V_{\text{eff}} & c\mathbf{p}^2 \\ c\mathbf{p}^2 & \mathcal{V} - 2m_e c^2 \mathbf{p}^2 \end{pmatrix} \quad (41)$$

where the transformation is clearly inspired by the restricted KB condition. where $\mathcal{V} = (\boldsymbol{\sigma} \cdot \mathbf{p}) q_e V_{\text{eff}} (\boldsymbol{\sigma} \cdot \mathbf{p})$. The idea was proposed by Dyall [61] in the context of the so-called modified Dirac equation

$$\begin{pmatrix} q_e V_{\text{eff}} & c\mathbf{p}^2 \\ c\mathbf{p}^2 & \mathcal{V} - 2m_e c^2 \mathbf{p}^2 \end{pmatrix} \begin{pmatrix} \varphi_i \\ \tilde{\chi}_i \end{pmatrix} = \epsilon_i \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{p}^2 \end{pmatrix} \begin{pmatrix} \varphi_i \\ \tilde{\chi}_i \end{pmatrix}, \quad (42)$$

which we should call here the modified Fock equation. Note that the lower component $\tilde{\chi}_i$ is no longer the small component of the original Dirac spinor. It is called the pseudo-large component. Now, the same basis set can be employed for the large and pseudo-large components of the modified Fock operator.

However, the modified Fock equation, Eq. 42, changes the overlap metric as can be seen on the right hand side. It is therefore more convenient to introduce a unitary transformation to ensure the important KB condition and also to preserve the identity as the metric. This is achieved with the help of a special operator s ,

$$s = \frac{\boldsymbol{\sigma} \cdot \mathbf{p}}{p}, \text{ with } p \equiv \sqrt{\mathbf{p}^2}. \quad (43)$$

because its square is the two-dimensional unit matrix, $s^2 = \mathbf{1}$. The KB unitary transformation [56] then reads

$$U_{\text{KB}} = \begin{pmatrix} 1 & 0 \\ 0 & s \end{pmatrix}. \quad (44)$$

and yields the KB-transformed Fock operator

$$\tilde{f} = U_{\text{KB}} f U_{\text{KB}}^\dagger = \begin{pmatrix} q_e V_{\text{eff}} & cp \\ cp & \frac{v}{p^2} - 2m_e c^2 \end{pmatrix}. \quad (45)$$

Now, the corresponding transformed Fock equation does no longer need a non-identity metric. In the following discussion, we denote the form \tilde{O} with a tilde on top of the symbol as the KB-transformed operator of an original operator O

$$\tilde{O} = U_{\text{KB}} O U_{\text{KB}}^\dagger. \quad (46)$$

It is crucial to employ the KB-transformed operators in the implementation of exact-decoupling methods.

The operator s in Eq. 43 is a quite special operator. The calculation of the square root of \mathbf{p}^2 needed for p seems to be difficult to evaluate considering the usual definition of the momentum operator as a differential operator. Moreover, there exists a restriction on the evaluation of operator p^{-1} arising from the KB condition. The operator s must preserve the KB condition such that the space spanned by $\{s\lambda_k\}$ must be equivalent to $\{\boldsymbol{\sigma} \cdot \mathbf{p}\lambda_k\}$. From the form of operator s we understand that the KB condition thus reduces to requiring that the space spanned by $\{p^{-1}\lambda_k\}$ must be equivalent to $\{\lambda_k\}$. Therefore, p^{-1} must be defined within the basis functions space $\{\lambda_k\}$, which a priori is not a condition trivially fulfilled in position space (only in momentum space it is).

In general, it is easy to calculate the action of operators that are algebraic functions of \mathbf{p}^2 in the space of eigenfunctions of \mathbf{p}^2 , $\{\theta_i\}$. Then, the operator \mathbf{p}^2 can be replaced by its eigenvalues

$$\mathbf{p}^2 \theta_i = p_i^2 \theta_i \implies f(\mathbf{p}^2) \theta_i = f(p_i^2) \theta_i. \quad (47)$$

For a finite basis-function space, $\{\lambda_k\}$, Hess [15, 32, 62] suggested that the exact momentum eigenfunctions θ_i are to be replaced by the eigenfunctions of the matrix representation of \mathbf{p}^2 , $\{\langle \lambda_k | \mathbf{p}^2 | \lambda_l \rangle\}$. A transformation into this basis is easily achieved as the non-relativistic kinetic energy matrix, which is available in every quantum chemistry program package, is proportional to \mathbf{p}^2 and can be diagonalized after multiplication by $-2m_e$. Within this scheme, the KB condition is satisfied since any $p^{-1}\lambda_k$ belongs to the space $\{\lambda_k\}$.

In particular, all unitary transformations applied to the relativistic Fock operator must preserve the KB condition. The explicit form of the KB-transformed free-particle Foldy–Wouthuysen transformation [29] reads

$$\tilde{U}_0 = \begin{pmatrix} \sqrt{\frac{E_0 + m_e c^2}{2E_0}} & \sqrt{\frac{E_0 - m_e c^2}{2E_0}} \\ -\sqrt{\frac{E_0 - m_e c^2}{2E_0}} & \sqrt{\frac{E_0 + m_e c^2}{2E_0}} \end{pmatrix}. \quad (48)$$

It only consists of operators which are algebraic functions of \mathbf{p}^2 . Therefore, they can be evaluated within the Hess scheme. Obviously, \tilde{U}_0 will not violate the KB condition. For the DKH expansion algorithm to exactly decouple the relativistic Fock operator, the subsequent U_k ($k > 0$) transformations as well as their components W_k must be evaluated within the KB-transformed space. In other words, one must calculate the matrices of \tilde{U}_k and \tilde{W}_k instead of their untransformed forms.

The traditional arbitrary-order DKH approach did not evaluate the KB-transformed \tilde{U}_k , but expanded the final Fock operator in terms of low-level operators [36]. The high-level intermediates, which are very useful to reduce the computational costs, were not used. Instead, the traditional DKH approach then leads to an exponentially scaling algorithm with increasing order of the expansion, which can make it hard to approach infinite-order results in practice. By contrast, the DKH method using KB-transformed operators scales only polynomially [39] so that the calculation of infinite-order results is feasible for any element from the periodic table.

As a side remark, we should note that the expansion formulation in terms of low-level operators is not directly evaluated. In order to reduce the number of matrix operators needed to evaluate the Hamiltonian, the resolution of identity operator $s^2 = 1$ is inserted into proper positions so that only a small number of operator matrices is required for the evaluation of the two-component DKH operator. In its scalar-relativistic variant, the identity to be employed reduces to \mathbf{p}^2/p^2 and only the matrix representation of $\mathbf{p} \cdot \tilde{V} \mathbf{p}$ is needed in addition to \tilde{T} and \tilde{V} (see Sect. 5.2). We

found the insertion approach to be numerically equivalent to the approach using the KB-transformed operators

In the BSS approach, the operator R is also not directly evaluated. In fact, the BSS approach proposes a set of iteration equations for an operator Y to replace R . The operator Y is defined as

$$Y = sR = \frac{\sigma \cdot p}{p} R. \quad (49)$$

This definition turned to be the most crucial point for the implementation of the BSS method in a computer program. However, the reason why this step was invoked appears to be mostly historical from our current perspective—namely, it paralleled the procedure of the DKH approach [17], namely to avoid the evaluation of matrix of operators which include an odd number of $\sigma \cdot p$ operators. By contrast, here, we start from a more fundamental point of view, namely from the KB condition and the non-relativistic limit, whose true importance has not been recognized in previous work on the BSS approach.

Not surprisingly, in implementations of the X2C method the X matrix operator is evaluated from the $C_L^{(+)}$ matrix which consists of basis set expansion coefficients of the pseudo-large components instead of the small components. This means that the X matrix is evaluated in the KB-transformed basis-function space.

Exact-decoupling methods have been given many names by different authors mostly for historical reasons. Actually, there exist only three variants of exact-decoupling methods up to now. These three variants and their main protagonists shall be briefly reviewed in the following three subsections. In Sect. 6, we shall demonstrate how efficient the three variants of exact-decoupling methods are. We will then also demonstrate that, if properly implemented, all three variants are almost equally efficient.

4.1 One-step transformation

If the exact-decoupling method is algorithmically achieved by only one unitary transformation, we shall call it a “one-step transformation”. On formal grounds this notation may be ambiguous as the single unitary transformation can be decomposed or combined from more than one unitary transformation. This is the reason why we write “algorithmically achieved” in order to clearly state that it is a matter of implementation into a computer program. Currently, the only example for a ‘one-step transformation’ is the U_X -transformation, but there might exist other analytical expressions that can be implemented in a single step.

The X2C method implements a one-step transformation. However, the discussion of the X2C method usually starts from the modified Dirac equation proposed by Dyall [61], but the modified Dirac equation method is equivalent to the

four-component method with KB-transformed basis functions. Very closely related to this method is the normalized-elimination of the small component (NESC) method [63–67] also proposed by Dyall, which is an electron-only method, but with eigenfunctions expressed in terms of the large components of electronic 4-spinors. In one NESC paper [66], Dyall discussed the transformation (with a renormalization matrix) to pure two-component wave functions, and this version of the NESC method has almost all characteristics of the current X2C method except for the construction scheme of the X matrix. Dyall employed an energy- or eigenfunction-dependent form to evaluate the X matrix iteratively [68], although a non-iterative scheme as in X2C could also be formulated in the NESC framework. It must be noted that when the essential ideas of the X2C approach had been worked out in 2005, it turned out that Filatov and co-workers [68–70] had actually come to similar conclusions considering an iterative NESC approach [71, 72].

In a series of papers by Kutzelnigg and Liu [47, 49, 50] the iterative way to construct the X matrix was discussed in detail. These authors suggested the non-iterative construction scheme for many-electron calculations. Later, the non-iterative construction approach was implemented into the BDF program [51, 52] by Peng and Liu for molecular calculations. The method was first called XQR (exact quasi-relativistic) or infinite-order quasi-relativistic method. This XQR method is a pure two-component method employing Dyall’s renormalization matrix. A one-step transformation method named IOTC (infinite-order two-component) was implemented into the DIRAC program [53] by Iliaš and Saue. [Note that the name IOTC has also been used for other exact-decoupling approaches (see Sect. 3.4 above) and it may easily cause misunderstandings.] Iliaš and Saue adopted almost the same algorithm as the one implemented in BDF except for Dyall’s renormalization matrix (and they used a numerically more stable expression for the evaluation of the X operator). In their implementation, the basis functions are first converted into an orthonormalized set and every matrix is then evaluated within the orthonormalized set, while the implementation in the BDF program uses the matrices expressed in unnormalized basis functions. Dyall’s renormalization matrix turned out to be problematic [73] and a new implementation in the BDF program fixed this problem using a new renormalization matrix, which made the BDF implementation equivalent to the IOTC method of the DIRAC program. Finally, Liu and co-workers found the name XQR to be not suitable to describe this approach. The acronym “X2C” is now commonly used as the name for the one-step exact-decoupling transformation approach as a result from extensive discussions of Jensen, Kutzelnigg, Liu, Saue, and Visscher at the DFT-2007 conference in Amsterdam in August 2007. so far. Unfortunately, any

exact-decoupling method may be called an exact two-component method or infinite-order two-component method. However, the acronym X2C has been used only for this special one-step transformation algorithm. The current X2C method has the following features : (1) an U_X -type transformation is employed, (2) X is defined within a KB-transformed basis, (3) X is non-iteratively constructed.

4.2 Two-step transformation

The first transformation of a two-step method is the free-particle Foldy–Wouthuysen transformation which was considered necessary for the BSS approach [4]. As discussed above, the second step is then the U_X -type transformation as first proposed by Barysz et al. [45]. It was the first approach proposed to exactly decouple the relativistic Dirac Hamiltonian in an external electrostatic potential operator. The non-iterative construction of the Y -operator has been discussed in a paper by Kędziera and Barysz [74] but was already mentioned in the talk by Jensen [48]. Almost all other calculations published employed the iterative scheme. One exception that used the prescription of Jensen was given in Ref. [75] for calculations on PbO. The scheme is usually called the “BSS approach” [4, 45, 46], but later Barysz et al. preferred the abbreviation “IOTC”. Other names for their approach, which have been used, are IOFW (infinite-order Foldy–Wouthuysen) [76, 77] and IODK (infinite-order Douglas–Kroll) [43, 78]. Clearly, the adjective “infinite-order” is not appropriate to describe the latest version of the BSS approach, since this approach uses either an iterative or non-iterative algorithm (like the X2C method) to achieve exact decoupling. A notion referring to an “order” is rooted in the history of this approach, which was proposed in 1997 as an attempt to decouple to a certain order in $1/c$ [4, 45]. As the attribute “infinite-order” is no longer suitable and we continue to simply use the name “BSS method”.

The conventional BSS method has the following characteristics : (1) the free-particle Foldy–Wouthuysen transformation is first applied to the Fock operator, (2) a U_X -type transformation is employed in the second step, (3) the Y operator is either iteratively or non-iteratively constructed.

However, it appears that the two-step transformation has no advantages over the one-step transformation and hence the X2C scheme can be used instead.

4.3 Expansion of the transformation

The DKH method is the only one of this category. In the past, the acronyms DKH (or DK) referred to Hess’ original truncated DKH2 method (DKH to 2nd order in the electrostatic potential). However, one should clearly distinguish finite-order from infinite-order DKH results. Also, the still

used abbreviation “DK” for “Douglas–Kroll” should be avoided in favor of “DKH” to highlight Hess’ work without which the suggestion by Douglas and Kroll would probably not been known (apart from the fact that it was Hess who demonstrated how to employ the DK transformation in actual calculations). We should note that different parameterizations of U_k , i.e., different sets of $a_{k,i}$ expansion parameters, give different exactly decoupled Fock operators. The infinite-order DKH Fock operators are therefore not unique. However, at infinite-order, the results obtained for expectation values are, of course, the same, independent of the chosen parametrization. Only the infinite-order DKH spinors differ by a unitary transformation from one another.

The DKH method has the following characteristics : (1) the free-particle Foldy–Wouthuysen transformation is applied to the relativistic Fock operator, (2) the electrostatic potential is used as an order parameter for an order-by-order expansion of the relativistic Fock operator, (3) the U_k transformation matrices are parametrized in terms of off-diagonal anti-Hermitian W_k of k -th order in the electrostatic potential.

5 Approximations involved in many-electron calculations

So far, we have discussed the three existing variants for exact-decoupling methods. In principle, they are all exact two-component methods employing the full electrostatic potential V_{eff} . However, in practice, approximations are introduced in order to increase the efficiency (ideally without compromising the accuracy). The discussion to follow now is independent of how the exact-decoupling transformation U is obtained and thus holds for all exact-decoupling methods.

5.1 The cumbersome two-electron terms

In many-electron calculations, the effective electrostatic potential contains two terms

$$V_{\text{eff}} = V_{\text{ext}} + V_{ee}[\{\psi_i^{(+)}\}], \quad (50)$$

i.e. the external potential and the effective potential from electron–electron interactions which depends on the positive-energy spinors that enter the expression for the total electronic wave function (cf. Sect. 2) Therefore, the exact-decoupling transformations must be updated if the spinors have changed, e.g., upon their optimization in a self-consistent field procedure (or when the positions of the nuclei, i.e., the molecular structure is changed). In general, this change of the exact-decoupling transformation has been shown to be small [54, 79]. An exact-decoupling transformation constructed only for the external electrostatic potential, $V_{\text{eff}} \rightarrow V_{\text{ext}}$,

turned out to be an excellent approximation for molecular many-electron calculation. Within this approximation, the exact two-component Fock operator is approximated as

$$f^{(2+)} \approx \left(U[V_{\text{ext}}] (h_D + q_e V_{\text{ext}} + q_e V_{\text{ee}}[\psi_i^{(+)}) U[V_{\text{ext}}]^\dagger \right)^{++}, \quad (51)$$

where $(\dots)^{++}$ denotes the upper-left part of the transformed four-dimensional operator. Formally, the choice of $U = U[V_{\text{eff}}]$ for exact decoupling preserves equivalency to the four-component approach, while $U = U[V_{\text{ext}}]$ is equivalent to the external-field no-pair projection approximation.

Untouched by this approximation, the untransformed electron–electron interaction potential operator, $V_{\text{ee}}[\{\psi_i^{(+)}\}]$, still depends on all occupied positive-energy 4-spinors in the many-electron case. To exploit the advantages of the two-component approach, we need to compute it from the 2-spinors $\phi_i^{(+)}$. There exist two approaches to achieve this. One is the back-transformation approach. In an exact-decoupling method, we have available both the exact-decoupling transformation and the 2-spinor. It is then easy to back-transform the two-component orbital to a four-component one:

$$\psi_i^{(+)} = U^{-1} \begin{pmatrix} \phi_i^{(+)} \\ 0 \end{pmatrix} = U^\dagger \begin{pmatrix} \phi_i^{(+)} \\ 0 \end{pmatrix}. \quad (52)$$

However, this approach is only useful to calculate one-electron properties, but it is useless for calculating two-electron integrals, because the former requires two times a back-transformation, while latter requires four. If the back-transformation is applied four times, the computational costs will be higher than those of the corresponding four-component calculation.

If we directly replace the four-component spinor $\psi_i^{(+)}$ by its corresponding two-component spinor $\phi_i^{(+)}$ in the calculation of the V_{ee} operator, we introduce a picture change error (see below). To correct for this error, we may add a correction term, V_{cor} [42]. The correction term can be added either before the transformation is carried out,

$$f^{(2+)} \approx (U[V_{\text{ext}}] (h_D + V_{\text{ext}} + V_{\text{ee}}[\phi_i^{(+)}) + V_{\text{cor}}) U[V_{\text{ext}}]^\dagger)^{++}, \quad (53)$$

or as an a posteriori correction

$$f^{(2+)} \approx (U[V_{\text{ext}}] (h_D + V_{\text{ext}}) U[V_{\text{ext}}]^\dagger)^{++} + V_{\text{ee}}[\phi_i^{(+)}) + V_{\text{cor}}. \quad (54)$$

Note that the correction term V_{cor} is not the same in both equations.

Atomic mean field (AMFI) [80] and screened spin–orbit (SNSO) [81] approaches have been proposed to correct for

two-electron picture change errors, but so far they have only been applied to quasi-relativistic methods. An extension to exact-decoupling methods would be most desirable. However, we should emphasize that in many calculations the bare (untransformed) electron–electron interaction operator is employed in the two-component Fock operator

$$f^{(2+)} \approx \left(U[V_{\text{ext}}] (h_D + V_{\text{ext}}) U[V_{\text{ext}}]^\dagger \right)^{++} + V_{\text{ee}}[\phi_i^{(+)})]. \quad (55)$$

It is already known for a long time from quasi-relativistic calculations that this bare-potential approximation provides reliable results for valence electron properties (this is particularly true (see e.g., Refs. [82, 83]) for the scalar-relativistic variant introduced in the next section, but may be different for the truly two-component method including spin–orbit splitting [84]; but see also the numerical results section below). However, it might be not good enough for core-electron properties and gives large errors for the spin–orbit splitting of high-angular-momentum orbitals [42, 81].

5.2 Scalar-relativistic approximations

There exist scalar versions of the exact-decoupling methods, which have the huge advantage that they can be easily interfaced with a standard non-relativistic quantum chemistry program package. These scalar-relativistic versions allow for an efficient description of kinematic relativistic effects. However, scalar exact-decoupling is not uniquely defined, because there is no unique definition of the scalar full-relativistic Fock equation [85]. There exist two principal ways to obtain a scalar exact-decoupling Fock operator.

One is the “a priori” approach, where a scalar fully relativistic Fock equation is defined first, then follows the same steps as in the exact decoupling of the two-dimensional operator to obtain a scalar electron-only Fock operator. The commonly used scalar fully relativistic Fock equation is obtained by replacing the two-component \mathcal{V} term in the Fock equation, Eq. 42, by a scalar operator $(\mathbf{p} \cdot V\mathbf{p})$. The scalar exact-decoupling method would then provide the same eigenvalues. Another option is the “a posteriori” approach. Once the two-component electron-only Fock operator has been obtained, the spin-dependent terms are discarded (by virtue of Dirac’s relation) to obtain the scalar version.

5.3 Local approximations

Because of the fact that all operators in molecular electronic structure calculations are evaluated in a one-electron basis set, the matrix operations required by the exact-decoupling methods require a computational effort that scales with the molecular size rather than with the number

of heavy atoms, which contribute most to the numerical relativistic effect. For large molecules, efficient local approximations are required. Naturally, one may restrict the unitary transformation to those matrix elements of basis functions that are located at a heavy atom (see e.g., the work by Peralta and Scuseria [86, 87]). A more systematic analysis has been provided by Thar and Kirchner [88]. The main conclusion from these studies is that the unitary transformation can be restricted to atom-same-atom diagonal blocks of the relativistic Fock operator without loss of numerical accuracy. However, a more rigorous localization scheme, which would also include all relevant atom-other-atom off-diagonal blocks of the relativistic Fock operator, is desired.

5.4 Transformed expectation values

To correctly evaluate an expectation value of a molecular property operator P from a two-component wave function the operator P must be transformed as well

$$\begin{aligned} \langle \psi_i^{(+)} | P | \psi_i^{(+)} \rangle &= \langle U \psi_i^{(+)} | U P U^\dagger | U \psi_i^{(+)} \rangle \\ &= \langle \phi_i^{(+)} | (U P U^\dagger)^{++} | \phi_i^{(+)} \rangle \\ &= \langle U^\dagger \phi_i^{(+)} | P | U^\dagger \phi_i^{(+)} \rangle. \end{aligned} \quad (56)$$

If the transformation of the property operator is neglected, a picture change error [89, 90] is introduced, whose magnitude depends on the type of property considered [91–97]. In general, the picture change error is large for core properties. If the two-component result does not match the four-component reference even though the property operator has been properly transformed, then this is most likely because of other approximations discussed in this section.

6 Numerical comparison of the three exact-decoupling variants

For a detailed numerical one-to-one comparison of the three exact-decoupling methods, (infinite-order) DKH, X2C, and BSS, we have implemented them into the MOLCAS program (see Sect. 7). However, only scalar-relativistic versions of these methods are currently available in MOLCAS. For this reason, the explicit order does not need to be denoted. To also shed light on spin-orbit effects, we carried out atomic calculations for the radon atom as an example (see Sect. 7).

6.1 Scaling behavior

The efficiency of the three decoupling methods are compared in calculations of a test one-electron atomic system

(Rn⁸⁵⁺). 100, 200, and 300 even-tempered Gaussian basis functions were used. Since a 64-bit calculation turned out to be too fast for this test, all calculations were carried out with 128-bit precision (this also cures the failure of the diagonalization routine for large basis sets in 64-bit precision). The calculations were performed on the Opteron 250 CPU. The evaluation of one-electron integrals has not been included in the measurement of the CPU time, only the transformation steps have been counted. Since the calculations are dominated by matrix multiplications and diagonalization, all methods are of the order $\mathcal{O}(aN^2 + bN^3)$ where N denotes the number of basis functions. The formal scaling analysis is confirmed by the data given in Table 1, where the ratios are $7.8 \approx 2^3$ and $26.2 \approx 3^3$ for increasing the basis set from $N = 100$ to $N = 200$ and $N = 300$. This demonstrates that the diagonalization dominates the computational effort, i.e., that the prefactor a is rather small. The ratio between different methods is then a constant factor. The approximate DKH2 method is the fastest one, but it does, of course, not achieve *exact* decoupling. The computation time of the BSS method is almost the same as (still approximate) DKH8, while X2C is a little bit faster than BSS since the additional free-particle FW transformation is missing in the X2C approach.

6.2 The hydrogen-like Rn⁸⁵⁺ heavy ion

Table 2 presents results for the $1s$ state of the one-electron heavy ion Rn⁸⁵⁺ obtained with the different variants of exact-decoupling methods as well as the four-component Dirac equation solution denoted as DEQ. As explained above, we will use the abbreviation ‘DKH’ for results converged with respect to the order in the external potential. We employed a 35th order DKH scheme, i.e., DKH35, in all cases presented here and below, which yields results

Table 1 Computation times (in seconds) of various decoupling methods (for Rn⁸⁵⁺)

	$N = 100$	$N = 200$	$N = 300$
DKH2	1.8	12.6 (7.1)	41.5 (23.5)
DKH4	2.6	19.3 (7.3)	63.9 (24.3)
DKH6	4.4	33.2 (7.5)	110.8 (25.0)
X2C	7.0	54.2 (7.7)	181.1 (25.9)
DKH8	7.4	57.1 (7.8)	191.0 (26.0)
BSS	7.5	58.3 (7.8)	194.5 (26.0)
DKH10	12.6	97.8 (7.7)	326.0 (25.8)
DKH12	19.3	151.1 (7.8)	505.6 (26.2)
DKH14	29.5	230.3 (7.8)	771.4 (26.1)
DKH16	41.3	322.4 (7.8)	1,082.3 (26.2)

N the number of basis functions. Time ratios with respect to $N = 100$ are given in parentheses

Table 2 Energy eigenvalue and expectation values of *untransformed* operators $\{r^n, n = -2, -1, 1, 2\}$ of the $1s$ state in the one-electron heavy ion Rn^{85+}

	DEQ	DKH	X2C	BSS
ϵ_i	-4,154.6625406	-4,154.6625406	-4,154.6625406	-4,154.6625406
r^{-2}	31,523.20	50,897.75	51,129.21	51,391.38
r^{-1}	110.1254	126.3773	125.1675	126.4942
r	1.488335E-02	1.389296E-02	1.396132E-02	1.389185E-02
r^2	3.080482E-04	2.761091E-04	2.779321E-04	2.760894E-04

The energy eigenvalue ϵ_i is obtained to be exactly the same for all methods. However, note that the $\{r^n, n = -2, -1, 1, 2\}$ are affected by picture change errors to highlight the fact that two-component eigenfunctions are obtained that are identical only up to another unitary transformation. All data are in Hartree atomic units

well converged to the infinite-order result. All eigenvalues are identical to the original DEQ (in the same finite basis set) as one would have expected for exact-decoupling methods. Of course, this is possible because we do not invoke any approximations for the electron–electron interaction as this is absent in a one-electron system. The equivalence of the energy eigenvalues is, of course, not a surprising result since we have already discussed the equivalence of the three variants for exact decoupling for any finite basis set.

In order to study the difference of eigenfunctions, we report picture change error affected expectation values of radial momenta $\{r^n, n = -2, -1, 1, 2\}$. If we would properly transform the operators, all expectation values would be the same as they should be. Instead, we evaluate the expectation value as an integral of the squared two-component eigenfunction multiplied by the proper power of r, r^n . The lower the power n the more weight is given to the core part of the squared eigenfunction. From the data in Table 2, we can see that there exist discrepancies among different exact-decoupling methods which indicates that their eigenfunctions are indeed different. The fact that the differences observed for different operators r^n are of the same order indicates that discrepancies exist in the whole range of the two-component eigenfunctions. Also, the large difference between the four-component and two-component results shows that the picture change error is not negligible.

Table 3 reports additional (picture change error affected) expectation values for the operator r^{-1} for different states of Rn^{85+} . It can be seen that larger deviations from the DEQ reference result are found for core orbitals. The DKH results are found to be very close to BSS results, especially for the outer core orbitals, which may be due to the free-particle Foldy–Wouthuysen transformation that is the first step in both schemes. *We should stress again that, if the operators r^n would have been properly transformed or if the two-component eigenstates were back-transformed to the four-component picture, the exact decoupling methods yield expectation values are identical to the DEQ*

Table 3 Picture change error affected expectation values of the operator r^{-1} for ground and excited states of the Rn^{85+} hydrogen-like ion

	DEQ	DKH	X2C	BSS
$1s_{1/2}$	110.12545	126.37731	125.16747	126.49422
$1p_{1/2}$	29.275384	28.034662	27.975650	28.036643
$1p_{3/2}$	22.643647	23.005977	22.989564	23.006007
$1d_{3/2}$	10.237159	10.166282	10.165467	10.166282
$1d_{5/2}$	9.7717553	9.8145082	9.8138649	9.8145083
$1f_{5/2}$	5.5349079	5.5216606	5.5216008	5.5216606
$1f_{7/2}$	5.4423748	5.4519652	5.4519099	5.4519652

All values are in Hartree atomic units

reference which is the reason, why we did not report them in Tables 2 and 3.

This last point cannot be overemphasized. Within the finite basis set used, we obtain *exactly* the same energies and expectation values for all states and methods considered. It turned out that basis set convergence of the r^{-1} and r^{-2} momenta for one-electron heavy ions is much better than first observed in Ref. [91]. The persistent (small) deviation of high-order DKH results from the *exact* DEQ reference for these operators was believed in Ref. [91] to be an artifact of the finite basis set used. However, we discovered a bug in the original implementation of Ref. [91], which caused a wrong prefactor in front of one of the commutators of the third-order DKH property operator expression. As a consequence, the comparatively small deviation from the DEQ result completely vanishes upon correction of the erroneous prefactor indicating again the excellent convergence of the DKH series.

6.3 A many-electron case: the Rn atom

In Table 4, we present results of DFT calculations of the ground state of the neutral radon atom. The orbital energies of exact decoupling methods now differ from each other and from the four-component results due to the picture

Table 4 Total and selected B3LYP as well as Hartree–Fock orbital energies of the neutral Rn atom

	4c	DKH + SNSO	DKH	X2C	BSS
B3LYP results					
Total	−23,611.4636	−23,597.9850 (0.057)	−23,598.9075 (0.053)	−23,599.8921 (0.049)	−23,598.9087 (0.053)
1s _{1/2}	−3,619.9163	−3,612.3095 (0.210)	−3,611.4756 (0.233)	−3,612.2372 (0.212)	−3,611.4679 (0.233)
3s _{1/2}	−162.3668	−162.1464 (0.136)	−162.1294 (0.146)	−162.1402 (0.140)	−162.1294 (0.146)
6s _{1/2}	−0.8575	−0.8568 (0.089)	−0.8568 (0.090)	−0.8568 (0.086)	−0.8568 (0.090)
2p _{1/2}	−634.9644	−633.8152 (0.181)	−635.6840 (0.113)	−635.7755 (0.128)	−635.6833 (0.113)
2p _{3/2}	−533.8440	−533.1097 (0.138)	−532.3910 (0.272)	−532.4396 (0.263)	−532.3909 (0.272)
2p _{SO}	101.1203	100.7055 (0.410)	103.2929 (2.149)	103.3359 (2.191)	103.2923 (2.148)
6p _{1/2}	−0.4160	−0.4160 (0.006)	−0.4189 (0.683)	−0.4189 (0.693)	−0.4189 (0.682)
6p _{3/2}	−0.2789	−0.2787 (0.057)	−0.2779 (0.333)	−0.2779 (0.332)	−0.2779 (0.333)
6p _{SO}	0.1372	0.1373 (0.133)	0.1409 (2.747)	0.1410 (2.779)	0.1409 (2.746)
5d _{3/2}	−1.8692	−1.8723 (0.166)	−1.8862 (0.907)	−1.8861 (0.901)	−1.8862 (0.907)
5d _{5/2}	−1.7045	−1.7046 (0.008)	−1.6962 (0.484)	−1.6961 (0.492)	−1.6962 (0.484)
5d _{SO}	0.1648	0.1677 (1.793)	0.1900 (15.305)	0.1900 (15.310)	0.1900 (15.305)
4f _{5/2}	−8.2980	−8.3035 (0.066)	−8.3759 (0.938)	−8.3752 (0.931)	−8.3759 (0.938)
4f _{7/2}	−8.0377	−8.0455 (0.097)	−7.9930 (0.556)	−7.9924 (0.563)	−7.9930 (0.556)
4f _{SO}	0.2603	0.2580 (0.877)	0.3828 (47.080)	0.3828 (47.077)	0.3828 (47.080)
MAE		0.293	0.397	0.377	0.397
MRE (%)		0.088	0.392	0.392	0.392
Hartree–Fock results					
Total	−23,602.1044	−23,591.3125 (0.046)	−23,592.1434 (0.042)	−23,593.2115 (0.038)	−23,592.1283 (0.042)
1s _{1/2}	−3,641.1973	−3,635.6614 (0.152)	−3,634.9353 (0.172)	−3,635.5917 (0.154)	−3,634.9235 (0.172)
3s _{1/2}	−166.8331	−166.6451 (0.113)	−166.6291 (0.122)	−166.6378 (0.117)	−166.6289 (0.122)
6s _{1/2}	−1.0714	−1.0704 (0.094)	−1.0704 (0.091)	−1.0704 (0.090)	−1.0704 (0.091)
2p _{1/2}	−642.3301	−642.0151 (0.049)	−643.9010 (0.245)	−643.9966 (0.259)	−643.9001 (0.244)
2p _{3/2}	−541.1023	−540.5336 (0.105)	−539.8050 (0.240)	−539.8547 (0.231)	−539.8049 (0.240)
2p _{SO}	101.2278	101.4815 (0.251)	104.0960 (2.833)	104.1419 (2.879)	104.0952 (2.833)
6p _{1/2}	−0.5403	−0.5411 (0.151)	−0.5440 (0.688)	−0.5441 (0.700)	−0.5440 (0.688)
6p _{3/2}	−0.3840	−0.3836 (0.089)	−0.3827 (0.319)	−0.3827 (0.318)	−0.3827 (0.319)
6p _{SO}	0.1563	0.1575 (0.742)	0.1613 (3.162)	0.1613 (3.200)	0.1613 (3.161)
5d _{3/2}	−2.1897	−2.1940 (0.197)	−2.2085 (0.863)	−2.2084 (0.857)	−2.2085 (0.863)
5d _{5/2}	−2.0165	−2.0156 (0.047)	−2.0066 (0.491)	−2.0065 (0.498)	−2.0066 (0.491)
5d _{SO}	0.1731	0.1784 (3.037)	0.2019 (16.634)	0.2019 (16.640)	0.2019 (16.634)
4f _{5/2}	−9.1939	−9.2009 (0.076)	−9.2763 (0.896)	−9.2757 (0.889)	−9.2763 (0.896)
4f _{7/2}	−8.9282	−8.9313 (0.034)	−8.8764 (0.580)	−8.8758 (0.587)	−8.8764 (0.580)
4f _{SO}	0.2657	0.2696 (1.481)	0.3999 (50.506)	0.3999 (50.503)	0.3999 (50.506)
MAE		0.203	0.363	0.346	0.363
MRE (%)		0.077	0.396	0.395	0.396

The relative errors (in %) with respect to the four-component reference values are presented in parentheses. The mean absolute error (MAE) and mean relative error (MRE) are indicated for the complete set of occupied orbitals. The *SO* (spin–orbit splitting) entry is the energy difference between orbitals $j = l + 1/2$ and $j = l - 1/2$. All values are in Hartree atomic units

change error in the two-electron Coulomb interaction term. For the mean absolute errors (MAEs) of the complete set of occupied electron orbitals, the results of DKH, X2C, and BSS calculations are 0.397, 0.377 and 0.397 Hartree for B3LYP (and 0.363, 0.346 and 0.363 Hartree for Hartree–Fock), respectively. These deviations are basically the

same. All variants have the same mean relative errors (MREs) for all occupied orbitals with a value of 0.397% (B3LYP). The differences of the total energies are also at the same level. It is evident that, for the calculation of many-electron systems without a proper inclusion of the two-electron interaction operator, current exact-decoupling

methods provide the same accuracy for electronic energies, but do not fully reproduce the reference result.

Neglecting the picture change of the two-electron interaction operator also leads to large errors for the spin-orbit splitting, especially for orbitals of high-angular momentum. As we can see from Table 4, the relative errors of spin-orbit splitting of the $6p$, $5d$, and $4f$ orbitals are roughly 2.7, 15.3, and 47.1% for B3LYP (and 3.2, 16.6, and 50.5% for Hartree-Fock), respectively, when compared with the four-component reference. The spin-orbit splitting of the f shell has an error larger than 50% for B3LYP.

The SNSO [81] approach proposed by Boettger is a simple method to correct the two-electron picture change error and can be applied for atomic systems. We employed the same parameters for our SNSO infinite-order DKH calculation as Boettger did for the DKH2 approximation (it is obvious that such parameters are not very suitable for the exact-decoupling methods and we will improve on it in future work). We find that the DKH method with SNSO correction significantly reduces the relative errors of spin-orbit couplings, especially for the high-angular-momentum orbitals. The relative error of $6p$, $5d$, and $4f$ orbitals decreases to 0.1, 1.8, and 0.9% for B3LYP (and 0.7, 3.0, and 1.5% for Hartree-Fock), respectively. The MAE and MRE are also improved. The MRE of all occupied orbitals is decreased to 0.088% in the case of B3LYP and to 0.077% in the case of Hartree-Fock. However, the total energy is not improved because the contribution of high-angular momenta is too small.

The SNSO correction approach includes only the correction of spin-orbit terms and thus no scalar-relativistic terms are involved (by contrast to the approach of van Wüllen [42] mentioned earlier). The error in total electronic energy is mainly determined by the innermost $1s$ orbital, which accounts for 56% of the total energy error. If a scalar-relativistic picture change correction term can also be included to account for the deficiencies in the approximate treatment of the electron-electron interaction, we may expect that it will improve the total electronic energy as well as the inner core orbitals. If one is only interested in energy differences, as is usually the case in chemical applications, which may mainly come from the valence orbitals, a spin-orbit picture change correction scheme may be already good enough. The SNSO approach combined with the X2C or BSS methods yields the same results as for DKH. We therefore reported only the SNSO-DKH results in Table 4.

The analysis of energies does not provide a complete picture of the accuracy of the exact-decoupling methods. We therefore also report results of other properties. Again, we utilize data for the expectation values of operator r^{-1} as listed in Table 5. The leftmost column shows the results

without picture change correction. This leads to 5.2% relative error of the total quantities. The error mainly stems from the $1s$ orbital, which has 14.5% relative error. Even for the MRE of all occupied orbitals, it is 1.14% for B3LYP (and 1.13% for Hartree-Fock) and obviously not negligible.

Clearly, in the step of calculating the one-electron expectation values, the picture change correction, i.e., the unitary transformation of the $R^{-1} \equiv \sum_i r_i^{-1}$ operator, is necessary. As we can see from Table 5, the picture change correction reduces the relative error of total expectation value from 5.2 to 0.1%. The MAE and MRE are also reduced from 0.4857 and 1.144% to 0.0101 and 0.156% for B3LYP (and 0.4838 and 1.132% to 0.0069 and 0.051% for Hartree-Fock), respectively. There errors are close to or even better than those obtained for energies. Note that the picture change error not only affects the operator R^{-1} , but also the two-component spinors through the neglect of the transformation of V_{ee} in the optimization of the spinors. Adding the SNSO correction term improves further the accuracy of expectation values. However, the total expectation value is not improved for the same reason as discussed above. We even observe that the results for $2p_{1/2}$ become worse, although $2p_{3/2}$ turns out to be more accurate. For high-angular momentum orbitals d and f all results are improved.

6.4 Contact densities

Contact densities are most sensitive to the proper set-up of transformation of operators [93, 94, 98–100]. Picture change affected results are dramatically wrong. Especially, the contact density of $p_{1/2}$ would be zero without a proper treatment of the picture change. Our results for the contact density of the Rn atom in Table 6 show that all exact decoupling methods provide results of the same accuracy. Not unexpectedly, the SNSO correction does not improve the accuracy of the contact densities as it has been designed to correct the spin-orbit splitting, while the contact density is dominated by the $1s$ orbital, which has a contribution of more than 80%.

Since the RnH molecule considered for an illustrative calculation of relative contact densities with the Rn atom as reference is somewhat artificial (we unexpectedly obtained with Hartree-Fock and with B3LYP an increased contact density for the Rn contact density in the molecule; as can be seen for Hartree-Fock in Table 7), we also studied contact densities and contact-density shifts for the heavy water analog of the sixth period of the periodic table, i.e., for PoH_2 , for which we obtained negative shifts. The most important result from Table 7 is that DKH and X2C yield the same results for the contact-density shift and very similar absolute contact densities.

Table 5 B3LYP and Hartree-Fock expectation values of the r^{-1} operator for selected orbitals of the neutral Rn atom as well as for the total operator, $R^{-1} \equiv \sum_i r_i^{-1}$

	4c	DKH + SNSO	DKH	X2C	DKH [pce]
B3LYP results					
Total	699.505703	698.833118 (0.096)	698.944382 (0.080)	698.991685 (0.073)	735.502906 (5.146)
1s _{1/2}	109.179301	109.071618 (0.099)	109.059783 (0.109)	109.070963 (0.099)	125.051113 (14.537)
3s _{1/2}	9.969738	9.962444 (0.073)	9.961970 (0.078)	9.962248 (0.075)	10.466635 (4.984)
6s _{1/2}	0.695418	0.695001 (0.060)	0.694979 (0.063)	0.694997 (0.060)	0.699724 (0.619)
2p _{1/2}	27.194679	27.093849 (0.371)	27.200967 (0.023)	27.205183 (0.039)	26.138550 (3.884)
2p _{3/2}	21.081260	21.072395 (0.042)	21.043522 (0.179)	21.045230 (0.171)	21.347845 (1.265)
6p _{1/2}	0.579729	0.579411 (0.055)	0.580965 (0.213)	0.581003 (0.220)	0.579486 (0.042)
6p _{3/2}	0.492555	0.492325 (0.047)	0.491740 (0.166)	0.491751 (0.163)	0.492103 (0.092)
5d _{3/2}	1.247426	1.247309 (0.009)	1.250741 (0.266)	1.250737 (0.265)	1.249353 (0.154)
5d _{5/2}	1.202089	1.202033 (0.005)	1.199814 (0.189)	1.199806 (0.190)	1.200621 (0.122)
4f _{5/2}	2.974877	2.972309 (0.086)	2.982564 (0.258)	2.982533 (0.257)	2.979985 (0.172)
4f _{7/2}	2.929191	2.930927 (0.059)	2.923331 (0.200)	2.923299 (0.201)	2.925170 (0.137)
MAE		0.0083	0.0101	0.0099	0.4857
MRE (%)		0.067	0.156	0.155	1.144
Hartree-Fock results					
Total	699.258976	698.707959 (0.079)	698.819315 (0.063)	698.865968 (0.056)	735.341956 (5.160)
1s _{1/2}	109.178769	109.059584 (0.109)	109.048561 (0.119)	109.058925 (0.110)	125.027648 (14.516)
3s _{1/2}	9.980194	9.972547 (0.077)	9.972099 (0.081)	9.972330 (0.079)	10.475999 (4.968)
6s _{1/2}	0.676254	0.675837 (0.062)	0.675876 (0.056)	0.675890 (0.054)	0.680187 (0.582)
2p _{1/2}	27.149872	27.085486 (0.237)	27.192072 (0.155)	27.196654 (0.172)	26.131226 (3.752)
2p _{3/2}	21.071816	21.066569 (0.025)	21.037648 (0.162)	21.039426 (0.154)	21.341624 (1.280)
6p _{1/2}	0.574960	0.575095 (0.023)	0.576472 (0.263)	0.576515 (0.270)	0.575063 (0.018)
6p _{3/2}	0.492764	0.492489 (0.056)	0.491984 (0.158)	0.491998 (0.156)	0.492334 (0.087)
5d _{3/2}	1.243663	1.243946 (0.023)	1.247251 (0.288)	1.247249 (0.288)	1.245899 (0.180)
5d _{5/2}	1.198849	1.198650 (0.017)	1.196512 (0.195)	1.196505 (0.196)	1.197296 (0.130)
4f _{5/2}	2.975837	2.973897 (0.065)	2.984120 (0.278)	2.984090 (0.277)	2.981547 (0.192)
4f _{7/2}	2.930898	2.932343 (0.049)	2.924769 (0.209)	2.924738 (0.210)	2.926603 (0.147)
MAE		0.0069	0.0117	0.0114	0.4838
MRE (%)		0.051	0.170	0.170	1.132

The relative errors (in %) with respect to the four-component data are presented in parentheses. The mean absolute error (MAE) and mean relative error (MRE) are indicated for the complete set of occupied orbitals. 'pce' denotes picture-change affected results. All values are in Hartree atomic units

7 Conclusion and outlook

In this paper, we reviewed the current status of exact-decoupling methods applied to the relativistic Fock operator. Three different approaches—DKH, X2C, and BSS—exist for this purpose and they are all intimately related. In addition to this discussion of all important formal aspects, we then set out to provide numerical results which are obtained on the same basis (same program platform, same basis set, same electronic structure method). The main results of this study, of which some have already been obtained in previous work as cited above, may be summarized as follows.

(a) The exact decoupling of the relativistic Fock operator can be achieved with either DKH, X2C,

or BSS. The iteration scheme within the X2C and BSS approaches may suffer from convergence problems, but can be cured by the non-iterative scheme.

- (b) DKH, X2C, and BSS calculations in a finite basis set produce the same eigenvalues as the four-component reference (in the same KB basis set). However, their eigenfunctions may differ (but are related by to one another by a unitary transformation).
- (c) In many-electron calculations, the exact decoupling transformation is usually carried out with the external potential only, thus introducing an approximation. Employing untransformed two-electron potentials is computationally very beneficial, but also introduces a

Table 6 B3LYP and Hartree–Fock contact densities and contributions from individual orbitals of the neutral Rn atom in Hartree atomic units

	4c	DKH + SNSO	DKH	X2C	DKH [pce]
B3LYP results					
Total	3,813,828.11	3,812,599.15 (0.032)	3,814,303.57 (0.012)	3,815,690.02 (0.049)	11,155,986.64 (192.514)
1s _{1/2}	1,550,219.61	1,550,969.85 (0.048)	1,550,367.62 (0.010)	1,550,950.37 (0.047)	4,607,441.57 (197.212)
2s _{1/2}	248,537.74	248,697.85 (0.064)	248,617.89 (0.032)	248,689.67 (0.061)	739,534.03 (197.554)
3s _{1/2}	58,180.21	58,241.41 (0.105)	58,224.13 (0.075)	58,239.10 (0.101)	173,192.71 (197.683)
4s _{1/2}	15,342.95	15,361.35 (0.120)	15,356.94 (0.091)	15,360.79 (0.116)	45,680.35 (197.729)
5s _{1/2}	3,527.57	3,532.14 (0.129)	3,531.22 (0.103)	3,532.09 (0.128)	10,503.85 (197.764)
6s _{1/2}	551.01	551.73 (0.131)	551.61 (0.109)	551.74 (0.133)	1,640.81 (197.782)
2p _{1/2}	22,656.27	21,456.97 (5.293)	22,615.67 (0.179)	22,630.44 (0.114)	0.00 (100.000)
3p _{1/2}	5,966.57	5,655.97 (5.206)	5,956.62 (0.167)	5,959.76 (0.114)	0.00 (100.000)
4p _{1/2}	1,561.98	1,481.19 (5.173)	1,560.03 (0.125)	1,560.82 (0.075)	0.00 (100.000)
5p _{1/2}	331.14	314.11 (5.141)	330.97 (0.051)	331.13 (0.001)	0.00 (100.000)
6p _{1/2}	38.99	37.00 (5.100)	39.08 (0.222)	39.10 (0.274)	0.00 (100.000)
Hartree–Fock results					
Total	3,802,285.57	3,789,453.14 (0.337)	3,791,548.74 (0.282)	3,792,519.57 (0.257)	11,094,058.60 (191.773)
1s _{1/2}	1,546,350.48	1,542,225.22 (0.267)	1,541,792.70 (0.295)	1,542,205.45 (0.268)	4,583,791.56 (196.426)
2s _{1/2}	247,684.15	246,976.59 (0.286)	246,934.60 (0.303)	246,966.92 (0.290)	734,830.91 (196.681)
3s _{1/2}	57,987.93	57,844.73 (0.247)	57,836.71 (0.261)	57,842.34 (0.251)	172,111.81 (196.806)
4s _{1/2}	15,121.89	15,087.81 (0.225)	15,086.38 (0.235)	15,087.72 (0.226)	44,894.20 (196.882)
5s _{1/2}	3,339.99	3,332.53 (0.223)	3,332.59 (0.222)	3,332.88 (0.213)	9,917.13 (196.921)
6s _{1/2}	499.71	498.47 (0.248)	498.58 (0.225)	498.62 (0.217)	1,483.69 (196.912)
2p _{1/2}	22,383.11	21,341.25 (4.655)	22,482.12 (0.442)	22,507.49 (0.556)	0.00 (100.000)
3p _{1/2}	5,898.06	5,627.65 (4.585)	5,923.69 (0.435)	5,929.55 (0.534)	0.00 (100.000)
4p _{1/2}	1,528.81	1,459.34 (4.544)	1,536.26 (0.487)	1,537.73 (0.583)	0.00 (100.000)
5p _{1/2}	311.92	297.89 (4.499)	313.70 (0.570)	314.00 (0.667)	0.00 (100.000)
6p _{1/2}	36.74	35.11 (4.429)	37.04 (0.818)	37.08 (0.920)	0.00 (100.000)

The relative errors (in %) with respect to the four-component values are presented in parentheses. ‘pce’ denotes picture-change affected results

Table 7 Hartree–Fock contact densities for Rn and Po atoms and for the molecules RnH (bond distance 108.7 pm) and PoH₂ (Po–H distance 177.3 pm)

Method	Rn	RnH	Δ RnH	Po	PoH ₂	Δ PoH ₂
sfDKH [pce]	11,097,679.77	11,097,810.30	130.53	9,448,555.63	9,448,442.35	–113.28
sfDKH	3,744,321.59	3,744,366.39	44.80	3,197,017.33	3,196,979.51	–37.82
sfX2C	3,745,217.87	3,745,262.68	44.81	3,197,772.34	3,197,734.51	–37.83

Difference densities are denoted as Δ RnH = RnH – Rn, Δ PoH₂ = PoH₂ – Po. ‘sf’ denotes the scalar-relativistic spin-free approximations of the exact-decoupling methods. ‘pce’ denotes picture-change-affected results

picture change error, which may be compensated by an effective correction term.

- (d) The SNSO correction of the two-electron picture change error improves significantly the accuracy of the spin–orbit splitting. Since no scalar-relativistic correction is included in this ansatz, total expectation values are not improved.
- (e) Picture change corrections of property operators are mandatory—especially for core properties such as the contact density.

The discussion in this work has highlighted various directions of future developments for exact-decoupling methods. The basic theory is well established, but a couple of practical issues for actual calculations are still to be solved. Examples are:

- (a) The development of an intelligent infinite-order DKH method, which automatically truncates the expansion at desired accuracy without the pre-input of the desired order.

- (b) The efficiency of transformations of property operators should be improved. This is most apparent for the position-dependent density operator.
- (c) Two-electron picture change correction schemes for exact decoupling methods require more work. Developments could be based on ideas of the AMFI and SNSO approaches.

Acknowledgments This work has been financially supported by ETH Zürich and by SNF Grant No. 200020-132542/1.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix: Computational methodology

For the molecular calculations presented in this paper, we have been implemented the scalar-relativistic polynomial-cost DKH algorithm as well as X2C and BSS into the MOLCAS programme package [37] by docking the module described in Ref. [39] to the existing interface to the exponentially scaling DKH module [36]. Calculations with truly two-component versions are performed with the atomic program presented in Ref. [39], where also the four-component approach is available. Dyall's TZ basis set was employed for Rn and Po in uncontracted form yielding a (30s26p17d11f) primitive basis for both. For H, Dunning's aug-cc-pVTZ basis set was used (6s3p2d)/(4s3p2d) [101]. The (finite) nuclear charge distribution was modeled by a Gaussian distribution [102]. In the DFT calculations we applied the B3LYP hybrid density functional [103–105]. Different Rn–H distance were tested for RnH (from 171.6 to 108.7 pm) and the contact density turned out to be always positive. For PoH₂ we chose Po–H bond lengths of 177.3 pm and an angle of 89.6°.

References

- Dirac PAM (1958) The principles of quantum mechanics, 4th edn. Oxford University Press, Oxford
- Reiher M, Wolf A (2009) Relativistic quantum chemistry. Wiley-VCH, Weinheim
- Akhiezer AI, Berestetskii VB (1965) Quantum electrodynamics, vol XI of interscience monographs and texts in physics and astronomy. Interscience Publishers, New York
- Reiher M, Wolf A (2004) J Chem Phys 121:2037
- Reiher M, Hinze J (2003) Four-component ab initio methods for electronic structure calculations of atoms, molecules, and solids, volume relativistic effects in heavy-element chemistry and physics. Wiley, Chichester, pp 61–88
- Mittleman MH (1981) Phys Rev A 24:1167

- Sucher J (1980) Phys Rev A 22:348
- Heully J-L, Lindgren I, Lindroth E, Mårtensson-Pendrill A-M (1986) Phys Rev A 19:4426
- Talman JD (1986) Phys Rev Lett 57:1091
- LaJohn L, Talman JD (1992) Chem Phys Lett 189:383
- Thyssen J, Fleig T, Jensen HJA (2008) J Chem Phys 129:034109
- Chang C, Pelissier M, Durand P (1986) Phys Scr 34:394
- van Lenthe E, Baerends EJ, Snijders JG (1993) J Chem Phys 99:4597
- van Lenthe E, Baerends EJ, Snijders JG (1994) J Chem Phys 101:9783
- Hess BA (1986) Phys Rev A 33:3742
- Hess BA, Marian CM (2000) In: Jensen P, Bunker PR (eds) Computational molecular spectroscopy. Wiley, Chichester, p 169
- Barysz M, Sadlej AJ (2001) J Mol Struct (THEOCHEM) 573:181
- Wolf A, Reiher M, Hess BA (2002) Two-component methods and the generalized Douglas–Kroll transformation, theoretical and computational chemistry. Elsevier, Amsterdam, pp 622–663
- Wolf A, Reiher M, Hess BA (2004) Transgressing theory boundaries: the generalized Douglas–Kroll transformation. World Scientific, Singapore
- Reiher M (2006) Theor Chem Acc 116:241
- Liu W (2010) Mol Phys 108:1679
- Barysz M (2010) Two-component relativistic theories, vol 10 of challenges and advances in computational chemistry and physics. Springer, Dordrecht
- Reiher M (2011) WIREs computational molecular science. doi: 10.1002/wcms.67
- Nakajima T, Hirao K (2011) Chem Rev doi:10.1021/cr200040s
- Hinze J, Biegler-König F (1990) Numerical relativistic and non-relativistic MCSCF for atoms and molecules. Elsevier, Amsterdam, pp 405–446
- Reiher M, Hinze J (1999) J Phys B: At Mol Opt Phys 32:5489
- Heully J-L, Lindgren I, Lindroth E, Lundqvist S, Mårtensson-Pendrill A-M (1986) J Phys B 19:2799
- Kutzelnigg W (1997) Chem Phys 225:203
- Foldy LL, Wouthuysen SA (1950) Phys Rev 78:29
- Tani S (1951) Prog Theor Phys 6:267
- Douglas M, Kroll NM (1974) Ann Phys (NY) 82:89
- Hess BA (1985) Phys Rev A 32:756
- Nakajima T, Hirao K (2000) J Chem Phys 113:7786
- Wolf A, Reiher M, Hess BA (2002) J Chem Phys 117:9215
- van Wüllen C (2004) J Chem Phys 120:7307
- Reiher M, Wolf A (2004) J Chem Phys 121:10945
- Aquilante F et al (2009) J Comput Chem 31:224
- Werner H-J et al (2009) Molpro, version 2009.1, a package of ab initio programs, see <http://www.molpro.net>
- Peng D, Hirao K (2009) J Chem Phys 130:044102
- Samzow R, Hess BA, Jansen G (1992) J Chem Phys 96:1227
- Park C, Almlöf JE (1994) Chem Phys Lett 231:269
- van Wüllen C, Michauk C (2005) J Chem Phys 123:204113
- Seino J, Hada M (2008) Chem Phys Lett 461:327
- Luber S, Malkin Ondik I, Reiher M (2009) Chem Phys 356:205
- Barysz M, Sadlej AJ, Snijders JG (1997) Int J Quantum Chem 65:225
- Barysz M, Sadlej AJ (2002) J Chem Phys 116:2696
- Kutzelnigg W, Liu W (2005) J Chem Phys 123:241102
- Jensen HJA (2005, April) Talk on conference on relativistic effects in heavy elements—REHE 2005, Mülheim
- Kutzelnigg W, Liu W (2006) Mol Phys 104:2225
- Liu W, Kutzelnigg W (2007) J Chem Phys 126:114107
- Liu W, Peng D (2006) J Chem Phys 125:044102
- Peng D, Liu D, Xiao Y, Cheng L (2007) J Chem Phys 127:104106

53. Iliáš M, Saue T (2007) *J Chem Phys* 126:064102
54. Sikkema J, Visscher L, Saue T, Ilias M (2009) *J Chem Phys* 131:124116
55. Kim Y-K (1967) *Phys Rev* 154:17
56. Kutzelnigg W (1984) *Int J Quantum Chem* 25:107
57. Lee YS, McLean AD (1982) *J Chem Phys* 76:735
58. Ishikawa Y, Binning R, Sando K (1983) *Chem Phys Lett* 101:111
59. Stanton R, Havriliak S (1984) *J Chem Phys* 81:1910
60. Dyall KG, Grant IP, Wilson S (1984) *J Phys B At Mol Phys* 17:493
61. Dyall KG (1994) *J Chem Phys* 100:2118
62. Jansen G, Hess BA (1989) *Phys Rev A* 39:6016
63. Dyall KG (1997) *J Chem Phys* 106:9618
64. Dyall KG (1998) *J Chem Phys* 109:4201
65. Dyall KG, Enevoldsen T (1999) *J Chem Phys* 111:10000
66. Dyall KG (2001) *J Chem Phys* 115:9136
67. Dyall KG (2002) *J Comput Chem* 23:786
68. Filatov M, Dyall KG (2007) *Theor Chem Acc* 117:333
69. Filatov M, Cremer D (2003) *J Chem Phys* 119:11526
70. Filatov M, Cremer D (2005) *J Chem Phys* 122:064104
71. Filatov M (2006) *J Chem Phys* 125:107101
72. Kutzelnigg W, Liu W (2006) *J Chem Phys* 125:107102
73. Liu W, Peng D (2009) *J Chem Phys* 131:031104
74. Kędziera D, Barysz M (2007) *Chem Phys Lett* 446:176
75. Ilias M, Jensen HJA, Kellö V, Roos BO, Urban M (2005) *Chem Phys Lett* 408:210
76. Fukui H, Baba T, Shiraishi Y, Imanishi S, Kudo K, Mori K, Shimoji M (2004) *Mol Phys* 102:641
77. Seino J, Hada M (2007) *Chem Phys Lett* 442:134
78. Seino J, Uesugi W, Hada M (2010) *J Chem Phys* 132:164108
79. Peng D unpublished results
80. Hess BA, Marian CM, Wahlgren U, Gropen O (1996) *Chem Phys Lett* 251:365
81. Boettger JC (2000) *Phys Rev B* 62:7809
82. Hess BA, Kaldor U (2000) *J Chem Phys* 112:1809
83. Wolf A, Reiher M, Hess BA (2004) *J Chem Phys* 120:8624
84. Mayer M, Krüger S, Rösch N (2001) *J Chem Phys* 115:4411
85. Visscher L, van Lenthe E (1999) *Chem Phys Lett* 306:357
86. Peralta JE, Scuseria GE (2004) *J Chem Phys* 120:5875
87. Peralta JE, Uddin J, Scuseria GE (2005) *J Chem Phys* 122:084108
88. Thar J, Kirchner B (2009) *J Chem Phys* 130:124103
89. Baerends EJ, Schwarz WHE, Schwerdtfeger P, Snijders JG (1990) *J Phys B At Mol Phys* 23:3225
90. Kellö V, Sadlej AJ (1998) *Int J Quantum Chem* 68:159
91. Wolf A, Reiher M (2006) *J Chem Phys* 124:064103
92. Mastalerz R, Barone G, Lindh R, Reiher M (2007) *J Chem Phys* 127:074105
93. Mastalerz R, Lindh R, Reiher M (2008) *Chem Phys Lett* 465:157
94. Mastalerz R, Widmark P-O, Roos BO, Lindh R, Reiher M (2010) *J Chem Phys* 133:144111
95. Pernpointner M, Schwerdtfeger P (1998) *Chem Phys Lett* 295:347
96. Dyall KG (2000) *Int J Quantum Chem* 78:412
97. Kellö V, Sadlej AJ (2001) *J Mol Struct (THEOCHEM)* 547:35
98. Reiher M (2007) *Faraday Discuss* 135:97
99. Fux S, Reiher M (2011) *Struct Bonding*. doi: [10.1007/430_2010_37](https://doi.org/10.1007/430_2010_37)
100. Knecht S, Fux S, van Meer R, Visscher L, Reiher M, Saue T (2011) *Theor Chem Acc* 129:631
101. Dunning TH Jr (1989) *J Chem Phys* 90:1007
102. Visscher L, Dyall KG (1997) *At Data Nucl Data Tables* 67:207
103. Becke AD (1993) *J Chem Phys* 98:5648
104. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37:785
105. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) *J Phys Chem* 98:11623