Course Scheme

Introduction — some inadequacies of nonrelativistic theory. The course itself also attempts to provide an introduction to many aspects of the subject.

Relativistic — review of classical relativistic mechanics and dynamics, to provide background for relativistic quantum theory.

Quantum — quantization of the relativistic Hamiltonian, both first and second. Quantum electrodynamics is also introduced in the derivation of a Hamiltonian.

Chemistry — focussing in the end on what is of importance to quantum chemists, and eventually, giving evidence and understanding of effects of relativity on bonding, structure and energetics.

What's wrong with nonrelativistic theory?

Nonrelativistic theory works well for light elements, but fails to predict or adequately explain some phenomena for heavy elements:

- Trends in ionization potential and electron affinity with n and occupation number
- Spin-orbit splitting of p, d, ...subshells
- Spin and magnetic effects
- Stabilization of lower valence states for $6p$ block, and higher valence for $5d$ block
- Trends in dissociation energy with n

If accurate properties are required even for light elements, or spin-dependent properties are required, a relativistic theory of some description is essential.

Ionization potential trends

Ionization potential trends

Classical Mechanics

The equations of motion for a particle can be derived from the principle of least action. The action S is defined by

$$S = \int_{t_a}^{t_b} \mathcal{L}(q_n, \dot{q}_n, t)\mathrm{d}t$$

where $\mathcal{L}$ is the Lagrangian. The equations of motion are

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_n}\right) - \left(\frac{\partial \mathcal{L}}{\partial q_n}\right) = 0$$

The momentum conjugate to q_n is given by

$$p_n = \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_n}\right)$$

and the force by

$$F_n = \dot{p}_n = \left(\frac{\partial \mathcal{L}}{\partial q_n}\right).$$

The Hamiltonian is given by

$$\mathcal{H} = \sum_n \dot{q}_n p_n - \mathcal{L}$$

The value of the Lagrangian approach is that it does not invoke Galilean relativity except to define the specific form of the Lagrangian, so it may be transferred to use in special relativity.

Special relativity

We define two frames of reference, frame K which is at rest with respect to an observer, and frame K' which is moving with velocity $\mathbf{v}$ in the x direction. The Galilean coordinate transformation between frames is

$$x' = x - vt$$

Under such a transformation

- Newton's laws are invariant
- Maxwell's equations are **not** invariant

Consequently, either

- There is a preferred (absolute) frame of reference for electromagnetic phenomena, which should be experimentally observable,

or

- Newton's or Maxwell's equations need to be reformulated.

Experiments negated the former. Of the latter, it was found that it was Newton's laws which needed reformulation.

Special relativity

The basic postulates of special relativity are:

- All physical laws are invariant, i.e. there is **no** preferred frame of reference.
- The speed of light is the same in all frames.

The second of these may be expressed in terms of the distance travelled by light between two points in a given frame:

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 = 0$$
$$(x_2' - x_1')^2 + (y_2' - y_1')^2 + (z_2' - z_1')^2 - c^2(t_2' - t_1')^2 = 0$$

Assuming a transformation of general form between the coordinates, including the time coordinates, the Lorentz transformations between the frames K and K' may be deduced:

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}; \quad y' = y; \quad z' = z;$$
$$t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}$$

from which it follows that

- Length contracts with increasing speed
- Time dilates with increasing speed
- Length and time are interrelated

Special relativity

The symmetry between the length and time coordinates can be seen if we define a new length coordinate:

$$w = ct,$$

and dimensionless quantities β and γ

$$\beta = v/c, \quad \gamma = \frac{1}{\sqrt{1-\beta^2}}$$

and rewrite the Lorentz transformation as:

$$\begin{aligned} x' &= \gamma(x - \beta w) \\ w' &= \gamma(w - \beta x). \end{aligned}$$

Thus, length and time must always appear in the same way in physical laws.

Similarly, for a particle moving with velocity u in K and u' in K' the transformation is

$$\begin{aligned} u'_x &= (u_x - v)/(1 - u_x v/c^2) \\ u'_y &= u_y/\gamma(1 - u_x v/c^2) \\ u'_z &= u_z/\gamma(1 - u_x v/c^2) \end{aligned}$$

Note that here, the y and z components of velocity are affected, where in the coordinate transformations there was no effect. This is because of the appearance of the differential of the time in the velocity expression.

Special relativity

For a frame K' moving with arbitrary velocity $\mathbf{v}$ with respect to frame K, the general forms of the coordinate, time and velocity transformations are

$$\begin{aligned}\mathbf{r}' &= \gamma(\mathbf{r} - \mathbf{v}t) + (\gamma - 1)\frac{\mathbf{v} \times (\mathbf{v} \times \mathbf{r})}{v^2} \\ &= \mathbf{r} + \mathbf{v}[\frac{(\gamma - 1)\mathbf{v} \cdot \mathbf{r}}{v^2} - \gamma t]\end{aligned}$$

$$t' = \gamma(t - (\mathbf{v} \cdot \mathbf{r})/c^2)$$

$$\mathbf{u}' = \frac{\mathbf{u} - \mathbf{v}}{1 - \mathbf{u} \cdot \mathbf{v}/c^2} + \frac{\gamma - 1}{\gamma v^2} \cdot \frac{\mathbf{v} \times (\mathbf{v} \times \mathbf{u})}{1 - \mathbf{u} \cdot \mathbf{v}/c^2}$$

These reduce to the Galilean transformations in the limit $c \to \infty$ - this is the nonrelativistic limit of classical relativity.

4-vectors

Since $w = ct$ can be considered as a coordinate like the three space coordinates, and since the Lorentz transformation mixes all four, it is convenient to work in terms of 4-component vectors, or **4-vectors**. The conventions used are

- 4-vectors: normal roman type: a,
- Space-like (3-)vectors: bold face: $\mathbf{a}$,
- Scalar variables: italics: a.

The time-like component is written either as the zeroth component of a 4-vector, a_0, or the fourth component, a_4. In the latter position it is usually multiplied by i. Here we will adopt the former convention.

4-vectors

4-vectors are classified into *co*variant and *contra*variant forms.

- The components of *co*variant vectors transform like derivatives $\partial/\partial x$. Their indices are given as *sub*scripts, e.g. a_μ. The covariant form of a is $\mathrm{a}_\mu = (a_0, -\mathbf{a})$.
- The components of *contra*variant vectors transform like coordinates x. Their indices are given as *super*scripts, e.g. a^μ. The contravariant form of a is $\mathrm{a}^\mu = (a_0, \mathbf{a})$.

The notation a_μ represents the whole vector in terms of a general component of the vector. 4-vectors are usually written in row form for convenience, but they should be considered as column vectors in normal matrix algebra. Thus the contravariant form

$$\mathrm{a}^\mu = (a_0, \mathbf{a})$$

should be considered the equivalent of

$$\mathrm{a} = \begin{pmatrix} a_0 \\ \mathbf{a} \end{pmatrix}.$$

The covariant form can be considered a row vector. To avoid any possibility of confusion, the subscript or superscript will always be included when using the "4-vector" notation, and it will be omitted when using the "matrix" notation.

4-vectors

In expressions involving more than one 4-vector, repeated indices are summed over; one of the summation indices must be subscript and the other superscript. Thus the scalar product of two 4-vectors is

$$\mathrm{a}_\mu \mathrm{b}^\mu = a_0 b_0 - \mathbf{a} \cdot \mathbf{b}.$$

The transformation between contravariant and covariant forms is made by using the *metric tensor* $g_{\mu\nu}$.

$$\mathrm{g}_{\mu\nu} = \mathrm{g}^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Thus

$$\mathrm{a}_\mu = \mathrm{g}_{\mu\nu} \mathrm{a}^\nu; \quad \mathrm{a}^\mu = \mathrm{g}^{\mu\nu} \mathrm{a}_\nu.$$

4-vectors

The connection with with matrix algebra can now be seen. The scalar product of 4-vectors

$$\mathrm{a}_\mu \mathrm{b}^\mu = \mathrm{a}^\nu \mathrm{g}_{\nu\mu} \mathrm{b}^\mu$$

is equivalent in matrix form to the norm of a vector with the metric g

$$\mathrm{a}^T \ \mathrm{g} \ \mathrm{b},$$

or to the product of a co- and contravariant vector

$$\begin{pmatrix} a_0 & -\mathbf{a}^T \end{pmatrix} \begin{pmatrix} b_0 \\ \mathbf{b} \end{pmatrix}.$$

Note: Moss does not make use of the metric or the distinction between covariant and contravariant vectors. Instead, he adopts a simple index summation convention, and the negative sign is accounted for by the use of an imaginary time coordinate. However, the connection of this algebra with the normal matrix algebra is not obvious, because the scalar product of two vectors in Moss's scheme does not involve the conjugate transpose of one of the vectors.

4-vectors

We may now define a world point in terms of a position 4-vector

$$\mathrm{x}^\mu = (ct, x, y, z)$$

The squared length of this vector is

$$s^2 = \mathrm{x}_\mu \mathrm{x}^\mu = c^2 t^2 - x^2 - y^2 - z^2$$

This quantity is zero for the propagation of a photon from the origin, defining the "light cone".

- Positive values correspond to time-like 4-vectors, and negative values to space-like vectors.
- A Lorentz transformation cannot transform a spacelike vector into a time-like vector.
- The direction of time — future or past — is determined by the sign of ct, as expected.

Relativistic dynamics

The principle of least action can now be applied to obtain a relativistic Lagrangian. The action is taken to be a minimum between two world points. The conditions to be applied are

- The action must be independent of the frame of reference, and therefore must be a scalar,
- The integrand must be a scalar as well as a first-order differential.

The only scalar that fits is

$$\begin{aligned} \mathrm{d}s &= \sqrt{c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2} \\ &= c\ \mathrm{d}t\sqrt{1 - u^2/c^2} \end{aligned}$$

and the action is

$$S = k\int_{t_a}^{t_b} \mathrm{d}s.$$

The constant k is found to be $-mc$, and the Lagrangian is

$$\mathcal{L} = -mc^2\sqrt{1 - u^2/c^2}.$$

Relativistic dynamics

From the Lagrangian may be deduced the relativistic momentum

$$\mathbf{p} = \left(\frac{\partial \mathcal{L}}{\partial \mathbf{u}}\right) = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}}$$

and the Hamiltonian

$$\mathcal{H} = \mathbf{p} \cdot \mathbf{u} - \mathcal{L} = \frac{mc^2}{\sqrt{1 - u^2/c^2}}.$$

Squaring this and substituting for the velocity from the momentum expression,

$$\mathcal{H}^2 = m^2c^4 + c^2p^2$$

Defining the relativistic 4-momentum as

$$\mathrm{p}^\mu = (\mathcal{H}/c, \mathbf{p})$$

we see that the magnitude of the relativistic 4-momentum is conserved, since

$$\mathrm{p}_\mu \mathrm{p}^\mu = \mathcal{H}^2/c^2 - p^2 = (mc)^2$$

Note: In the definition of $\mathbf{p}$, the mass is sometimes taken to be a relativistically variant quantity, $M = m/\sqrt{1 - u^2/c^2}$. Then the Hamiltonian, or energy, is simply $\mathcal{H} = E = Mc^2$. Here m will always be taken to be the rest mass.

Electromagnetic interactions

So far we have dealt only with a free particle. The introduction of electromagnetic fields requires an addition to the action. The additional term is found to be

$$S_{em} = -q \int_a^b \mathrm{A}_\mu \mathrm{dx}^\mu$$

where q is the charge and A_μ is the covariant 4-potential

$$\mathrm{A}_\mu = (\phi/c, -\mathbf{A}).$$

Converting the differential to one involving only the time,

$$\mathrm{dx}^\mu = \mathrm{d}t(c, \mathbf{u})$$

we may define a new Lagrangian,

$$\mathcal{L} = -mc^2\sqrt{1 - u^2/c^2} + q\mathbf{A} \cdot \mathbf{u} - q\phi.$$

Electromagnetic interactions

The canonical momentum derived from this Lagrangian is

$$\mathbf{p} = \left(\frac{\partial \mathcal{L}}{\partial \mathbf{u}}\right) = \frac{m\mathbf{u}}{\sqrt{1 - u^2/c^2}} + q\mathbf{A} = \boldsymbol{\pi} + q\mathbf{A}$$

where $\boldsymbol{\pi}$ is the mechanical momentum, and the Hamiltonian is

$$\mathcal{H} = \mathbf{p} \cdot \mathbf{u} - \mathcal{L} = \frac{mc^2}{\sqrt{1 - u^2/c^2}} + q\phi,$$

which, when replacing u using the canonical momentum, gives

$$\mathcal{H} = c\sqrt{m^2c^2 + [\mathbf{p} - q\mathbf{A}]^2} + q\phi$$

where the square of the vector quantity is to be taken as a scalar product.

Electromagnetic fields

The electric field strength $\mathbf{E}$ and the magnetic flux density $\mathbf{B}$ are related to the potentials by

$$\mathbf{B} = \nabla \times \mathbf{A}$$

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}$$

These can be seen to form the components of an antisymmetric contravariant second-rank 4-tensor F

$$\mathrm{F}^{\mu\nu} = \partial^{\nu}\mathrm{A}^{\mu} - \partial^{\mu}\mathrm{A}^{\nu}$$

where ∂_{μ} is the covariant gradient 4-vector

$$\partial_{\mu} = (\frac{\partial}{\partial(ct)}, \nabla)$$

This tensor is the electromagnetic field tensor, and can be written explicitly as

$$\mathrm{F}^{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix}$$

It gives rise to a relativistic invariant:

$$\mathrm{F}_{\mu\nu}\mathrm{F}^{\mu\nu} = 2(B^2 - E^2/c^2)$$

which is related to the Lagrangian for an electromagnetic field:

$$\mathcal{L}_{em} = -\frac{1}{2}\epsilon_0(c^2B^2 - E^2)$$

Electromagnetic fields

Straightforward application of the gradient operator on the definition of $\mathbf{E}$ and $\mathbf{B}$ give two of Maxwell's equations,

$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{E} = -\left(\frac{\partial \mathbf{B}}{\partial t}\right)$$

which may be written in terms of the electromagnetic field tensor,

$$\frac{\partial F_{\mu\nu}}{\partial x_\xi} + \frac{\partial F_{\nu\xi}}{\partial x_\mu} + \frac{\partial F_{\xi\mu}}{\partial x_\nu} = 0.$$

The remaining pair of equations is

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0$$
$$\nabla \times \mathbf{B} = \frac{1}{c^2}\left[\frac{\mathbf{j}}{\epsilon_0} + \frac{\partial \mathbf{E}}{\partial t}\right]$$

where ρ is the charge density and $\mathbf{j}$ is the current flux. These may also be expressed in terms of 4-vectors and tensors as

$$\partial_\mu \mathrm{F}^{\mu\nu} = -\mathrm{j}^\nu/c^2\epsilon_0$$

where the contravariant 4-current is

$$\mathrm{j}^\nu = (c\rho, \mathbf{j})$$

Gauge transformations

The electromagnetic field vectors are defined up to a gauge transformation by the vector potential. If we make the transformation

$$\tilde{\mathrm{A}}^{\mu} = \mathrm{A}^{\mu} + \partial^{\mu} f$$

which is equivalent to

$$\tilde{\mathbf{A}} = \mathbf{A} - \nabla f; \quad \tilde{\phi} = \phi + \frac{\partial f}{\partial t}$$

it can be shown that $\mathbf{E}$ and $\mathbf{B}$ are invariant to this transformation. The Lagrangian changes by

$$-q\mathbf{u} \cdot \nabla f - q\left(\frac{\partial f}{\partial t}\right) = -q\left(\frac{\mathrm{d}f}{\mathrm{d}t}\right).$$

The equations of motion are not altered by addition of a total time derivative, but of course since the Lagrangian is different in form, so will the Hamiltonian be also. The choice of gauge is therefore important for the form of the equations to be used. The common gauge choice is the Coulomb gauge,

$$\nabla \cdot \mathbf{A} = 0.$$

Another choice, which is relativistically invariant, is the Lorentz (or Feynman) gauge

$$\partial_{\mu} \mathrm{A}^{\mu} = \nabla \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t} = 0.$$

Relativistic potentials

For a stationary point charge, the potential is given by

$$\mathbf{A} = 0, \qquad \phi = q/4\pi\epsilon_0 r$$

The potential seen by an observer due to a moving charge is obtained by a Lorentz transformation of the 4-potential, which gives

$$\mathbf{A}' = \mathbf{u}\phi/c^2, \qquad \phi' = \gamma\phi$$

The coordinates in ϕ are given with respect to the frame of the moving charge; we require them in the frame of the observer, which is achieved by the following method.

- In order for the observer to observe the charge and hence its potential, a signal must be transmitted from the charge to the observer, which propagates at velocity c.
- The time of observation of the signal is taken to be $t = t' = 0$. The signal was therefore emitted at times $t = -r/c$ and $t' = -r'/c$.
- At these times, a Lorentz transformation may be performed to relate the two:

  $$t = \gamma(t' - \mathbf{u}\cdot\mathbf{r}'/c^2).$$

  Therefore the distance in the frame of the charge is

  $$r = -ct = \gamma(r' + \mathbf{u}\cdot\mathbf{r}'/c)$$

- The potential can then be expressed in the observer's frame as

  $$\phi = q/4\pi\epsilon_0(r' + \mathbf{u}\cdot\mathbf{r}'/c)$$

Relativistic potentials

The potential derived above is called the Liénard-Wiechert potential, and is expressed in terms of the distance to the charge at the time of emission of the signal, not at the time of observation, for which the velocity of the charge may have changed. The potential is thus expressed in terms of the so-called *retarded* values of $\mathbf{r}$ and $\mathbf{u}$, i.e. at a time earlier than that of observation.

The concept of the finite speed of propagation of potentials is a fundamental one in relativistic electrodynamics, and will be met again later; in particular the propagation of potentials at the speed of light.

Relativistic interactions

The interaction between two charged particles can be derived from the Lagrangian. Making use of the expression for $\mathbf{A}$ from above, the Lagrangian for particle 1 is

$$\mathcal{L} = -m_1 c^2 \sqrt{1 - u_1^2/c^2} + q_1 \phi(\mathbf{u}_2 \cdot \mathbf{u}_1/c^2 - 1).$$

Addition of a term involving particle 2 only cannot affect the equation of motion for particle 1, so we may write

$$\begin{aligned}\mathcal{L} = &-m_1 c^2 \sqrt{1 - u_1^2/c^2} - m_2 c^2 \sqrt{1 - u_2^2/c^2} \\ &+ q_1 q_2 V(\mathbf{u}_2 \cdot \mathbf{u}_1/c^2 - 1).\end{aligned}$$

Now, $Vq_2 = \phi$ is the retarded potential for particle 1 due to particle 2, and as Vq_1 is not in general equal to the retarded potential for particle 2 due to particle 1, it cannot be the complete Lagrangian for the interacting system.

Relativistic interactions

An approximate Lagrangian may be derived by assuming

- The velocities do not change significantly during the time of transmission, therefore the Lorentz transformation can be performed at the time of observation, giving

$$\phi = \frac{\gamma q}{4\pi\epsilon_0 r}\left[1 + \left(\frac{\gamma \mathbf{u}\cdot\mathbf{r}}{cr}\right)^2\right]^{-1/2}$$

- The velocities are small compared to c so that an expansion with truncation at low order is valid. This gives

$$\mathbf{A} = \frac{q}{4\pi\epsilon_0 cr}[\mathbf{u}/c + \mathcal{O}((u/c)^3)]$$
$$\phi = \frac{q}{4\pi\epsilon_0 r}[1 + u^2/2c^2 - (\mathbf{u}\cdot\mathbf{r})^2/2c^2r^2 + \mathcal{O}((u/c)^4)]$$

- The Coulomb gauge is employed; with a gauge transformation with $f = q(\mathbf{u}\cdot\mathbf{r})/2c^2r$ the potentials are

$$\mathbf{A} = \frac{q}{8\pi\epsilon_0 cr}[\mathbf{u}/c + (\mathbf{u}\cdot\mathbf{r})\mathbf{r}/r^2c + \mathcal{O}((u/c)^3)]$$
$$\phi = \frac{q}{4\pi\epsilon_0 r}[1 + \mathcal{O}((u/c)^4)]$$

Relativistic interactions

Making use of these potentials, the Lagrangian can now be written as

$$\mathcal{L} = -m_1 c^2 \sqrt{1 - u_1^2/c^2} - m_2 c^2 \sqrt{1 - u_2^2/c^2}$$
$$- \frac{q_1 q_2}{4\pi\epsilon_0} \left[\frac{1}{r} - \frac{\mathbf{u}_1 \cdot \mathbf{u}_2}{c^2 r} - \frac{(\mathbf{u}_1 \cdot \mathbf{r})(\mathbf{u}_2 \cdot \mathbf{r})}{c^2 r^3} + \mathcal{O}((u/c)^4) \right]$$

It is obvious that this Lagrangian, derived for particle 1 and its interaction with particle 2, is unchanged on interchange of the coordinates of the two particles, and therefore will serve for both. It is a total Lagrangian for the system.

From this Lagrangian, the momenta of the two particles and the Hamiltonian may be derived. The inversion of the expression for the momenta to obtain the velocities for substitution in the Hamiltonian expression is complicated; since we have already approximated the potentials, this inversion may also be approximated to an appropriate order, giving the approximate Hamiltonian

$$\mathcal{H} = \sum_{i=1,2} [m_i c^2 + p_i^2/2m_i - p_i^4/8m_i^3 c^2] + \frac{q_1 q_2}{4\pi\epsilon_0 r}$$
$$- \frac{q_1 q_2}{8\pi\epsilon_0 m_1 m_2 c^2} \left[\frac{\mathbf{p}_1 \cdot \mathbf{p}_2}{r} + \frac{(\mathbf{p}_1 \cdot \mathbf{r})(\mathbf{p}_2 \cdot \mathbf{r})}{r^3} \right].$$

which is correct to approximately $\mathcal{O}(mc^2(p/mc)^6)$.

Relativistic interactions

For a charge distribution it is more useful to derive the contribution to the Lagrangian from the current 4-vector. This contribution is

$$\mathcal{L}_{int} = \int \mathrm{j}_\mu \mathrm{A}^\mu \ \mathrm{d}\tau$$

where the integration is taken over the spatial variables. Note that charge conservation is given by the equation

$$\partial_\mu \ \mathrm{j}^\mu = \nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0.$$

Quantization and invariance conditions

Quantum mechanical equations are obtained from the classical equations by use of the correspondence principle,

$$\mathrm{p}_\mu \to i\hbar\partial_\mu \quad \equiv \quad \mathcal{H} \to i\hbar\frac{\partial}{\partial t}; \quad \mathbf{p} \to -i\hbar\nabla.$$

In the transition from classical to quantum mechanics, we still require that the equations are invariant to a Lorentz transformation, and also are gauge invariant. The Lorentz invariance is guaranteed if the equations can be expressed in terms of 4-vectors and relativistic invariants. The gauge invariance is assured if the wave function in the new gauge is multiplied by a phase factor,

$$\tilde{\psi} = \psi \exp(-iqf/\hbar),$$

provided also that the equations are expressed in terms of 4-vectors, since this ensures the cancellation of the phase factors.

Quantization

The nonrelativistic Hamiltonian in a static potential is

$$\mathcal{H} = T + V = p^2/2m + q\phi.$$

Using the correspondence principle, we get the time-dependent Schrödinger equation

$$-\frac{\hbar^2}{2m}\left[\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} + \frac{\partial^2\Psi}{\partial z^2} + V\Psi\right] = i\hbar\frac{\partial\Psi}{\partial t}.$$

This equation has x, y and z appearing *quadratically*, but t appearing *linearly*, which violates the principle of equivalence of spatial and temporal variables, and thus cannot be Lorentz invariant.

The relativistic Hamiltonian in a static potential is

$$\mathcal{H} = c\sqrt{m^2c^2 + p^2} + V$$

Use of the correspondence principle gives

$$[mc^2\sqrt{1 + \nabla^2/(mc)^2} + V]\Psi = i\hbar\frac{\partial\Psi}{\partial t}$$

This presents a problem: how to interpret the square root?

Perturbation series approach

One possible approach is to expand the square root in a power series. Thus, the Hamiltonian function becomes

$$\mathcal{H} = mc^2 + p^2/2m - p^4/8m^3c^2 + \ldots + V$$

The first two terms in the expansion are the rest energy and the nonrelativistic kinetic energy.

But does this expansion converge? The series is only valid if $p < mc$, which is not the case if the velocity exceeds $c/\sqrt{2}$, which is bound to happen to a particle moving in a Coulomb potential.

The expansion of the Hamiltonian in terms of the velocity *is* valid, but the substitution of the velocity in terms of the momentum must still be made in order to use the correspondence principle, and this leaves the momentum in the denominator. In this case we would arrive at an operator

$$\mathcal{H} = mc^2 + \frac{p^2}{2m[1 + p^2/m^2c^2]} + \frac{3p^4}{8m^3c^2[1 + p^2/m^2c^2]^2} + \ldots + V$$

Then the problem is to interpret, or apply, the inverse of the momentum operator.

The Klein-Gordon equation

The use of the Hamiltonian with its square-root operator, in its original or expanded form, does not transparently display the equivalence of length and time coordinates, and hence the Lorentz invariance of the Hamiltonian.

An alternative is to take the Hamiltonian in squared form. The relativistic invariance of the equations is better displayed if a general electromagnetic potential is considered. The classical form is

$$(\mathcal{H} - q\phi)^2 = (mc^2)^2 + c^2\left(\mathbf{p} - q\mathbf{A}\right)^2 .$$

Use of the correspondence principle gives the Klein-Gordon equation

$$\left[(i\hbar\frac{\partial}{\partial t} - q\phi)^2 - c^2(-i\hbar\nabla - q\mathbf{A})^2\right]\Psi = (mc^2)^2\Psi$$

or in 4-vector form,

$$c^2(i\hbar\partial_\mu - q\mathrm{A}_\mu)(i\hbar\partial^\mu - q\mathrm{A}^\mu)\Psi = (mc^2)^2\Psi$$

The Klein-Gordon equation

The Klein-Gordon equation is a second-order differential equation, with the following properties:

- It is manifestly Lorentz invariant.
- For a particle of zero mass and charge, it reduces to the electromagnetic field equations.
- Because of the square, it admits solutions of both positive and negative mass/energy.
- The square also means that for a single particle, the potentials are two-body potentials, and for many particles, a four-body potential would be introduced. This is clearly not useful!
- $\Psi^*\Psi$ is in general time-dependent, and therefore cannot be interpreted as a probability density.
- The charge is conserved but the particle number is not.
- It does not introduce the concept of spin.

Thus, although it has been used successfully to describe mesons, it is not generally useful as a relativistic quantum mechanical equation.

The Dirac equation

Dirac postulated that the argument of the square root operator could be written as a perfect square,

$$\pi^2 + m^2c^2 = (\boldsymbol{\alpha} \cdot \boldsymbol{\pi} + \beta mc)^2 \,,$$

where $\boldsymbol{\pi}$ is the mechanical momentum operator. Thus, the relativistic Hamiltonian can be written

$$\mathcal{H} - q\phi = c\boldsymbol{\alpha} \cdot (\mathbf{p} - q\mathbf{A}) + \beta mc^2$$

Then using the correspondence principle the Dirac equation is

$$\left(i\hbar\frac{\partial}{\partial t} - q\phi\right)\Psi = c\boldsymbol{\alpha} \cdot (-i\hbar\nabla - q\mathbf{A})\,\Psi + \beta mc^2\Psi$$

which may be displayed in Lorentz invariant form as

$$c\alpha^\mu(i\hbar\partial_\mu - q\mathrm{A}_\mu)\Psi = \beta mc^2\Psi$$

where

$$\alpha^\mu = (1, \boldsymbol{\alpha}).$$

Comparison of the free-particle Dirac Hamiltonian, $c\boldsymbol{\alpha} \cdot \mathbf{p} + \beta mc^2$, with the classical relativistic Hamiltonian, suggests that $c\boldsymbol{\alpha}$ should be interpreted as the *velocity* operator.

The Dirac matrices

In order for the perfect square to be obtained, the quantities $\boldsymbol{\alpha}$ and β must obey the following:

$$\begin{aligned} \alpha_i^2 &= \beta^2 = 1 \\ \alpha_i \alpha_j &= -\alpha_j \alpha_i, \ i \neq j, \\ \alpha_i \beta &= -\beta \alpha_i. \end{aligned}$$

or, with $\beta = \alpha_4$,

$$\alpha_i \alpha_j + \alpha_j \alpha_i = 2\delta_{ij}$$

- Since the algebra of the α matrices is non-commutative, scalars cannot satisfy these relations. Therefore an appropriate non-commutative algebra must be found to satisfy these conditions.
- Matrix algebra is non-commutative, and hence is appropriate. The minimum matrix rank required by these conditions is 4.
- An alternative is to use a Clifford algebra with the appropriate order. In this case it would have to be order 8, i.e. one would use octernions.

The use of a Clifford algebra will be returned to later. The standard choice is that of the 4×4 matrices.

The Dirac matrices

The matrix representation of α and β is

$$\alpha_k = \begin{pmatrix} 0_2 & \sigma_k \\ \sigma_k & 0_2 \end{pmatrix}, \; k = 1, 2, 3; \quad \beta = \begin{pmatrix} 1_2 & 0_2 \\ 0_2 & -1_2 \end{pmatrix}$$

where 1_2 and 0_2 are the 2×2 unit and null matrices, and σ_k are the Pauli spin matrices,

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

The 4-vector a^μ must therefore be a vector of 4×4 matrices, and the zeroth component must be interpreted as 1_4, the 4×4 unit matrix. Note that the 4-vector does *not* contain β; it is defined as α_4 only for the purposes of the algebra above.

This representation of the alpha matrices is called the *standard representation*, and is defined only up to a similarity transformation,

$$\alpha'_\mu = \mathbf{Q}^{-1} \alpha_\mu \mathbf{Q}$$

which for invariance requires that

$$\Psi' = \mathbf{Q}^{-1} \Psi$$

The Dirac wave function

Because of the appearance of 4×4 matrices in the Dirac equation, the wave function must be a 4-component vector:

$$\Psi = \begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} \Psi^L \\ i\Psi^S \end{pmatrix}$$

- Ψ^L and Ψ^S are termed the *large* and *small* components of the wave function. The reason for this nomenclature will become clear later.
- The factor of i has been introduced for convenience in atomic calculations. At other stages it will not be introduced.
- Spin has been introduced in vector form. Components 0 and 2 represent α spin, and components 1 and 3 represent β spin.
- Ψ^L and Ψ^S are two-component spinors, or 2-spinors, also simply called spinors. Ψ is a four-component spinor, or 4-spinor, also called a bispinor.

The Dirac and Klein-Gordon equations

Writing the Dirac equation as

$$[c\alpha^\mu \pi_\mu - \beta mc^2]\Psi = 0$$

and the Klein-Gordon equation as

$$c^2 \pi_\mu \pi^\mu \Psi = (mc^2)^2 \Psi$$

we can see that the two are related, by operating on the Dirac equation with the operator corresponding to a particle of negative mass

$$[c\alpha_\mu \bar{\pi}^\mu + \beta mc][c\alpha^\mu \pi_\mu - \beta mc^2]\Psi = 0.$$

where

$$\bar{\pi}^\mu = (E/c, -\mathbf{p}).$$

Thus a solution of the Dirac equation is also a solution of the Klein-Gordon equation.

The time-independent Dirac equation

For a stationary state, the temporal part of the Dirac equation can be factored out just as in the Schrödinger equation, to leave a time-independent Dirac equation. This must be done in a particular frame of reference, which is chosen to be the Born-Oppenheimer frame. In a static potential $\mathbf{A} = 0$,

$$-i\hbar c\alpha.\nabla\Psi + (\beta mc^2 + V)\Psi = E\Psi$$

Writing Ψ in terms of large and small components,

$$\begin{pmatrix} (V - E + mc^2) & -i\hbar c(\sigma.\nabla) \\ \\ -i\hbar c(\sigma.\nabla) & (V - E - mc^2) \end{pmatrix} \begin{pmatrix} \Psi^L \\ \\ i\Psi^S \end{pmatrix} = 0$$

or as a pair of coupled equations,

$$(V - E + mc^2)\,\Psi^L + \hbar c(\sigma.\nabla)\,\Psi^S = 0,$$

$$-\hbar c(\sigma.\nabla)\Psi^L + (V - E - mc^2)\,\Psi^S = 0.$$

Nonrelativistic limit of the Dirac equation

If we eliminate the small component using

$$\Psi^S = \hbar c(V - E - mc^2)^{-1}(\sigma.\nabla)\Psi^L,$$

we obtain an equation for the large component,

$$(V - E + mc^2)\Psi^L = -\hbar^2 c^2(\sigma.\nabla)(V - E - mc^2)^{-1}(\sigma.\nabla)\Psi^L$$

Noting that

$$(\mathbf{A} - \mathbf{B})^{-1} = -\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{A}(\mathbf{A} - \mathbf{B})^{-1}$$

we may write, with $2mc^2 = \mathbf{B}$ and $(V - E + mc^2) = \mathbf{A}$,

$$[(V - E + mc^2) - \frac{\hbar^2}{2m}(\sigma.\nabla)(\sigma.\nabla)]\Psi^L$$

$$= \frac{\hbar^2}{2m}(\sigma.\nabla)\left(\frac{V - E + mc^2}{V - E - mc^2}\right)(\sigma.\nabla)\Psi^L$$

Noting that $(\sigma.\nabla)^2 = 1_2\nabla^2$, we arrive at the equation

$$(T + V - E + mc^2)\Psi^L = \frac{\hbar^2}{2m}(\sigma.\nabla)\left(\frac{V - E + mc^2}{V - E - mc^2}\right)(\sigma.\nabla)\Psi^L$$

Nonrelativistic limit of the Dirac equation

The left hand side (LHS) of this equation is the Schrödinger equation with the rest mass included. To align the energy scale, a phase factor may be introduced into the wave function (which is only defined up to an arbitrary phase)

$$\Psi' = \Psi \exp(-imc^2 t/\hbar)$$

so that

$$E' = E - mc^2$$

and

$$\beta' = \beta - 1_4$$

Then we may write

$$(T + V - E')\Psi^L = \frac{\hbar^2}{2m}(\sigma.\nabla)\left(\frac{V - E'}{V - E' - 2mc^2}\right)(\sigma.\nabla)\Psi^L$$

The right hand side (RHS) now clearly vanishes as $c \to \infty$, and the LHS reduces to the Schrödinger equation. Thus, the large component of the wave function reduces to the nonrelativistic wave function in spinor form in the limit of infinite velocity of light.

One note of caution needs to be added to this statement. If the potential V is a Coulomb potential, $V \sim 1/r$, the RHS does not vanish for infinite c at $r = 0$. Thus for a point nuclear model, the nonrelativistic limit is only obtained asymptotically.

Nonrelativistic limit of the Dirac equation

Rewriting the small component equation, the nonrelativistic limit for the small component is

$$\lim_{c\to\infty} c\Psi^S = -\frac{\hbar}{2m}(\sigma.\nabla)\Psi^L.$$

Thus the small component vanishes in the nonrelativistic limit. The same proviso applies here regarding a Coulomb potential.

Taking for example the NR hydrogenic atom wave function as an approximation to Ψ^L, we find that

$$\Psi^S \sim \frac{Z}{2mc}\Psi^L.$$

thus giving meaning to the names "large" and "small" components.

Dirac wave function for atoms

For atoms, the wave function may be factored in spherical polar coordinates

$$\Psi = \frac{1}{r} \begin{pmatrix} P_{n\kappa}(r)\, \chi_{\kappa,m}(\theta,\phi) \\ iQ_{n\kappa}(r)\, \chi_{-\kappa,m}(\theta,\phi) \end{pmatrix}$$

P and Q are the *radial* large and small components of the wavefunction; the angular functions χ are 2-component spinors:

$$\chi_{\kappa,m}(\theta,\phi) = \sum_{m_s=\pm 1/2} \langle \ell m_\ell s m_s | j m \rangle \; Y_\ell^{m_\ell}(\theta,\phi)\; \eta(m_s)$$

$\langle \ell m_\ell s m_s | j m \rangle$ is a Clebsch-Gordan coefficient, $Y_\ell^{m_\ell}(\theta,\phi)$ is a spherical harmonic, defined with the Condon and Shortley phase conventions,

$$Y_\ell^{m_\ell}(\theta,\phi) = (-1)^{m_\ell + |m_\ell|} P_\ell^{m_\ell}(\cos\theta) e^{i m_\ell \phi}$$

and the spin functions η are

$$\eta(1/2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad \eta(-1/2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Atomic quantum numbers

For atoms, the wavefunction is an eigenfunction of $\hat{H}$, $\hat{j}^2$, $\hat{j}_z$, $\beta\hat{\mathcal{P}}$, and $\beta\hat{\mathcal{K}}$, with quantum numbers n, j, m, $(-1)^\ell$ and κ respectively. $\hat{\mathcal{P}}$ is the inversion operator, and $\hat{\mathcal{K}}$ is defined by

$$\hat{\mathcal{K}} = \hat{\ell}^2 + \hat{s}^2 - \hat{j}^2 - 1 = -1 - \boldsymbol{\sigma} \cdot \hat{\boldsymbol{\ell}}$$

$$\hat{\mathcal{K}}\, \chi_{\kappa,m}(\theta,\phi) = \kappa\, \chi_{\kappa,m}(\theta,\phi)$$

$\hat{\mathcal{K}}$ is thus related to the spin-orbit interaction operator. The values of κ are related to j and ℓ:

$$\kappa = 2(\ell - j)(j + 1/2)$$

$2(j - \ell) = \pm 1$ is given the symbol a. Values of j, ℓ and κ are given below:

ℓ	0	1	1	2	2	3	3
j	1/2	1/2	3/2	3/2	5/2	5/2	7/2
κ	-1	1	-2	2	-3	3	-4
label	s	$\bar{p}$	p	$\bar{d}$	d	$\bar{f}$	f
or	s	p_-	p_+	d_-	d_+	f_-	f_+
or	$s_{1/2}$	$p_{1/2}$	$p_{3/2}$	$d_{3/2}$	$d_{5/2}$	$f_{5/2}$	$f_{7/2}$

Angular wave functions

With explicit expressions for the Clebsch-Gordan coefficients, and omitting the angular variables,

$$\chi_{\kappa,m} = \frac{1}{\sqrt{2\ell+1}} \begin{pmatrix} a\sqrt{\ell+\frac{1}{2}+am}\; Y_\ell^{m-1/2} \\ \sqrt{\ell+\frac{1}{2}-am}\; Y_\ell^{m+1/2} \end{pmatrix}$$

For $a=+1$, i.e. $j=\ell+\frac{1}{2}$, and with $m_\ell=m-\frac{1}{2}$,

$$\chi_{\kappa,m} = \frac{1}{\sqrt{2\ell+1}} \begin{pmatrix} \sqrt{\ell+m_\ell+1}\; Y_\ell^{m_\ell} \\ \sqrt{\ell-m_\ell}\; Y_\ell^{m_\ell+1} \end{pmatrix}$$

and for $a=-1$, i.e. $j=\ell-\frac{1}{2}$,

$$\chi_{\kappa,m} = \frac{1}{\sqrt{2\ell+1}} \begin{pmatrix} -\sqrt{\ell-m_\ell}\; Y_\ell^{m_\ell} \\ \sqrt{\ell+m_\ell+1}\; Y_\ell^{m_\ell+1} \end{pmatrix}$$

Angular wave functions

s	$\begin{pmatrix} s \\ 0 \end{pmatrix}$		
$\sqrt{3}\ \bar{p}$	$\begin{pmatrix} -p_0 \\ \sqrt{2}p_1 \end{pmatrix}$		
$\sqrt{3}\ p$	$\begin{pmatrix} \sqrt{2}p_0 \\ p_1 \end{pmatrix}$	$\begin{pmatrix} p_1 \\ 0 \end{pmatrix}$	
$\sqrt{5}\ \bar{d}$	$\begin{pmatrix} -\sqrt{2}d_0 \\ \sqrt{3}d_1 \end{pmatrix}$	$\begin{pmatrix} -d_1 \\ 2d_2 \end{pmatrix}$	
$\sqrt{5}\ d$	$\begin{pmatrix} \sqrt{3}d_0 \\ \sqrt{2}d_1 \end{pmatrix}$	$\begin{pmatrix} 2d_1 \\ d_2 \end{pmatrix}$	$\begin{pmatrix} d_2 \\ 0 \end{pmatrix}$

Properties of angular wave functions

- Densities depend on j and not on ℓ. Thus s and $\bar{p}$ 2-spinors are *both* spherical.
- Relativistic filled shell is spherical.
- Mixture of m_ℓ and $m_{\ell+1}$ character approaches equality as ℓ increases: loss of spatial directionality.
- Loss of nodal structure: "fat" orbitals. Large and small components will have nodes in different places, and spin-up and spin-down components of each of these will also have different nodal structure.

Polar density plots

Radial Dirac equation

By use of the relation

$$(\sigma.\nabla)\left[\frac{F(r)}{r}\chi_{\kappa m}\right] = \frac{-1}{r}\left[\frac{dF}{dr} + \frac{\kappa F}{r}\right]\chi_{-\kappa m}$$

we obtain the radial Dirac equation in Hartree atomic units $\hbar = e = m = 1$

$$\begin{pmatrix} E - V & c\left[\frac{\mathrm{d}}{\mathrm{d}r} - \frac{\kappa}{r}\right] \\ -c\left[\frac{\mathrm{d}}{\mathrm{d}r} + \frac{\kappa}{r}\right] & 2c^2 + E - V \end{pmatrix} \begin{pmatrix} P(r) \\ Q(r) \end{pmatrix} = 0$$

or as a pair of coupled equations:

$$(E - V)\,P + c\left[\frac{dQ}{dr} - \frac{\kappa Q}{r}\right] = 0$$

$$-c\left[\frac{dP}{dr} + \frac{\kappa P}{r}\right] + (2c^2 + E - V)\,Q = 0$$

Here the prime on the energy has been dropped, and henceforth the shifted energy will be employed unless otherwise specified.

Note: In many physics texts the units used are $\hbar = c = e = 1$. This is more convenient when particles of different masses are being considered. Here however we are interested in electrons, and possibly the use of c as a perturbation parameter, so Hartree atomic units are preferred.

Dirac radial wave functions

For a point nucleus, $V = -Z/r$, the solutions of the Dirac equation are

$$\begin{pmatrix} P(r) \\ \\ Q(r) \end{pmatrix} = \mathcal{N} e^{-\rho/2} \rho^{\gamma} \begin{pmatrix} \mathcal{N}_P (F_1 + F_2) \\ \\ \mathcal{N}_Q (F_1 - F_2) \end{pmatrix}$$

where

$$\mathcal{N}_P = \sqrt{2 + E/c^2}; \quad \mathcal{N}_Q = -\sqrt{-E/c^2};$$

$$\rho = 2\lambda r; \quad \lambda = -c^2 \mathcal{N}_P \mathcal{N}_Q = \sqrt{-E(2c^2 + E)}.$$

We define the radial quantum number n_r by

$$n_r = n - |\kappa|.$$

The apparent angular quantum number γ is

$$\gamma = +\sqrt{\kappa^2 - Z^2/c^2} < |\kappa|,$$

and the apparent principal quantum number N is

$$\begin{aligned} \mathrm{N} &= \sqrt{(n_r + \gamma)^2 + Z^2/c^2} \\ &= \sqrt{n^2 - 2n_r(|\kappa| - \gamma)} \\ &\leq n. \end{aligned}$$

Dirac radial wave functions

F_1 and F_2 are polynomials in ρ, given by

$$F_1 = (\mathrm{N} - \kappa) F(-n_r, 2\gamma + 1, \rho);$$
$$F_2 = -n_r F(1 - n_r, 2\gamma + 1; \rho).$$

$F(a, b; r)$ is a confluent hypergeometric function.

The normalization factor is

$$\mathcal{N} = \frac{\lambda}{\Gamma(2\gamma + 1)} \left[\frac{2\lambda}{\mathrm{N}(\mathrm{N} - \kappa)} . \frac{\Gamma(n_r + \gamma)}{\Gamma(n_r + 1)} \right]^{\frac{1}{2}} .$$

- The normalization factors for P and Q display the "large" and "small" character of these functions.
- $E = -2c^2$ is a limiting value, at which the large component vanishes. Similarly, at $E = 0$ the small component vanishes.
- γ is less than $\ell + 1$: s and $\bar{p}$ functions will have a *singularity* at the origin.

Dirac radial wave functions

The orbital eigenvalue $\epsilon_{n\kappa}$ is given by

$$\begin{aligned}
\epsilon_{n\kappa} &= -c^2 \left[1 - \sqrt{1 - \left(\frac{Z}{c\mathrm{N}} \right)^2} \right] \\
&= -\frac{Z^2}{\mathrm{N}^2} \left[1 + \sqrt{1 - \left(\frac{Z}{c\mathrm{N}} \right)^2} \right]^{-1} \\
&< -\frac{Z^2}{2n^2}
\end{aligned}$$

since $\mathrm{N} \leq n$. The mean radius is given by

$$\begin{aligned}
\langle r \rangle_{n\kappa} &= \frac{1}{2Z} \left[\frac{\gamma + n_r}{\mathrm{N}} (3\mathrm{N}^2 - \kappa^2) - \kappa \right] \\
&< \frac{1}{2Z} \left[3\mathrm{N}^2 - \kappa(\kappa + 1) \right] \\
&< \frac{1}{2Z} \left[3n^2 - \ell(\ell + 1) \right]
\end{aligned}$$

since $\ell(\ell+1) = \kappa(\kappa+1)$. A simple explanation of the decrease in the mean radius is found by considering the old quantum theory. Here, the angular momentum had to be an integer multiple of $\hbar$. Given that $\boldsymbol{\ell} = \mathbf{r} \times \mathbf{p}$ and that p increases with the velocity of the electron, r must decrease with velocity to maintain the quantization condition.

Properties of eigenstates

It is instructive to compare the solutions of the Schrödinger equation with those of the Dirac equation. The radial solution of the Schrödinger equation is

$$P^{NR}(r) = \mathcal{N}^{\mathcal{NR}} e^{-\rho/2} \rho^{\ell} L_{n+\ell}^{2\ell+1}(\rho)$$

where $\rho = 2r\sqrt{-2E}$; the Dirac solution (repeated here) is

$$\begin{pmatrix} P(r) \\ Q(r) \end{pmatrix} = \mathcal{N} e^{-\rho/2} \rho^{\gamma} \begin{pmatrix} \mathcal{N}_P(F_1 + F_2) \\ \mathcal{N}_Q(F_1 - F_2) \end{pmatrix}$$

where $\rho = 2r\sqrt{-E(2c^2 + E)}$. The nature of the spectrum is contained in these radial variables.

- For $E > 0$, ρ is imaginary in both cases. The solutions are oscillatory and represent a free electron.
- For $E < 0$, ρ is real in both cases. The solutions are decaying and represent bound electron states.
- For $E < -2c^2$ in the Dirac solutions, ρ is again imaginary. The solutions are oscillatory - but what do they represent?

Thus, in contrast to the Schrödinger spectrum, the spectrum of the Dirac Hamiltonian has two continua, one of positive energy and one of negative energy. If the external field (nuclear potential) is strong enough that an electron is in a state with $E < -2c^2$, there is enough energy to create an electron-positron pair. Thus they can be considered as states in which spontaneous pair production occurs.

Spectrum of Schrödinger and Dirac Hamiltonians

Large Z behaviour

But is it possible for an electron to have enough energy in a Coulomb field for pair production? Consider the Dirac eigenvalue, rewritten as

$$\epsilon_{n\kappa} = -c^2 \left[1 - \frac{n_r + \gamma}{\mathrm{N}} \right].$$

$\gamma = \sqrt{\kappa^2 - Z^2/c^2}$ becomes zero if $Z = c|\kappa|$ and is imaginary for higher Z. Thus for $Z/|\kappa| > c \simeq 137$,

- The energy, and consequently the radial variable ρ, becomes complex.
- A continuum component has entered the wavefunction, i.e. it has become a *decaying* rather than a *bound* state.

At the point where this happens, the energy is only $-mc^2$, not $-2mc^2$, so the decaying component is *not* due to spontaneous pair production. Rather it is due to the inadequate physics in the pure Coulomb field and the Dirac equation, which is addressed by quantum electrodynamics. The incorporation of a finite nuclear size removes the singularity at the origin and with it the decaying component for $Z/|\kappa| > c$. The eigenvalues then decrease smoothly as a function of Z until they dive into the negative continuum. Attempts have been made to observe this phenomenon by high-energy collisions of heavy ions.

Eigenvalues as a function of Z

Small r behaviour

We now turn to consider the behaviour of the solutions of the Dirac equation at the origin. We expand the radial functions in a power series:

$$P(r) = r^{\gamma_p}(p_0 + p_1 r + p_2 r^2 + \ldots)$$
$$Q(r) = r^{\gamma_q}(q_0 + q_1 r + q_2 r^2 + \ldots)$$

(note that P and Q do not necessarily have the same lowest order term), and also the potential V:

$$-rV(r) = v_0 + v_1 r + v_2 r^2 + \ldots .$$

For a point nucleus, $v_0 = Z,\ v_i = 0,\ i > 0$, while for a finite nucleus, $v_0 = v_2 = 0,\ v_1 \neq 0$. Substituting in the radial Dirac equation gives

$$r^{\gamma_p - 1}[v_0 p_0] + r^{\gamma_p}[(v_1 + E)p_0 - v_0 p_1] + \ldots$$
$$+ r^{\gamma_q - 1}[c(\gamma_q - \kappa)]q_0 + r^{\gamma_q}[c(\gamma_q - \kappa + 1)]q_1 + \ldots = 0$$

$$r^{\gamma_q - 1}[v_0 q_0] + r^{\gamma_q}[(2c^2 + v_1 + E)q_0 - v_0 q_1] + \ldots$$
$$- r^{\gamma_p - 1}[c(\gamma_p + \kappa)]p_0 - r^{\gamma_p}[c(\gamma_p + \kappa + 1)]p_1 - \ldots = 0$$

Equating powers of r, there are 3 cases to consider: $\gamma_p = \gamma_q$, $\gamma_p < \gamma_q$ and $\gamma_p > \gamma_q$.

Small r behaviour

Case 1. $\gamma_p = \gamma_q = \gamma$.

Equating lowest powers of r gives

$$v_0 p_0 + c(\gamma - \kappa) q_0 = 0$$
$$v_0 q_0 - c(\gamma + \kappa) p_0 = 0$$

from which, by eliminating q_0,

$$v_0^2 = c^2(\kappa^2 - \gamma^2)$$

This is the *point* nucleus case. Since $v_0 = Z$, we obtain the definition of γ given earlier. Note that if $v_0 = 0$, we have $\gamma = -\kappa$ *and* $\gamma = \kappa$ at the same time!

Case 2. $\gamma_p < \gamma_q$.

Then $v_0 p_0 = 0$, and thus $v_0 = 0$. We must have $\gamma_q = \gamma_p + 1$, and then

$$(v_1 + E) p_0 + c(\gamma_q - \kappa) q_0 = 0$$
$$c(\gamma_p + \kappa) p_0 = 0$$

from which $\gamma_p = -\kappa$, and thus $\kappa < 0$. This is the *finite* nucleus case for $\kappa < 0$.

Case 3. $\gamma_p > \gamma_q$.

Then $v_0 q_0 = 0$, and again $v_0 = 0$. Here $\gamma_p = \gamma_q + 1$, and

$$c(\gamma_q - \kappa) q_0 = 0$$
$$(2c^2 + v_1 + E) q_0 + c(\gamma_q + \kappa) p_0 = 0$$

from which $\gamma_q = \kappa$, and thus $\kappa > 0$. This is the *finite* nucleus case for $\kappa > 0$.

Properties of solutions

We now consider some of the features of the solutions for point and finite nuclear charge distributions.

Point nucleus case.

- P and Q have the same behaviour at origin.
- Both are series in r (not r^2).
- The solutions for $|\kappa| = 1$ have a singularity at the origin. This could be a serious problem in quantum chemistry where Gaussian basis functions are used to expand the wave function. The cusp which has to be represented in the Schrödinger solutions is now replaced by a singularity, which will inevitably make greater demands on the flexibility of the basis.
- For $\gamma > 1/2$, there is only one normalizable solution, i.e. the regular solution with leading power r^γ. This is called the "limit point" case.
- For $\gamma < 1/2$ both the regular (r^γ) and irregular ($r^{-\gamma}$) solutions are normalizable, and the general solution is a linear combination of the two. This is called the "limit circle" case. The extra condition that the integral over the potential must also be finite eliminates the irregular solution.

Properties of solutions

Finite nucleus case.

- P and Q have the same lowest order term as the corresponding nonrelativistic function at the origin. This means that the large component behaves like the NR function, $P/r \sim r^\ell$, but the small component behaviour depends on κ. For $\kappa < 0$ $Q/r \sim r^{\ell+1}$, for $\kappa > 0$, $Q/r \sim r^{\ell-1}$. The spin-orbit splitting can then be understood in terms of the expectation of the potential over the small component, which is larger for positive κ than for negative κ.
- For at least the first few terms, the solutions are either odd or even functions of r. If the nuclear potential is an even function for all r (e.g. for a Gaussian nuclear charge distribution), P and Q are odd or even for all r.
- At least inside the nucleus, P and Q are essentially *Gaussian* in shape. This means that in a method using a Gaussian basis set a nuclear charge distribution with a finite radius is preferred to a point nucleus: the basis then has the right behaviour at the origin, and the demands on the basis are smaller because of the cutoff in the potential.

Nuclear models

If a nuclear charge distribution with a finite radius is to be used, the question of the functional form of the distribution must be raised. The point nuclear model was simple: now one has to consider some form of the nuclear charge distribution which bears some relation to experimentally determined distributions. Some commonly used models are:

- The uniformly charged sphere,

$$\begin{aligned}\rho_{nuc}(r) &= \rho_0, \quad r < r_{nuc} \\ &= 0, \quad r > r_{nuc}.\end{aligned}$$

  This is the simplest model, and gives rise to a quadratic potential inside the nucleus and a Coulombic potential outside:

$$\begin{aligned}V_{nuc}(r) &= -\frac{Z}{2r_{nuc}}\left[3-\left(\frac{r}{r_{nuc}}\right)^2\right], \quad r < r_{nuc} \\ &= -\frac{Z}{r}, \quad r > r_{nuc}.\end{aligned}$$

Nuclear models

- The Fermi two-parameter distribution,

$$\rho_{nuc}(r) = \frac{\rho_0}{1 + \exp[(r - r_{nuc})/s]},$$

  which introduces a smooth decay of the nuclear distribution rather than a sudden drop to zero. The potential is difficult to represent in closed form. It reduces to the uniform distribution in the limit $s \to 0$. A variant of this distribution, the three-parameter distribution, multiplies it by a quadratic function of r. This model has been used to fit a range of nuclear scattering data.

- The Gaussian distribution,

$$\rho_{nuc}(r) = \rho_0 \exp[-\zeta_{nuc} r^2],$$

  which is a convenient choice for quantum chemical calculations, since the integrals are readily evaluated. For this choice, the nuclear charge distribution falls off too slowly for heavy atoms, but the effect is not large. The potential is simply

$$V_{nuc}(r) = -\frac{Z}{r} \operatorname{erf}(\zeta_{nuc}^{1/2} r).$$

Nuclear models

The principal parameter used in these distributions is the nuclear radius. For many nuclei the rms radius has been determined. These data could be used directly in calculations, but it is probably more useful in quantum chemistry, where the accuracy of the nuclear model is not very important, to use a fitted value. Two common fits to the cube root of the mass number are used:

$$r_{nuc} = 1.2A^{1/3}$$

$$r_{nuc} = 0.836A^{1/3} + 0.57$$

where r_{nuc} is given in fm. This value may be used directly in the uniform and Fermi distributions; for the Gaussian distribution it may be used to determine the exponent by matching the rms radii, so that

$$\eta_{nuc}^{1/2} = 3/2r_{nuc}.$$

In addition, the Fermi distribution has the parameter s which can be taken to be constant, determined from fits to scattering data.

Nuclear models

The difference between the models is not large, as shown by the following data for hydrogen-like mercury. The effect of the finite radius is approximately 2 E_{h}, but the difference between the models is only 0.01 E_{h}.

Nuclear model	Eigenvalue	Difference
Point	−3532.191 849	
		2.017 574
Uniform	−3530.174 275	
		0.007 881
Fermi	−3530.182 156	
		0.011 843
Gaussian	−3530.193 999	

Negative energy states

We now return to the problem of the negative energy continuum. If the negative continuum is empty, why doesn't the H atom decay radiatively into it? The radiative transition moment would connect the large component of the positive energy solution with the small component of the negative energy solution, both of which should be large in magnitude, giving a large transition moment. Dirac postulated that the negative continuum is fully occupied. The implications of this postulate are significant and wide-ranging:

- The excitation of an electron from the negative continuum leaves a positively-charged hole in the vacuum — a positron. The phenomenon of pair creation is thereby explained, and the energy required to create a pair with zero kinetic energy is $2mc^2$. The negative energy states are therefore connected with *positrons*.
- However, the filling of the negative continuum implies that the "vacuum" state is infinitely charged.
- As a consequence, what we measure is the *differences* between a state and the vacuum.
- The hydrogen atom then becomes an infinitely many-body problem! This is clearly not a satisfactory state of affairs.

The Dirac interpretation of the vacuum nevertheless has many features which are in fact correct. For a more satisfactory approach to the interpretation of the vacuum we must turn to quantum field theory and quantum electrodynamics.

Classical field theory

For a continuous mechanical system, the Lagrangian is expressed as an integral over a Lagrangian density, L,

$$\mathcal{L} = \int L(\eta, \dot{\eta}, \nabla\eta, t)\mathrm{d}^3 r$$

where η is the amplitude of the motion, and are functions of $\mathbf{r}$ and t. The equations of motion derived from L are

$$\frac{\partial L}{\partial \eta} - \partial_\mu \frac{\partial L}{\partial(\partial_\mu \eta)} = 0.$$

Note that the 4-vector $\partial L/\partial(\partial_\mu \eta)$ is a contravariant vector. The equations of motion are Lorentz invariant provided L is a scalar density, which is also Lorentz invariant. The canonical momentum density and Hamiltonian density are defined in terms of η as

$$\pi = \frac{\partial L}{\partial \dot{\eta}}; \qquad H = \pi\dot{\eta} - L$$

Thus the η take the place of the generalized coordinates in the mechanics of particles. The η are the variables of a *scalar field.* One can similarly define a vector field in which each field variable has equations of motion and momenta as described, and in which the Hamiltonian has a sum over components of the field variables - thus

$$H = \pi_\mu \dot{\eta}^\mu - L$$

One example of a vector field is the electromagnetic field, with $\eta_\mu = A_\mu$.

Quantization of a field

The quantization conditions for a particle may be expressed in commutator form as

$$[q_i, p_j] = i\hbar\delta_{ij}; \quad [q_i, q_j] = [p_i, p_j] = 0; \quad i, j = x, y, z.$$

For a field, one then requires a pair consisting of a field variable and its conjugate momentum variable. These are then interpreted as operators with the same commutation relations. Thus we have η replacing q and π replacing p, but now these are functions, so the appropriate commutator relations are

$$\begin{aligned}
&[\eta_s(\mathbf{r}, t), \pi_{s'}(\mathbf{r}', t)] = i\hbar\delta_{ss'}\delta^3(\mathbf{r} - \mathbf{r}');\\
&[\eta_s(\mathbf{r}, t), \eta_{s'}(\mathbf{r}', t)] = [\pi_s(\mathbf{r}, t), \pi_{s'}(\mathbf{r}', t)] = 0.
\end{aligned}$$

To determine π we need the Lagrangian density, and therefore we need a representation of the field variable. Since η is the amplitude of the motion, it is natural to take the wave function, ψ, as the field variable. The Lagrangian density for the Dirac equation is taken to be

$$L = i\hbar\psi^\dagger\dot{\psi} - V\psi^\dagger\psi - c\psi^\dagger\boldsymbol{\alpha}\cdot(i\hbar\nabla - q\mathbf{A})\psi$$

The canonical momentum is then

$$\pi = \frac{\partial L}{\partial\dot{\psi}} = i\hbar\psi^\dagger.$$

Note that this is the same for both the Schrödinger and Dirac equations, in which the only term which contains $\dot{\psi}$ is the same. This is not true for the Klein-Gordon equation.

Picture change

To develop further, it is necessary also to transform from the Schrödinger picture to the Heisenberg picture. In the Schrödinger picture, a state vector $|a_S(t)\rangle$ is related to the vector at time $t = 0$ by

$$|a_S(t)\rangle = e^{-i\hat{\mathcal{H}}t/\hbar}|a_S(0)\rangle,$$

and the time derivative of a matrix element by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle b_S(t)|\Omega|a_S(t)\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\langle b_S(0)|e^{i\hat{\mathcal{H}}t/\hbar}\Omega e^{-i\hat{\mathcal{H}}t/\hbar}|a_S(0)\rangle \\ &= \langle b_S(0)|e^{i\hat{\mathcal{H}}t/\hbar}\frac{\mathrm{d}\Omega}{\mathrm{d}t}e^{-i\hat{\mathcal{H}}t/\hbar}|a_S(0)\rangle \\ &+ \frac{1}{i\hbar}\langle b_S(0)|\left[e^{i\hat{\mathcal{H}}t/\hbar}\Omega e^{-i\hat{\mathcal{H}}t/\hbar}, \hat{\mathcal{H}}\right]|a_S(0)\rangle.\end{aligned}$$

The operators in the Schrödinger picture, e.g. $\hat{\mathcal{H}}$, are generally time-independent.

In the Heisenberg picture, it is the state vectors which are taken to be time-independent,

$$|a_H(t)\rangle = e^{i\hat{\mathcal{H}}t/\hbar}|a_S(t)\rangle = |a_S(0)\rangle,$$

and the time-dependence is transferred to the operators,

$$\Omega_H = e^{i\hat{\mathcal{H}}t/\hbar}\Omega_S e^{-i\hat{\mathcal{H}}t/\hbar}.$$

The matrix elements in the Schrödinger and Heisenberg pictures are of course identical.

Picture change

Since the state vectors in the Heisenberg picture do not depend on time, the time derivative can be taken inside the matrix element, so that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle b_H|\Omega_H|a_H\rangle = \langle b_H|\dot{\Omega}_H|a_H\rangle + \frac{1}{i\hbar}\langle b_H|[\Omega_H, \hat{\mathcal{H}}]|a_H\rangle.$$

where

$$\dot{\Omega}_H = e^{i\hat{\mathcal{H}}t/\hbar}\frac{\partial\Omega_S}{\partial t}e^{-i\hat{\mathcal{H}}t/\hbar} \equiv \frac{\partial\Omega_H}{\partial t}.$$

This is valid for a general bra and ket, so that Heisenberg's equations of motion can be written

$$\frac{\mathrm{d}\Omega_H}{\mathrm{d}t} = \frac{\partial\Omega_H}{\partial t} + \frac{1}{i\hbar}[\Omega_H, \hat{\mathcal{H}}].$$

This is the quantized equivalent to the classical mechanics formulation in terms of Poisson brackets,

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + \{F, \mathcal{H}\}.$$

Heisenberg's equations of motion as given above are appropriate for bosons; for fermions, the commutator must be replaced with an anticommutator, and the field variables must obey anticommutation relations:

$$\frac{\mathrm{d}\Omega_H}{\mathrm{d}t} = \frac{\partial\Omega_H}{\partial t} + \frac{1}{i\hbar}[\Omega_H, \hat{\mathcal{H}}]_+.$$

$$[\eta_s(\mathbf{r}, t), \pi_{s'}(\mathbf{r}', t)]_+ = i\hbar\delta_{ss'}\delta^3(\mathbf{r} - \mathbf{r}');$$
$$[\eta_s(\mathbf{r}, t), \eta_{s'}(\mathbf{r}', t)]_+ = [\pi_s(\mathbf{r}, t), \pi_{s'}(\mathbf{r}', t)]_+ = 0.$$

Quantized wave equations

The Hamiltonian is derived from the Hamiltonian density,

$$\hat{H} = \int H \, \mathrm{d}^3 r = \int \psi^\dagger \hat{\mathcal{H}} \psi \, \mathrm{d}^3 r$$

This is now the energy of the field, not the same as the Hamiltonian for a particle. The equations of motion are

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\partial \psi}{\partial t} + \frac{1}{i\hbar}[\psi, \hat{H}]_+.$$

Since ψ is the amplitude in the Heisenberg picture, it does not depend explicitly on time, and therefore the partial derivative is zero. The anticommutator is easily evaluated to give

$$\begin{aligned}
[\psi, \hat{H}]_+ &= \int [\psi, (\psi^\dagger)']_+ \; \hat{\mathcal{H}} \psi' \, \mathrm{d}^3 r' \\
&= \int \delta^3(\mathbf{r} - \mathbf{r}') \; \hat{\mathcal{H}} \psi' \, \mathrm{d}^3 r' \\
&= \hat{\mathcal{H}} \psi
\end{aligned}$$

and thus the time-dependent wave equation is obtained,

$$i\hbar \frac{\mathrm{d}\psi}{\mathrm{d}t} = \hat{\mathcal{H}} \psi.$$

This holds for both the Dirac and Schrödinger Hamiltonians.

Representation of field operators

We now consider a specific representation of the field operators $\psi(\mathbf{r}, t)$. If we have a complete orthonormal set of spatial functions $u_k(\mathbf{r})$ we may expand the field in terms of these functions with time-dependent coefficients,

$$\psi(\mathbf{r}, t) = \sum_k a_k(t) u_k(\mathbf{r}), \qquad \psi^\dagger(\mathbf{r}, t) = \sum_k a_k^\dagger(t) u_k^*(\mathbf{r})$$

However, since we are working in the Heisenberg picture, the field operators are time-independent, so that a_k can be considered to be operators in some space. This space is called *Fock space*. The Hamiltonian may be written

$$\hat{H} = \sum_{kl} a_k^\dagger \int u_k^* \hat{\mathcal{H}} u_l \, \mathrm{d}^3 r \; a_l = \sum_{kl} a_k^\dagger H_{kl} a_l$$

In this representation the equivalence with the matrix formulation of quantum mechanics is obvious.

Note that the field operator represents the *entire* wave field, not just one particular solution. Therefore it is *independent* of the particular representation chosen for the expansion. Transformation between representations may be achieved by rotating the spatial functions and counter-rotating the time-dependent coefficients.

Summary of QFT

We now summarize some of the ideas and implications of quantum field theory.

- The quantization of a field is called *second quantization*, in contrast to the quantization of particle motion, which is called *first quantization*. The particular form of the operators introduced above is only one realization of the second quantized operators, but it is the most common.
- The operators a_k and $a_k^\dagger$ are called *annihilation* and *creation* operators. These obey the anticommutation relations for fermions and the commutation relations for bosons. They are often written for convenience in terms of the indices as k and $k^\dagger$.
- Any first-quantized operator can be represented by the space integral of its density with the field operators. One important operator is the *number operator*, N, which gives the total number of particles in the system.

$$N = \int \psi^\dagger \psi \ \mathrm{d}^3 r = \sum_{kl} a_k^\dagger \int u_k^* u_l \ \mathrm{d}^3 r \ a_l$$
$$= \sum_{kl} a_k^\dagger \delta_{kl} a_l = \sum_k a_k^\dagger a_k$$

- If the set of functions used to expand the field operators is not complete, then the second quantized operators are not an exact representation of the first quantized operators but a projection onto a finite subspace.

N-particle states

The (one-electron) operators derived from QFT consist of a product of a creation operator, a matrix element and an annihilation operator, in that order. For an interaction to exist, there must be a particle in the state which corresponds to the annihilation operator. Thus we can represent a state K containing n particles by a product of creation operators acting on the vacuum state.

$$|K\rangle = \prod_{k}^{n} a_k^\dagger |vac\rangle$$

Here, the vacuum state will be represented $|vac\rangle$ or $|\ \rangle$ to save confusion later with a reference state with index 0, $|0\rangle$. For bosons, a creation operator can occur more than once in the product, but for fermions the appearance of a creation operator more than once gives a zero result due to the anticommutation rules.

Some important results

We quote here some important results of QFT, which will be used later. The rotation of an orthonormal basis can be achieved by the use of the unitary operator

$$\hat{U} = e^{i\hat{\lambda}}$$

where $\hat{\lambda}$ is a Hermitian operator,

$$\hat{\lambda} = \int \psi^\dagger \Lambda \psi \; \mathrm{d}^3 r = \sum_{r\,s} \lambda_{r\,s} a_r{}^\dagger \, a_s \, .$$

The operators are transformed by

$$\begin{aligned} \hat{O} &= \hat{U} \; \hat{O} \; \hat{U}^\dagger \\ &= e^{i\hat{\lambda}} \; \hat{O} \; e^{-i\hat{\lambda}} \\ &= 1 + i[\hat{\lambda}, \hat{O}] + \frac{i^2}{2!}[\hat{\lambda}, [\hat{\lambda}, \hat{O}]] + \ldots \end{aligned}$$

The annihilation operators transform to

$$a'_p = \sum_q U_{p\,q} a_q \, ;$$

the creation operators transform according to the hermitian conjugate. The transformation $\hat{U}$ applied to an N-particle state generates a state with the same occupation numbers with respect to the new basis.

Negative energy states

We now return to the negative energy states of the Dirac equation. If we take the complex conjugate of the Dirac equation for an electron, we have

$$c(\alpha^\mu)^*[-i\hbar\partial_\mu + e\mathrm{A}_\mu]\Psi^* = \beta mc^2\Psi^*$$

Although α^μ is Hermitian, the complex conjugate is not the same as the original because $\alpha_y^* = -\alpha_y$. If we define an operator

$$\hat{\mathcal{C}}' = i\beta\alpha_y$$

we find that

$$\hat{\mathcal{C}}'\beta = -\beta\hat{\mathcal{C}}'; \qquad \hat{\mathcal{C}}'(\alpha^\mu)^* = \alpha^\mu\hat{\mathcal{C}}'$$

and applying this to the conjugated Dirac equation,

$$c\alpha^\mu[i\hbar\partial_\mu - e\mathrm{A}_\mu]\hat{\mathcal{C}}'\Psi^* = \beta mc^2\hat{\mathcal{C}}'\Psi^*.$$

which is the Dirac equation for a positron. Thus, $\hat{\mathcal{C}}'\Psi^*$ is the wave function for a positron, and $\hat{\mathcal{C}} = \hat{\mathcal{C}}'\hat{\mathcal{K}}_0$ is called the *charge conjugation* operator. $\hat{\mathcal{K}}_0$ is the complex conjugation operator.

We may therefore interpret the negative energy solutions as solutions of the Dirac equation for positrons. The appearance of a negative sign for the mass can be reinterpreted by this means as a positive sign of the charge, and this is what is done in quantum electrodynamics. The explicit development of this idea in second quantization follows.

QED field operator

In the representation of the second-quantized Dirac Hamiltonian, the complete set of states must be used to obtain a proper representation of the first-quantized Dirac Hamiltonian. This includes the negative energy states. Thus the field operator sums over both:

$$\Psi = \sum_{p^+} a_{p^+} u_{p^+} + \sum_{p^-} a_{p^-} u_{p^-} .$$

Now, the negative energy solutions are to be reinterpreted as the complex conjugates of positron solutions in a representation that maintains orthogonality to the electron solutions. The complex conjugation converts the annihilation operator into a creation operator. Thus, the field operator becomes

$$\Psi = \sum_{p} \left[a_p \, u_p + \tilde{a}_p{}^\dagger \, \tilde{u}_p \right]$$

where the tildes indicate positronic entities. This operator conserves charge, because it combines the creation of positrons with the annihilation of electrons.

QED Hamiltonian operator

The Hamiltonian operator is

$$\hat{H} = \sum_{pq} \left[H_{pq}\ a_p{}^\dagger a_q + H_{p\tilde{q}}\ a_p{}^\dagger \tilde{a}_q{}^\dagger + H_{\tilde{p}q}\ \tilde{a}_p\, a_q + H_{\tilde{p}\tilde{q}}\ \tilde{a}_p\, \tilde{a}_q{}^\dagger \right]$$

There are several points to be noted about this operator:

- The second term creates an electron-positron pair, and the third term annihilates an electron-positron pair. This means that the Hamiltonian connects states with different particle numbers, i.e. particle number is *not* conserved, though charge is. The existence of these terms embodies the idea of an infinitely many-body problem which arose from the filling of the negative energy states in Dirac's interpretation.
- The order of the operators in the fourth term means that the vacuum expectation value of this operator is *not* zero, but

  $$\langle\ |\hat{H}|\ \rangle = \sum_{pq} H_{\tilde{p}\tilde{q}} \langle\ |\tilde{a}_p\, \tilde{a}_q{}^\dagger|\ \rangle = \sum_{pq} H_{\tilde{p}\tilde{q}} \delta_{pq} = \sum_{p} H_{\tilde{p}\tilde{p}}$$

  Thus the vacuum has an energy which is equal to the sum of the energies of the negative solutions of the Dirac equation, as is expected from Dirac's interpretation. Note that the matrix elements are the same as in the Dirac equation, so the sum is negative and infinite.

QED Hamiltonian operator

To avoid the negative infinite vacuum energy, the vacuum expectation value is subtracted from the Hamiltonian to define a new, QED Hamiltonian

$$\hat{H}_{QED} = \sum_{pq} \big[H_{pq}\ a_p{}^\dagger a_q + H_{p\tilde{q}}\ a_p{}^\dagger \tilde{a}_q{}^\dagger + H_{\tilde{p}q}\ \tilde{a}_p a_q$$
$$+ H_{\tilde{p}\tilde{q}}\ (\tilde{a}_p \tilde{a}_q{}^\dagger - \langle\ |\tilde{a}_p \tilde{a}_q{}^\dagger|\ \rangle) \big]$$

The last term can be rewritten using the anticommutation relations as

$$\tilde{a}_p \tilde{a}_q{}^\dagger - \langle\ |\tilde{a}_p \tilde{a}_q{}^\dagger|\ \rangle = \tilde{a}_p \tilde{a}_q{}^\dagger - \delta_{\tilde{p}\tilde{q}} = -\tilde{a}_q{}^\dagger \tilde{a}_p$$

and thus the Hamiltonian is

$$\hat{H}_{QED} = \sum_{pq} \big[H_{pq}\ a_p{}^\dagger a_q + H_{p\tilde{q}}\ a_p{}^\dagger \tilde{a}_q{}^\dagger + H_{\tilde{p}q}\ \tilde{a}_p a_q - H_{\tilde{p}\tilde{q}}\ \tilde{a}_q{}^\dagger \tilde{a}_p \big]$$

The subtraction of the vacuum expectation value is therefore equivalent to permuting all annihilation operators to the right as if all anticommutators vanished. This is called *normal ordering*, and operators which are normal-ordered are often enclosed in colons, thus

$$\hat{H}_{QED} = :\hat{H}:$$

QED Hamiltonian and charge operators

Note that the vacuum expectation of the Hamiltonian is now zero (by construction), i.e. the vacuum has zero energy. Note also that the positronic term now has a negative sign. The energy of a positron state $|\tilde{p}\rangle = \tilde{a}_p{}^\dagger \,|\, \rangle$ is now

$$\langle \tilde{p}|\hat{H}_{QED}|\tilde{p}\rangle = -H_{\tilde{p}\tilde{p}} > 0.$$

Thus, the negative energy (or mass) electron solutions have now become positive energy positron solutions.

To gain some more insight into the transition to the QED picture, we define the total charge operator

$$\begin{aligned}
\hat{Q} &= -e \int \Psi^\dagger \Psi \; \mathrm{d}^3 r \\
&= -e \sum_{pq} \left[\delta_{pq} \; a_p{}^\dagger a_q \; + \delta_{p\tilde{q}} \; a_p{}^\dagger \tilde{a}_q{}^\dagger \; + \delta_{\tilde{p}q} \; \tilde{a}_p \, a_q \; + \delta_{\tilde{p}\tilde{q}} \; \tilde{a}_p \, \tilde{a}_q{}^\dagger \right] \\
&= -e \sum_{p} \left[a_p{}^\dagger a_p \; + \tilde{a}_p \, \tilde{a}_p{}^\dagger \right]
\end{aligned}$$

The vacuum expectation of this operator is infinite because of the last term: the negative energy states are all filled. However, the normal-ordered operator,

$$:\hat{Q}:= -e \sum_{p} \left[a_p{}^\dagger a_p \; - \tilde{a}_p{}^\dagger \, \tilde{a}_p \right]$$

has a zero vacuum expectation, and correctly gives the charge of a state with a given number of electrons or positrons.

Vacuum definition

The subtraction of the vacuum expectation value from the Hamiltonian makes the operator *vacuum dependent*, i. e. the definition of the vacuum depends on the definition of the basis in the expansion of the field operator. The argument may be developed as follows.

- If the Hamiltonian is diagonal, the terms which create and annihilate pairs vanish, and we are left with an operator which conserves particle number. It is in this basis that the interpretation of particles is made.
- A transformation of the basis may be made with the exponential operator $\hat{u} = \exp(i\hat{\lambda})$. The normal-ordered form of $\hat{\lambda}$ is

  $$:\hat{\lambda}:= \sum_{pq} \left[\lambda_{pq}\ a_p{}^\dagger a_q + \lambda_{p\tilde{q}}\ a_p{}^\dagger \tilde{a}_q{}^\dagger + \lambda_{\tilde{p}q}\ \tilde{a}_p a_q - \lambda_{\tilde{q}\tilde{p}}\ \tilde{a}_p{}^\dagger \tilde{a}_q\right]$$

  which contains pair creation and destruction terms. Exponentiating this operator will give an operator containing terms which create one, two, ... pairs.
- The vacuum state in the new basis is derived by operating on the vacuum in the old basis with this exponential operator. Therefore the new vacuum contains a whole series of pair states from the old basis. The structure of the vacuum has changed.

Vacuum definition

- This may be seen by evaluating the number of "old" electrons and positrons in the new vacuum, which gives a non-zero result.

$$\langle vac'|\hat{N}^e|vac'\rangle = \langle vac|\hat{U}^\dagger \sum_p a_p{}^\dagger a_p \hat{U}|vac\rangle$$

  The number of electrons is however equal to the number of positrons, so that charge is conserved.

- The energy of a particle in the new vacuum becomes a weighted sum of energies of states with one particle, a particle and one pair, etc., in the old vacuum.
- With respect to the old vacuum, the new vacuum is called a *dressed* vacuum and the states in it are called *dressed* states. The new particles are called *quasiparticles* because they are a composite of an "old" particle and a series of "old" pairs – they are dressed with a series of pairs rather than being bare particles.
- The new vacuum is also called a *polarized vacuum* because with respect to the old vacuum there is a charge polarization expressed by the presence of undressed pairs. The concept of vacuum polarization will be dealt with later.
- Finally, another way of looking at the change in the vacuum is that because the transformation mixes creation and annihilation operators, a normal-ordered operator in the new basis will definitely not be normal-ordered in the old basis.

Summary of QED concepts

Adoption of the QED picture has the following benefits:

- The negative energy states are reinterpreted as positrons.
- Positrons appear as positive energy particles with positive charge.
- The infinite vacuum energy is removed from the Dirac reinterpretation.
- There is no danger of radiative decay of electrons.

While doing this, it introduces the following features:

- Operators must be used in normal-ordered form.
- Charge is conserved but particle number is not necessarily conserved.
- Particles are therefore defined in a basis in which the Hamiltonian is diagonal. Such a basis is one in which particle number is conserved.
- The vacuum is defined relative to a given representation of the one-particle states.
- No prescription is given for choosing a representation.

The electron-electron interaction

From the discussion of classical relativity, it was clear that the Coulomb interaction between charged particles was not Lorentz invariant. We therefore wish to find an interaction which is invariant.

As foreshadowed in our consideration of classical relativity, the interaction of charged particles in QED is mediated by the emission and absorption of photons. The derivation of the Hamiltonian for the interaction is obtained from perturbation theory. This is usually expressed in terms of Feynman diagrams, which are a pictorial way of representing the interaction as a function of time.

Feynman diagrams are written with time on the vertical axis. Straight lines represent electrons. The direction of propagation is given by an arrow: an electron which is propagating backwards in time is interpreted as a positron. Squiggly lines represent photons. The number of particles at any given time can be found by drawing a horizontal line across the diagram. In some cases it is useful to distinguish between bare particles (i.e. free particles) and dressed particles. The dressed particles appear with a double line.

Feynman diagrams

Feynman diagrams

There are three possible diagrams for the exchange of one photon. The first is the "normal" interaction, where a particle interacts with another particle. The second incorporates the creation of an electron-positron pair (the closed loop). This represents the interaction of the electron with the vacuum, and is known as *vacuum polarization.* In the third diagram, the electron is interacting with itself. This term is called the *self energy* term. The self energy and vacuum polarization terms together make up what is called the Lamb shift.

The diagrams for two-photon exchange are numerous and much more complicated. A few of these are also given. The first is a "ladder" diagram: two electrons exchange a photon at one time, and then exchange another a bit later. These ladder diagrams can be summed to infinite order by incorporating the one-photon interaction into a variational procedure. The second is a diagram in which a pair is created and annihilated. One electron emits a photon which creates a pair. The positron from this pair annihilates the second electron, creating a photon which is reabsorbed by the second electron. In the third diagram, two electrons simultaneously emit a photon each, and later each absorbs the photon from the other electron.

Time-dependent perturbation theory

We now turn to the mathematical form which these embody, and to do this we use time-dependent perturbation theory in the interaction picture, where

$$i\dot{\psi} = V(t)\psi$$

The potential for the interaction can be given in Lorentz invariant form as

$$V = -q \int \mathrm{j}^{\mu} \mathrm{A}_{\mu} \mathrm{d}^3 r$$

We must deal with the time-dependence, which is achieved in the following manner. Taking a small time increment,

$$\psi(t + \delta t) = e^{-i\delta t \ V(t)} \psi(t)$$

For any time interval, we may take the product of the operators and allow the interval to go to zero:

$$\psi(t_2) = \prod_{t_i} e^{-i\delta t_i \ V(t_i)} \psi(t_1) \rightarrow \exp\left(-i \int_{t_1}^{t_2} V(t) \ \mathrm{d}t\right) \psi(t_1).$$

Time-dependent perturbation theory

However, this transition requires that the time intervals and the operators commute, which is not necessarily the case. Therefore, we introduce a *time-ordering operator* $\hat{T}$ which ensures the correct chronological sequence of perturbations. Writing the operator which carries ψ from t_1 to t_2 as $\hat{S}$, we have

$$\begin{aligned}\hat{S} &= \hat{T}\exp\left(-i\int V(t)\,\mathrm{d}t\right)\\ &= \sum_k \frac{(-i)^k}{k!}\int \mathrm{d}t_1\int \mathrm{d}t_2\ldots\int \mathrm{d}t_k\hat{T}\{V(t_1)V(t_2)\ldots V(t_k)\}\end{aligned}$$

Substituting the form of the interaction we get

$$\hat{S} = \hat{T}\exp\left(iq\int \mathrm{j}^\mu \mathrm{A}_\mu \mathrm{d}^4x\right)$$

The integrand is clearly Lorentz invariant - but what of $\hat{T}$?

If the world points x_2 and x_1 are separated by a time-like interval $(\mathrm{x}_2 - \mathrm{x}_1)^2 > 0$ then $t_2 > t_1$ in any frame of reference. However if the interval is space-like, there can be no causal connection between them, and therefore the operators referring to them must commute, because non-commutativity implies that the quantities cannot be measured simultaneously and therefore the events must be connected. Thus in the space-like region the chronological order can be restored if it gets out of order in a Lorentz transformation.

Electron-electron interaction

Now, the lowest term in the sum involves the interaction of only one current with the electromagnetic field, which would correspond to photon emission. Therefore the second term in the sum corresponds to the lowest order electron-electron interaction,

$$\hat{S}^{(2)} = -\frac{qq'}{2}\int\int \mathrm{d}^4\mathrm{x}\ \mathrm{d}^4\mathrm{x}'\hat{T}\big[\mathrm{j}^\mu(\mathrm{x})\mathrm{A}_\mu(\mathrm{x})\mathrm{j}^\nu(\mathrm{x}')\mathrm{A}_\nu(\mathrm{x}')\big]$$

This corresponds to the lowest order Feynman diagram for photon exchange. The photon and electron operators commute, so we may write

$$\hat{S}^{(2)} = -\frac{qq'}{2}\int\int \mathrm{d}^4\mathrm{x}\ \mathrm{d}^4\mathrm{x}'\hat{T}\big[\mathrm{j}^\mu(\mathrm{x})\mathrm{j}^\nu(\mathrm{x}')\big]\hat{T}\big[\mathrm{A}_\mu(\mathrm{x})\mathrm{A}_\nu(\mathrm{x}')\big]$$

Now, to calculate the Fock space matrix element of this operator we consider the initial and final states of the interaction. The inital state has two electrons, as has the final state, and neither has any photons. Therefore the matrix element for the photons is

$$D_{\mu\nu}(\mathrm{x}-\mathrm{x}') = i\langle\ |\hat{T}\big[\mathrm{A}_\mu(\mathrm{x})\mathrm{A}_\nu(\mathrm{x}')\big]|\ \rangle,$$

which is called the *photon propagator*.

Electron-electron interaction

Similarly one may expand the current density in terms of the wave operators. A factor of 2 arises from the expansion, and the exchange due to the anticommutation of the creation and annihilation operators also appears, so that we get

$$\langle 34|\hat{T}\big[\mathrm{j}^{\mu}(\mathrm{x})\mathrm{j}^{\nu}(\mathrm{x}')\big]|12\rangle = c^2\big[(\psi_4{}^{\dagger}\alpha^{\mu}\psi_2)(\psi_3'{}^{\dagger}\alpha^{\nu}\psi_1') \\ -(\psi_4{}^{\dagger}\alpha^{\mu}\psi_1)(\psi_3'{}^{\dagger}\alpha^{\nu}\psi_2')\big]$$

where the numbers label the incoming and outgoing particles.

Thus we have the interaction of two current densities via a photon propagator, which must be integrated over 4-space to give the interaction matrix element and hence the potential for the interaction. The development of this is long and involves the Green's functions for the photon and the electron in momentum space. The final result is an effective interaction matrix element which can be given as

$$S^{(2)}_{pr,sq} = \int \mathrm{d}^3 r_1 \int \mathrm{d}^3 r_2 \; u_p^{\dagger}(\mathbf{r}_1) u_r^{\dagger}(\mathbf{r}_2) \; V(\omega, r_{12}) u_s(\mathbf{r}_2) u_q(\mathbf{r}_1).$$

where ω is the frequency of the exchanged photon.

Electron-electron interaction

The potential for the interaction is derived in the Feynman (or Lorentz) gauge. It is given by

$$V^F(\omega, r_{12}) = \frac{e^2}{4\pi\epsilon_0} \frac{1 - \boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{r_{12}} e^{i|\omega| r_{12}/c}$$

Although the results should be independent of the gauge, the Coulomb gauge is usually preferred in calculations. Thus the Coulomb gauge form of the interaction is

$$V^C(\omega, r_{12}) = V^F(\omega, r_{12}) + \frac{e^2}{4\pi\epsilon_0} (\boldsymbol{\alpha}_1 \cdot \nabla_1)(\boldsymbol{\alpha}_2 \cdot \nabla_2) f(|\omega|/c, r_{12})$$

where the function f is

$$f(\lambda, r) = \frac{1 + i\lambda r - e^{i\lambda r}}{\lambda^2 r}$$

The additional term in the Coulomb interaction is properly called the *gauge* term because it arises from the gauge transformation of the 4-potential.

The interaction has a real and an imaginary part. The imaginary part gives the energy levels a width, and is often neglected, leaving the real interaction operator.

Electron-electron interaction

The frequency of the exchanged photon in this interaction is given by conservation of energy in terms of the one-particle eigenvalues as

$$\omega \equiv \omega_{rs} = \epsilon_s - \epsilon_r \,,$$

This is the interaction of the current of electron 1 with the potential generated by electron 2. It is a retarded potential, and is not symmetric to the interchange of electrons 1 and 2. One could equally derive the retarded potential for the interaction of the current of electron 2 interacting with the potential generated by electron 1, which has the frequency

$$\omega \equiv \omega_{qp} = \epsilon_p - \epsilon_q$$

A procedure which is at least correct to $(1/c^2)$ is to take the real part of the average potential

$$\bar{V}(\omega, r_{12}) = \frac{1}{2}\mathrm{Re}[V(\omega_{rs}, r_{12}) + V(\omega_{qp}, r_{12})]$$

The Breit and Gaunt interactions

The Lorentz invariant electron-electron interaction operator just derived contains differences of orbital energies. This means that each matrix element will have different photon frequencies, and the calculation of integrals is then an enormous task. The oscillatory term in the interaction is only large if ωr_{12} is large. For the interaction of electrons, and in particular for chemical energies, the frequencies are small, and the contributions from large r_{12} will be damped by the $1/r_{12}$ operator, becoming negligible. Thus, to obtain the effect in lowest order, it is reasonable to let $\omega/c \rightarrow 0$. The result is

$$V^F(0, r_{12}) = \frac{1 - \boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{4\pi\epsilon_0 r_{12}}$$

$$V^C(0, r_{12}) = \frac{1 - \boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2}{4\pi\epsilon_0 r_{12}} + \frac{(\boldsymbol{\alpha}_1 \times \mathbf{r}_{12}) \cdot (\boldsymbol{\alpha}_2 \times \mathbf{r}_{12})}{8\pi\epsilon_0 r_{12}^3}$$

The last term is the gauge term. The addition to the Coulomb interaction in the Feynman gauge is the Gaunt interaction, and in the Coulomb gauge it is the Breit interaction.

Note that if one is interested in pair creation effects, such as in the second of the two-photon diagrams, the frequency will always be of order $2c^2$, and therefore the use of the full frequency dependence is *mandatory.*

The Breit interaction

Rewriting the Coulomb gauge interaction as

$$V^C(0, r_{12}) = \frac{1}{4\pi\epsilon_0 r_{12}} - \frac{1}{8\pi\epsilon_0 r_{12}} \left[\boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}_2 + \frac{(\boldsymbol{\alpha}_1 \cdot \mathbf{r}_{12})(\boldsymbol{\alpha}_2 \cdot \mathbf{r}_{12})}{r_{12}^2} \right],$$

a clear connection with the retarded potential from classical relativity can be seen, with $\boldsymbol{\alpha} = \mathbf{u}/c$.

In some places, the last term in square brackets is referred to as the retardation correction, but this is not correct, because the whole term in square brackets arises from a retarded interaction. Moreover, the retardation is expressed here by the finite photon frequency, which does not contribute to this interaction. It is better to describe the Coulomb gauge interaction as a sum of the Coulomb interaction, a current-current interaction and an interaction derived from the interaction of two angular momenta.

One further point about the Feynman and Coulomb gauge interactions. It has been shown that the Coulomb-Breit interaction is correct to $\mathcal{O}(c^{-2})$. The Coulomb-Gaunt interaction is not correct to this order, and does not become correct by summing ladder diagrams. The missing term comes from the crossed two-photon exchange diagram. Thus, in a situation where one is interested in an accurate calculation of low-order effects, the Coulomb gauge interaction is preferable.

The Lamb shift

We have dealt with the first of the one-photon diagrams. The other two are the vacuum polarization and self-energy terms. Since the classical self-energy of an electron is infinite, and the vacuum is infinitely charged in the Dirac picture, it is not surprising that these terms are difficult to handle. The operators derived for the self energy and the vacuum polarization have divergences in the photon frequency, either at large k or at small k. The divergences are removed by redefining the mass and charge of the electron, a process called *renormalization*. Since the mass and charge measured experimentally incorporate these terms, but in the equations derived they do not, there is an extra term to be included in the Hamiltonian density which arises from the difference between the measured and the bare charge. In essence, this involves subtracting a term which arises from the bare vacuum. Put another way, one is measuring not the absolute values of these terms, but the difference in them imposed by the presence of a field. This in fact solves a problem in QED, in which the vacuum is redefined in the presence of a field, so that as the field changes, the vacuum changes too. There is then a problem of comparison of the energies of states in different external fields (for example, different molecular geometries). This is solved by taking the bare vacuum as a reference. The differences in energy are very small, and for molecular calculations can be neglected. It should be noted that QED effects are of lowest order α^3, whereas the Breit interaction contains terms of order α^0 and α^2 (α is the fine structure constant).

Variational theory

No satisfactory means of deriving a variational theory based on QED has yet been achieved. However, it is always possible to construct a one-particle basis by variational methods, since these can be represented as the summation of certain perturbation series to infinite order. Thus one may proceed with safety to construct a mean-field theory which can be used as a basis for the perturbation theory of QED. In particular, the inclusion of the Breit interaction in the mean-field calculations ensures that the terms of $\mathcal{O}(\alpha^2)$ are included to infinite order in QED.

Conjugated Dirac equation

Earlier, when we considered charge conjugation, we took the complex conjugate of the Dirac equation, and interpreted the negative energy solutions as positrons. But what about the positive energy solutions?

If we take the conjugated Dirac equation,

$$c(\alpha^\mu)^*[-i\hbar\partial_\mu + e\mathrm{A}^*_\mu]\Psi^* = \beta mc^2\Psi^*$$

and apply to this the operator

$$\hat{\mathcal{K}}' = -i\sigma_y^{(4)},$$

where $\sigma_y^{(4)}$ is the 4×4 matrix

$$\sigma_y^{(4)} = \begin{pmatrix} \sigma_y & 0_2 \\ 0_2 & \sigma_y \end{pmatrix},$$

and for which

$$\hat{\mathcal{K}}'\boldsymbol{\alpha}^* = -\boldsymbol{\alpha}\hat{\mathcal{K}}'; \qquad \hat{\mathcal{K}}'\beta = \beta\hat{\mathcal{K}}'$$

the result is

$$c\alpha_\mu[-i\hbar\partial_\mu + e\mathrm{A}^*_\mu]\hat{\mathcal{K}}'\Psi^* = \beta mc^2\hat{\mathcal{K}}'\Psi^*.$$

Time-reversal

If we define a new time coordinate which is the negative of the original, $\bar{t} = -t$, we find that the momentum $\mathbf{p}$ stays the same, but p_0 changes sign. If the vector potential changes sign, $\mathbf{A}(-t) = -\mathbf{A}^*(t)$, and the scalar potential A_0 stays the same, the above equation becomes

$$c\alpha_\mu[i\hbar\bar{\partial}^\mu + e\bar{\mathrm{A}}^\mu]\hat{\mathcal{K}}\Psi = \beta mc^2\hat{\mathcal{K}}\Psi$$

where $\hat{\mathcal{K}} = \hat{\mathcal{K}}'\hat{\mathcal{K}}_0$ is called the *time-reversal* operator, and the bars indicate that the time-reversed form of the operator is to be taken. Thus the function $\bar{\Psi} = \hat{\mathcal{K}}\Psi$ is a solution of the time-reversed Dirac equation.

Now, in the absence of a vector potential, i.e. $\mathbf{A} = 0$, the time-independent Dirac operator commutes with $\hat{\mathcal{K}}$. Thus $\bar{\Psi}$ is a solution of the time-independent Dirac equation with the same energy as Ψ. The one-electron solutions of the Dirac equation are therefore *doubly degenerate*. This is Kramers' theorem, and the pairs of such states are called *Kramers pairs*.

For future reference, it should be noted that the double-degeneracy at the one-electron level depends on the forward- and backward-time solutions being determined in the same potential.

Time-reversal operator

The presence of the complex conjugation operator in $\hat{\mathcal{K}}$ means that it is an antilinear operator, i.e.

$$\hat{\mathcal{K}}(a\Psi) = a^* \hat{\mathcal{K}}\Psi,$$

and even though it commutes with the Dirac Hamiltonian, it does not give rise to any observable and thus has no eigenvalues.

If we apply $\hat{\mathcal{K}}$ a second time to the Dirac equation, we would expect to return to the original equation — for, after all, a double reversal of time should give the original equation? However, the square of $\hat{\mathcal{K}}$ is

$$\hat{\mathcal{K}}^2 = -i\sigma_y^{(4)}\hat{\mathcal{K}}_0(-i\sigma_y^{(4)})\hat{\mathcal{K}}_0 = (-i\sigma_y^{(4)})^2\hat{\mathcal{K}}_0^2 = -1_4$$

Thus, although the original equation is indeed obtained, it is with a sign change of the wave function. This will be pursued further in the section on double groups.

Time-reversal and spin

At this point, it is of value to examine the effect of $\hat{\mathcal{K}}$ on a spin function. We may write $\hat{\mathcal{K}} = -i\sigma_y \otimes 1_2 \hat{\mathcal{K}}_0$. The spin functions in vector form were written

$$\alpha \equiv \eta(1/2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \qquad \beta \equiv \eta(-1/2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Obviously, the only part of the operator that matters for the purpose of the spin functions is

$$\hat{\mathcal{K}}_S = -i\sigma_y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

from which we deduce that

$$\hat{\mathcal{K}}_S \ \alpha = \beta, \qquad \hat{\mathcal{K}}_S \ \beta = -\alpha$$

Thus, time-reversal corresponds to a spin-flip, and the construction of many-particle states is made from combinations of both forward- and backward-time wave functions.

Field operators

As for the case of charge conjugation, we may construct a new field operator from the solutions of the Dirac equation for both positive and negative time,

$$\Psi = \sum_p \left[a_p \, u_p(\mathbf{r}) + a_{\bar{p}} \, u_{\bar{p}}(\mathbf{r}) \right]$$

where the operator is now in the un-reinterpreted Dirac picture. In the QED picture, we would need to reinterpret the positron solutions.

The field operator is invariant under time reversal, because we have

$$\hat{\mathcal{K}} a_p = a_{\bar{p}} \,, \qquad \hat{\mathcal{K}} a_{\bar{p}} = -a_p$$

and also

$$\hat{\mathcal{K}} u_p = u_{\bar{p}}, \qquad \hat{\mathcal{K}} u_{\bar{p}} = -u_p.$$

Second-quantized *Kramers-restricted* operators may now be determined with this field operator in the usual manner,

$$\hat{F} = \int \Psi^\dagger \hat{\mathcal{F}} \Psi \; \mathrm{d}^3 r.$$

Operators with time reversal

Before considering the effect of time reversal on the second-quantized operators, we consider the behaviour of some first quantized operators under time reversal.

- The scalar potential for a Coulomb field, A_0, is symmetric.
- The Dirac velocity operator $\boldsymbol{\alpha}$ is antisymmetric. This means that $\hat{\mathcal{K}}$ converts the current operator from covariant to contravariant form and vice versa.
- The momentum operator is antisymmetric, and hence the combination $c\boldsymbol{\alpha} \cdot \mathbf{p}$ is symmetric, as it should be.
- The β matrix is symmetric.
- For a radiation field in the Coulomb gauge, the vector potential is real and hence symmetric under time-reversal. The interaction of a current with the vector radiation field is therefore antisymmetric.
- For a product of two $\boldsymbol{\alpha}$ matrices, time reversal of both gives a symmetric result, but time reversal of only one of them gives an antisymmetric result. Thus the Gaunt interaction is symmetric under time reversal as a whole, but the interaction with a nuclear spin, for example, gives an antisymmetric result unless time reversal of the nuclear spin is also included.

Operators with time reversal

The second quantized form of a single-particle operator $\hat{\mathcal{F}}$ in this basis of Kramers pairs, is

$$\hat{F} = \sum_{pq} \left[f_{pq}\ a_p{}^\dagger\, a_q + f_{p\bar{q}}\ a_p{}^\dagger\, a_{\bar{q}} + f_{\bar{p}q}\ a_{\bar{p}}{}^\dagger\, a_q + f_{\bar{p}\bar{q}}\ a_{\bar{p}}{}^\dagger\, a_{\bar{q}} \right]$$

The first-quantized operator $\hat{\mathcal{F}}$ may by symmetric or anti-symmetric under time reversal, and in addition may be hermitan or anti-hermitian, i.e.

$$\hat{\mathcal{F}}^\dagger = h\hat{\mathcal{F}}, \qquad \hat{\mathcal{K}}\hat{\mathcal{F}}\hat{\mathcal{K}}^{-1} = t\hat{\mathcal{F}}, \qquad h, t = \pm 1$$

Under time-reversal the matrix elements of $\hat{\mathcal{F}}$ have the following behaviour

$$\begin{aligned} \hat{\mathcal{K}} f_{pq} &= \langle \hat{\mathcal{K}} p | \hat{\mathcal{K}}\hat{\mathcal{F}}\hat{\mathcal{K}}^{-1} | \hat{\mathcal{K}} q \rangle = t f_{\bar{p}\bar{q}} = f^*_{pq} = h f_{qp} \\ \hat{\mathcal{K}} f_{p\bar{q}} &= -t f_{\bar{p}q} = f^*_{p\bar{q}} = h f_{\bar{q}p}. \end{aligned}$$

We may therefore collect the matrix elements in $\hat{F}$, with reindexing, to obtain

$$\begin{aligned} \hat{F}^{t,h} = \sum_{pq} \Big[& f_{pq}\ (a_p{}^\dagger\, a_q + th\ a_{\bar{q}}{}^\dagger\, a_{\bar{p}}) + \frac{1}{2} f_{p\bar{q}}\ (a_p{}^\dagger\, a_{\bar{q}} - th\ a_q{}^\dagger\, a_{\bar{p}}) \\ & + \frac{1}{2} f_{\bar{p}q}\ (a_{\bar{p}}{}^\dagger a_q - th\ a_{\bar{q}}{}^\dagger a_p) \Big] \end{aligned}$$

Kramers replacement operators

The operators in parentheses define some operators that are termed are termed the *Kramers single replacement operators*

$$
\begin{aligned}
X^s_{pq} &= a_p{}^\dagger\, a_q \;+ s\; a_{\bar{q}}{}^\dagger\, a_{\bar{p}} \\
X^s_{p\bar{q}} &= a_p{}^\dagger\, a_{\bar{q}} \;- s\; a_q{}^\dagger\, a_{\bar{p}} \\
X^s_{\bar{p}q} &= a_{\bar{p}}{}^\dagger a_q \;- s\; a_{\bar{q}}{}^\dagger a_p
\end{aligned}
$$

so that the one-electron operator can be written

$$
\hat{F}^{t,h} = \sum_{pq} \left[\; f_{pq}\; X^{th}_{pq} + \frac{1}{2} f_{p\bar{q}}\; X^{th}_{p\bar{q}} + \frac{1}{2} f_{\bar{p}q}\; X^{th}_{\bar{p}q} \;\right]
$$

These are *not* the same as the one-particle excitation operators,

$$
\begin{aligned}
E^s_{pq} &= i^{(1-s)/2}(a_p{}^\dagger\, a_q \;+ s\; a_{\bar{p}}{}^\dagger a_{\bar{q}}) \\
E^s_{p\bar{q}} &= i^{(1-s)/2}(a_p{}^\dagger\, a_{\bar{q}} \;- s\; a_{\bar{p}}{}^\dagger a_q).
\end{aligned}
$$

In terms of these operators, the one-electron operator is

$$
\hat{F}^{t,h} = \sum_{pq} \left[\; \mathrm{Re} f_{pq}\; E^t_{pq} + \mathrm{Im} f_{pq}\; E^{-t}_{pq} + \mathrm{Re} f_{p\bar{q}}\; E^t_{p\bar{q}} + \mathrm{Im} f_{p\bar{q}}\; E^{-t}_{p\bar{q}} \;\right]
$$

The symmetry with respect to Hermitian conjugation is contained entirely in the matrix elements in this representation.

Note also that the E operators are defined with respect to spinors, and not spin-orbitals, and are therefore *not* the same as the nonrelativistic one-particle excitation operators.

Comparison of X and E

The X operators have the following features:

- They give the most compact representation of an operator.
- There is no division of matrix elements into parts.
- The symmetry under time-reversal and hermitian conjugation is transparently displayed in the sign index of X.
- Therefore only one kind of X operator needs to be considered for a given property of the first quantized operator.
- The behaviour under time reversal is

$$\hat{\mathcal{K}} X^s_{pq} = s\ X^s_{qp}; \qquad \hat{\mathcal{K}} X^s_{p\bar{q}} = s\ X^s_{\bar{p}q}$$

  i.e. they transform into their conjugates.

The E operators have the following properties

- They give a representation of an operator in terms of real matrices.
- They are genuine excitation operators which can be used in the same way as in nonrelativistic theory.
- They are symmetric under time reversal:

$$\hat{\mathcal{K}} E^s_{pq} = E^s_{pq}; \qquad \hat{\mathcal{K}} E^s_{p\bar{q}} = E^s_{p\bar{q}}$$

The differences and uses of these operators will be pursued later.

Replacement operators

The single replacement operators may be defined in an elegant and compact manner through two auxiliary operators, the *single bar replacement operator*, $\hat{K}_p$, which is the time-reversal operator for a spinor with index p, and the *Kramers permutation operator*, $\hat{T}_{pq}$ whose action is given by

$$\hat{T}_{pq}\ a_p{}^\dagger\, a_q\ = a_{\bar{q}}{}^\dagger a_{\bar{p}}\,, \qquad \hat{T}_{pq}\ a_{\bar{p}}{}^\dagger a_q\ = -a_{\bar{q}}{}^\dagger a_p$$

and which may be considered as the application of the time-reversal operator followed by Hermitian conjugation. These two operators commute. The Kramers single replacement operators X^s_{pq} are then defined by

$$X^s_{pq} = \left(1 + s\hat{T}_{pq}\right) a_p{}^\dagger\, a_q$$

and the operation of $\hat{K}_p$ and $\hat{K}_q$ on this expression.

The commutator of two single replacement operators is given by application of single bar replacement operators to

$$\left[X^{s_1}_{pq}\,,\, X^{s_2}_{rs}\right] = (1+s_2\hat{T}_{pq})(1-s_1 s_2\hat{T}_{ps}\hat{T}_{rq})(\delta_{rq}a_p{}^\dagger\, a_s - \delta_{ps}a_r{}^\dagger\, a_q\,),$$

The Kramers permutation operator must be taken to apply to the Kronecker delta as well as the operators. Thus, for $s_1 = s_2 = +1$ we have

$$\left[X^+_{pq}\,,\, X^+_{rs}\right] = \delta_{rq}X^-_{ps} - \delta_{ps}X^-_{rq}.$$

Replacement operators

Kramers double replacement operators may be defined in analogous manner:

$$x^{s_1 s_2}_{pq,rs} = \left(1 + s_1 \hat{T}_{pq}\right) \left(1 + s_2 \hat{T}_{rs}\right) a_p{}^\dagger \, a_r{}^\dagger \, a_s \, a_q \; = x^{s_2 s_1}_{rs,pq},$$

and the x operators with bars derived from the application of the single bar replacement operators. The double replacement operators can be expressed in terms of the single replacement operators, but not in the same simple manner as in nonrelativistic theory:

$$\begin{aligned} x^{s_1 s_2}_{pq,rs} = X^{s_1}_{pq} X^{s_2}_{rs} &- \delta_{rq} \; a_p{}^\dagger \, a_s \; - s_1 \; \delta_{r\bar{p}} \; a_{\bar{q}}{}^\dagger a_s \\ &- s_2 \; \delta_{\bar{s}q} \; a_p{}^\dagger \, a_{\bar{r}} \; - s_1 s_2 \; \delta_{\bar{p}\bar{s}} \; a_{\bar{q}}{}^\dagger a_{\bar{r}} \, . \end{aligned}$$

The Kronecker delta terms with an odd number of bars have been retained both to show the structure and to enable the expressions for the other x operators to be derived using the single bar replacement operator.

The commutator of a single and a double replacement operator is given by

$$\begin{aligned} \left[X^{s_1}_{tu} \, , \, x^{s_2 s_3}_{pq,rs}\right] = \; & (1 + s_2 T_{pq})(1 + s_3 T_{rs}) \times \\ & \left[(1 + s_1 s_2 T_{pu} T_{tq})(\delta_{pu} a_t{}^\dagger \, a_r{}^\dagger \, a_s \, a_q \; - \delta_{tq} a_p{}^\dagger \, a_r{}^\dagger \, a_s \, a_u \,) \right. \\ & \left. + (1 + s_1 s_3 T_{ru} T_{ts})(\delta_{ru} a_t{}^\dagger \, a_p{}^\dagger \, a_q \, a_s \; - \delta_{ts} a_r{}^\dagger \, a_p{}^\dagger \, a_q \, a_u \,)\right] \end{aligned}$$

Matrix elements

We have already examined the matrix elements of a one-particle operator. The integrals have the following relations

$$f_{\bar{p}\bar{q}} = th\ f_{qp} = t\ f^*_{pq}; \qquad f_{\bar{p}q} = -th\ f_{\bar{q}p} = -t\ f^*_{p\bar{q}}.$$

Thus, the application of time reversal reduces the number of unique matrix elements by a factor of 2.

The Kramers permutation operator is effectively a unit operator when applied to a charge density, for

$$\begin{aligned}\hat{T}_{pq}\ u_p{}^\dagger u_q &= (u_{\bar{p}}{}^\dagger u_{\bar{q}})^* = u_{\bar{q}}{}^\dagger u_{\bar{p}} \\ &= \hat{\mathcal{K}}(u_{\bar{p}}{}^\dagger u_{\bar{q}}) = \hat{\mathcal{K}}(\hat{\mathcal{K}} u_p{}^\dagger u_q) = u_p{}^\dagger u_q\end{aligned}$$

because a charge density behaves as a boson, for which $\hat{\mathcal{K}}^2 = 1$. The relations between the 16 types of two-electron integrals arising from all possible combinations of barred and unbarred spinors can then be determined with the use of this operator and the single bar replacement operator. Thus, for the Coulomb interaction,

$$\begin{aligned}(pq|rs) &= \hat{T}_{rs}(pq|rs) = \hat{T}_{pq}(pq|rs) = \hat{T}_{pq}\hat{T}_{rs}(pq|rs) \\ &= (pq|\bar{s}\bar{r}) = (\bar{q}\bar{p}|rs) = (\bar{q}\bar{p}|\bar{s}\bar{r})\end{aligned}$$

giving a reduction of a factor of 4. Application of the single bar replacement operator to this expression yields expressions for all the other integral types, and time reversal reduces the number of unique integrals by a factor of 4.

Matrix elements

When applied to a current density, the Kramers permutation operator gives

$$\begin{aligned}\hat{T}_{pq}\ u_p{}^\dagger \boldsymbol{\alpha}\ u_q &= (u_{\bar{p}}{}^\dagger \boldsymbol{\alpha}\ u_{\bar{q}})^\dagger = u_{\bar{q}}{}^\dagger \boldsymbol{\alpha}\ u_{\bar{p}} \\ &= \hat{\mathcal{K}}(u_{\bar{p}}{}^\dagger \boldsymbol{\alpha}\ u_{\bar{q}}) = -\hat{\mathcal{K}}(\hat{\mathcal{K}} u_p{}^\dagger \boldsymbol{\alpha}\ u_q) = -u_p{}^\dagger \boldsymbol{\alpha}\ u_q\end{aligned}$$

Thus, for the Gaunt (or Breit) interaction,

$$(p\boldsymbol{\alpha}q|r\boldsymbol{\alpha}s) = -(p\boldsymbol{\alpha}q|\bar{s}\boldsymbol{\alpha}\bar{r}) = -(\bar{q}\boldsymbol{\alpha}\bar{p}|r\boldsymbol{\alpha}s) = (\bar{q}\boldsymbol{\alpha}\bar{p}|\bar{s}\boldsymbol{\alpha}\bar{r})$$

The number of integrals is again reduced by a factor of 4.

If the Coulomb and Gaunt interactions are combined into a single interaction, there is of course a loss of one of these relations, and then the only relation between the integrals involves permutation of both pair indices. However, the factor of 2 lost by this procedure is regained by the use of a single interaction instead of two separate interactions.

Note: The permutational symmetry of the integrals is reduced relative to the nonrelativistic case because the functions are complex. Thus we have

$$(pq|rs) = (qp|sr)^* \neq (qp|rs)$$

Particle interchange symmetry is of course retained.

Kramers-restricted Hamiltonian

With the use of these relations, the Kramers-restricted Hamiltonian may be written as

$$
\begin{aligned}
\hat{H} = \sum_{pq} \Big[h_{pq} X^{+}_{pq} + \frac{1}{2} \left(h_{\bar{p}q} X^{+}_{\bar{p}q} + h_{p\bar{q}} X^{+}_{p\bar{q}} \right) \Big] \\
+ \frac{1}{2} \sum_{pqrs} \Big[(pq|rs) x^{++}_{pq,rs} - (p\boldsymbol{\alpha}q|r\boldsymbol{\alpha}s) x^{--}_{pq,rs} \\
+ (\bar{p}q|rs) x^{++}_{\bar{p}q,rs} - (\bar{p}\boldsymbol{\alpha}q|r\boldsymbol{\alpha}s) x^{--}_{\bar{p}q,rs} \\
+ (p\bar{q}|rs) x^{++}_{p\bar{q},rs} - (p\boldsymbol{\alpha}\bar{q}|r\boldsymbol{\alpha}s) x^{--}_{p\bar{q},rs} \Big] \\
+ \frac{1}{4} \sum_{pqrs} \Big[(\bar{p}q|r\bar{s}) x^{++}_{\bar{p}q,r\bar{s}} - (\bar{p}\boldsymbol{\alpha}q|r\boldsymbol{\alpha}\bar{s}) x^{--}_{\bar{p}q,r\bar{s}} \Big] \\
+ \frac{1}{8} \sum_{pqrs} \Big[(\bar{p}q|\bar{r}s) x^{++}_{\bar{p}q,\bar{r}s} - (\bar{p}\boldsymbol{\alpha}q|\bar{r}\boldsymbol{\alpha}s) x^{--}_{\bar{p}q,\bar{r}s} \\
+ (p\bar{q}|r\bar{s}) x^{++}_{p\bar{q},r\bar{s}} - (p\boldsymbol{\alpha}\bar{q}|r\boldsymbol{\alpha}\bar{s}) x^{--}_{p\bar{q},r\bar{s}} \Big]
\end{aligned}
$$

Many-particle states

A many-particle (Fock space) state in unrestricted form is given by

$$|K\rangle = \prod_k^n a_k^\dagger |vac\rangle.$$

In Kramers-restricted form we may partition the creation operators into barred and unbarred sets, and anticommute the barred operators to the right, so that

$$|K\rangle = \prod_p a_p^\dagger \prod_q a_{\bar{q}}^\dagger |vac\rangle.$$

The two products of creation operators are in nonrelativistic theory termed alpha and beta strings. To avoid confusion with these operators but also to retain the similarity between Kramers pairs and spin-orbital pairs, we will term these A and B strings:

$$\hat{A}(N_A) = \prod_p^{N_A} a_p^\dagger$$

$$\hat{B}(N_B) = \prod_p^{N_B} a_{\bar{p}}^\dagger.$$

Then the many-particle state can be written

$$|K\rangle = \hat{A}(N_A)\hat{B}(N_B)|vac\rangle.$$

Many-particle states

For a closed shell system, it is often convenient to reorder these so that the Kramers pair operators are together. Thus we define a Kramers pair creation operator

$$\hat{O}_p = a_p{}^\dagger \, a_{\bar{p}}{}^\dagger$$

and a closed-shell state is then

$$|K\rangle = \prod_p \hat{O}_p |vac\rangle.$$

It is clear that $\hat{O}_p$ is symmetric under time-reversal, for

$$\hat{\mathcal{K}}\hat{O}_p = \hat{\mathcal{K}} a_p{}^\dagger \, a_{\bar{p}}{}^\dagger = -a_{\bar{p}}{}^\dagger a_p{}^\dagger = a_p{}^\dagger \, a_{\bar{p}}{}^\dagger = \hat{O}_p$$

Thus, all closed-shell states are symmetric under time-reversal.

Many-particle states

We now consider time-reversal on a general many-particle state (determinant), which we write in terms of A and B strings,

$$|I, N_A; J, N_B\rangle = \hat{A}(N_A)\hat{B}(N_B)|vac\rangle.$$

where I and J index a particular selection of A and B spinors. Operating with $\hat{\mathcal{K}}$ on this determinant produces the result

$$\begin{aligned}
\hat{\mathcal{K}}\ |I, N_A; J, N_B\rangle &= (-1)^{N_B}\ \hat{B}(I, N_A)\ \hat{A}(J, N_B)\ |vac\rangle \\
&= (-1)^{N_B(N_A+1)}\ \hat{A}(J, N_B)\ \hat{B}(I, N_A)\ |vac\rangle \\
&= (-1)^{N_B(N_A+1)}\ |J, N_B; I, N_A\rangle.
\end{aligned}$$

Operating again produces the result

$$\begin{aligned}
\hat{\mathcal{K}}^2\ |I, N_A; J, N_B\rangle &= \hat{\mathcal{K}}\ (-1)^{N_B}\ \hat{B}(I, N_A)\ \hat{A}(J, N_B)\ |vac\rangle \\
&= (-1)^{N_A+N_B}\ \hat{A}(I, N_A)\ \hat{B}(J, N_B)\ |vac\rangle \\
&= (-1)^{N}\ |I, N_A; J, N_B\rangle.
\end{aligned}$$

Many-particle states

From the above, the following points can be made about time reversal of many-electron states.

- Double time reversal produces a phase which depends on the number of electrons. If N is odd, there is a change of phase; if N is even, there is no change of phase.
- This is because a system with an odd number of fermions behaves like a fermion, but a system with an even number of fermions behaves like a boson.
- If we define $|\bar{M}\rangle = \hat{\mathcal{K}}|M\rangle$, then $\hat{\mathcal{K}}|\bar{M}\rangle = (-1)^N|M\rangle$ defines the phase relation for time-reversal of the time-reversed many-electron state.

For the definition of $|\bar{M}\rangle$, the phase is also determined by whether N is even or odd. If N is even, then

- $N_B = N_A + 2k$ for integer k
- N_B can be replaced by N_A
- $N_A(N_A + 1)$ must be even and the phase is $+1$

If N is odd,

- $N_B = N_A + 2k + 1$ for integer k
- N_B can be replaced by $N_A + 1$
- $N_B(N_A + 1)$ differs from the even case by $N_A + 1$ which determines the phase.

N-particle Hamiltonian

We now consider the behaviour of matrix elements of the Hamiltonian under time reversal.

$$\hat{\mathcal{K}}^2 \ \langle L|\hat{H}|M\rangle = \langle \bar{L}|\hat{H}|\bar{M}\rangle^* = \langle L|\hat{H}|M\rangle$$
$$\hat{\mathcal{K}}^2 \ \langle L|\hat{H}|\bar{M}\rangle = (-1)^N \langle \bar{L}|\hat{H}|M\rangle^* = \langle L|\hat{H}|\bar{M}\rangle$$

Thus the relations between the matrix elements of $\hat{H}$ are also determined by the parity of N.

For the case of N even, it is possible to construct a real basis, i.e. one in which the Hamiltonian matrix is real.

$$|M^s\rangle \ = \ \frac{i^{(1-s)/2}}{\sqrt{2(1+\delta_{M,\bar{M}})}}\left(\ |M\rangle + s|\bar{M}\rangle \ \right)$$

Both of these functions are symmetric under time reversal, by construction. Transforming the Hamiltonian matrix elements into this basis and making use of their behaviour under time reversal, we have

$$H_{L^+M^+} = \text{Re}\ (H_{LM} + H_{L\bar{M}})/G_{LM}$$
$$H_{L^+M^-} = -\text{Im}\ (H_{LM} - H_{L\bar{M}})/G_{LM}$$
$$H_{L^-M^+} = \text{Im}\ (H_{LM} + H_{L\bar{M}})/G_{LM}$$
$$H_{L^-M^-} = \text{Re}\ (H_{LM} - H_{L\bar{M}})/G_{LM}$$

where $G_{LM} \ = [(1+\delta_{L,\bar{L}})(1+\delta_{M,\bar{M}})]^{1/2}$ is a normalization factor required when one or both of the determinants is a closed shell.

Mean-field theory

Having developed much of the machinery for relativistic calculations, we now turn to the question of determining the one-particle functions in which the various operators are expanded. In other words, we are now concerned with developing a mean-field theory.

We will use the exponential operator $\hat{U} = \exp(i\hat{\lambda})$ which expresses a general rotation of the basis, but instead of the Hermitian $\hat{\lambda}$ we will use the anti-Hermitian $\hat{\kappa} = i\hat{\lambda}$. We choose a reference state (initial guess), $|0\rangle$, and apply the general rotation to it to obtain a new state, $|0'\rangle = \hat{U}|0\rangle$. The energy of the reference state in the rotated basis is simply

$$E' = \langle 0'|\hat{H}|0'\rangle = \langle 0|e^{-\hat{\kappa}}\hat{H}e^{\hat{\kappa}}|0\rangle$$

The operator can be expanded using the Baker-Campbell-Hausdorff expansion,

$$e^{-\hat{\kappa}}\hat{H}e^{\hat{\kappa}} = \hat{H} + \big[\hat{H},\hat{\kappa}\big] + \frac{1}{2!}\big[\big[\hat{H},\hat{\kappa},\big],\hat{\kappa}\big] + \ldots$$

so that the energy may finally be written

$$E' = \langle 0|\hat{H}|0\rangle + \langle 0|\big[\hat{H},\hat{\kappa}\big]|0\rangle + \frac{1}{2!}\langle 0|\big[\big[\hat{H},\hat{\kappa},\big],\hat{\kappa}\big]|0\rangle + \ldots$$

Structure of κ

We must now choose an explicit form for κ. In the first instance, we will take the Dirac picture, in which the negative-energy states are not reinterpreted as positrons; thus

$$\hat{\kappa} = \sum_{pq} \kappa_{pq} \; a_p{}^\dagger \, a_q$$

where the sum extends over both positive and negative energy states. Substituting $\hat{\kappa}$ into the first term in the energy expansion,

$$\langle 0| \big[\hat{H}, \hat{\kappa}\big] |0\rangle = \sum_{pq} \kappa_{pq} \big[\langle 0|\hat{H} \, a_p{}^\dagger \, a_q \, |0\rangle - \langle 0|a_p{}^\dagger \, a_q \, \hat{H}|0\rangle\big],$$

we may make the following observations:

- If p and q are both unoccupied indices, the action on the reference produces zero, and hence the term in square brackets will be zero.
- If p and q are both occupied indices, the action on the reference produces zero unless $p = q$. But in this case the result is the reference and again the term vanishes.

Similar arguments can be marshaled for the higher order terms.

Structure of κ

The conclusion is that the energy is invariant to rotations within the occupied spinor space and within the unoccupied space. The matrix elements of κ within these spaces may therefore be set to zero, and the only matrix elements to be considered are those which connect occupied and unoccupied spinors.

Now, the definition of the occupied and unoccupied spinors depends on the interpretation used.

- If the negative energy states are empty (the original or "empty Dirac" picture), the occupied spinors are the electron states, and the unoccupied spinors consist of the virtual electron spinors and the positron spinors.
- If the negative energy states are filled (the reinterpreted or "filled Dirac" picture), the occupied spinors must include these states, and the only empty spinors are the virtual electron spinors.

Thus the structure of $\hat{\kappa}$ will depend on whether the negative energy states are filled or not. Since the reinterpreted Dirac picture was unsatisfactory due to the infinitely charged vacuum, we must now look at the QED picture.

Structure of $\boldsymbol{\kappa}$

In the QED picture, the normal-ordered operator is

$$:\hat{\kappa}:= \sum_{pq} \left[\kappa_{pq}\ a_p{}^\dagger\, a_q + \kappa_{p\tilde{q}}\ a_p{}^\dagger\, \tilde{a}_q{}^\dagger + \kappa_{\tilde{p}q}\ \tilde{a}_p\, a_q - \kappa_{\tilde{p}\tilde{q}}\ \tilde{a}_q{}^\dagger\, \tilde{a}_p\right].$$

We now substitute each term in $\hat{\kappa}$ in turn into the first term in the energy expansion. For the first term, we draw the same conclusions as for the Dirac picture, (whether filled or empty). For the second term,

$$\sum_{pq} \kappa_{pq} \left[\langle 0|\hat{H}\, a_p{}^\dagger\, \tilde{a}_q{}^\dagger\, |0\rangle - \langle 0|a_p{}^\dagger\, \tilde{a}_q{}^\dagger\, \hat{H}|0\rangle\right],$$

- The first integral will be zero unless p is an unoccupied index.
- The second integral is always zero.

For the third term in $\hat{\kappa}$, the converse holds. The strange result for these two terms is that the energy apparently depends on the definition of the unoccupied orbitals. However, these terms involve pair creation and destruction, and therefore must involve empty orbitals. The fourth term is identically zero.

The conclusion from these arguments is the same as in the reinterpreted Dirac picture: the elements of $\boldsymbol{\kappa}$ are non-zero only when one index is a virtual electron index and the other is either an occupied electron or a positron index.

Structure of $\boldsymbol{\kappa}$

However, there is a problem with this approach. The product of two normal-ordered operators is not itself a normal-ordered operator, but must be put into normal order. Therefore, to consider the effect of $\hat{\kappa}$ only on the reference — even though it is normal-ordered — is not consistent with QED, in which all operators must be placed in normal order before their matrix elements are taken.

We must therefore consider the commutator as a whole. For simplicity, we choose the one-particle Dirac Hamiltonian, but the idea can be extended to the many-particle Hamiltonian. Here, the QED picture gives

$$
\begin{aligned}
:\big[\hat{h},\hat{\kappa}\big]:= &\sum_{pqr}\big(a_r{}^\dagger a_q + a_r{}^\dagger \tilde{a}_q{}^\dagger + \tilde{a}_r a_q - \tilde{a}_q{}^\dagger \tilde{a}_r\big)\big(h_{pq}\kappa_{rp} + h_{\tilde{p}q}\kappa_{r\tilde{p}}\big)\\
&+\sum_{pqs}\big(a_p{}^\dagger a_s + a_p{}^\dagger \tilde{a}_s{}^\dagger + \tilde{a}_p a_s - \tilde{a}_s{}^\dagger \tilde{a}_p\big)\big(h_{pq}\kappa_{qs} + h_{p\tilde{q}}\kappa_{\tilde{q}s}\big)
\end{aligned}
$$

and the Dirac picture gives

$$
\begin{aligned}
\big[\hat{h},\hat{\kappa}\big] = &\sum_{pqr}\big(a_r{}^\dagger a_q + a_r{}^\dagger \tilde{a}_q + \tilde{a}_r{}^\dagger a_q + \tilde{a}_r{}^\dagger \tilde{a}_q\big)\big(h_{pq}\kappa_{rp} + h_{\tilde{p}q}\kappa_{r\tilde{p}}\big)\\
&+\sum_{pqs}\big(a_p{}^\dagger a_s + a_p{}^\dagger \tilde{a}_s + \tilde{a}_p{}^\dagger a_s + \tilde{a}_p{}^\dagger \tilde{a}_s\big)\big(h_{pq}\kappa_{qs} + h_{p\tilde{q}}\kappa_{\tilde{q}s}\big)
\end{aligned}
$$

Structure of $\boldsymbol{\kappa}$

If we now take the reference expectation of this operator, we find that in the QED picture and in the empty Dirac picture, all terms except the first vanish, because the reference contains only electrons. But in the filled Dirac picture, all terms in principle survive. Thus the properly normal-ordered QED picture gives the same result for this term as the empty Dirac picture, and not the filled Dirac picture.

Furthermore, the reference expectation value ensures that one index of the matrix elements of $\boldsymbol{\kappa}$ must be an occupied index, and here the empty Dirac and the QED picture have the same occupied spinors, whereas the filled Dirac picture has the negative energy states occupied.

Comparing the normal-ordered operator with the Dirac picture operator, we may formulate the following rules for the picture change:

- Determine the operator in the empty Dirac picture.
- Re-interpret the negative energy operators.
- Place them in normal order.

Conclusions about pictures

The conclusion of this analysis is that the normal-ordered QED picture as presented here is equivalent to the empty Dirac picture. Since the reference expectation of the first commutator gives the same results in both pictures, the remaining commutators must also give the same result. Put another way, if the gradient of the energy with respect to the parameters is the same, all the higher derivatives must be the same as well. We may therefore proceed with the standard Dirac-Hartree-Fock method.

Stationarity conditions

We now consider the operators with time-reversal symmetry included. The energy is stationary when the gradient with respect to the parameters is zero, i.e.

$$\begin{aligned}\frac{\partial E'}{\partial \kappa_{tu}} &= \langle 0|\big[\hat{H}, X_{tu}^{-}\big]|0\rangle \\ &\quad + \frac{1}{2}\sum_{vw} \kappa_{vw} \Big[\langle 0|\big[\big[\hat{H}, X_{tu}^{-}\big], X_{vw}^{-}\big]|0\rangle \\ &\qquad\qquad + \langle 0|\big[\big[\hat{H}, X_{vw}^{-}\big], X_{tu}^{-}\big]|0\rangle\Big] + \ldots \; = 0\end{aligned}$$

At the stationary point, the values of the parameters must be zero, so the condition for a stationary point is that

$$g_{tu} = \langle 0|\big[\hat{H}, X_{tu}^{-}\big]|0\rangle = 0.$$

This is the gradient with respect to spinor rotations.

The nature of the stationary point is determined by the eigenvalues of the Hessian, which here is given as

$$K_{tu,vw} = \frac{1}{2}\Big\{\langle 0|\big[\big[\hat{H}, X_{tu}^{-}\big], X_{vw}^{-}\big]|0\rangle + \langle 0|\big[\big[\hat{H}, X_{vw}^{-}\big], X_{tu}^{-}\big]|0\rangle\Big\}$$

It may be shown (for example by considering the hydrogen atom) that the Hessian is *not* positive definite, but has as many negative eigenvalues as there are negative energy states. Thus the optimization of the wave function becomes a *minimax* problem.

Fock matrices

The gradient may be expressed in terms of the generalized Fock matrices,

$$g_{tu} = f_{tu} - f^*_{ut}$$

where the Coulomb Fock matrix is

$$\begin{aligned} f_{tu} = & \sum_q \Big\{ D^+_{tq}\ h_{uq} + D^+_{t\bar{q}}\ h_{u\bar{q}} \Big\} \\ & + \sum_{qrs} \Big\{ P^{++}_{tq,rs}\ (uq \mid rs) + P^{++}_{t\bar{q},rs}\ (u\bar{q} \mid rs) \\ & \quad + \frac{1}{2} \big[P^{++}_{t\bar{q},\bar{r}s}\ (u\bar{q} \mid \bar{r}s) + P^{++}_{t\bar{q},r\bar{s}}\ (u\bar{q} \mid r\bar{s}) \\ & \qquad + P^{++}_{tq,\bar{r}s}\ (uq \mid \bar{r}s) + P^{++}_{tq,r\bar{s}}\ (uq \mid r\bar{s}) \big] \Big\}. \end{aligned}$$

and D^+_{pr} and $P^{++}_{pr,st}$ are the time-reversal invariant one- and two-electron Kramers-reduced density matrices, respectively

$$\begin{aligned} D^+_{pq} &= \langle 0 \mid \hat{X}^+_{pq} \mid 0 \rangle, \\ P^{++}_{pq,rs} &= \langle 0 \mid \hat{x}^{++}_{pq,rs} \mid 0 \rangle. \end{aligned}$$

The Fock matrix for one barred index can be derived by using the single bar replacement operator.

The addition of the Breit or Gaunt interaction is made by replacing the integrals and substituting $P^{--}_{pq,rs}$ for $P^{++}_{pq,rs}$.

Fock matrices

For a closed shell system, the density matrices reduce to

$$\begin{aligned} D^{s}_{pq} &= \delta_{pq} + s\ \delta_{\bar{q}\bar{p}}, \\ P^{s_1 s_2}_{pq,rs} &= (\delta_{pq} + s_1\delta_{\bar{q}\bar{p}})(\delta_{rs} + s_2\delta_{\bar{s}\bar{r}}) \\ &\quad - \delta_{ps}\delta_{qr} - s_1\delta_{\bar{p}r}\delta_{\bar{q}s} - s_2\delta_{p\bar{r}}\delta_{q\bar{s}} - s_1 s_2\delta_{\bar{p}\bar{s}}\delta_{\bar{q}\bar{r}}, \end{aligned}$$

with barred density matrices derived as usual by the use of $\hat{K}_p$. The Fock matrices become

$$F_{tu} = h_{ut} + \sum_r \left\{2(ut|rr) - (ur|rt) - (u\bar{r}|\bar{r}t)\right\}$$

$$F_{t\bar{u}} = h_{\bar{u}t} + \sum_r \left\{2(\bar{u}t|rr) - (\bar{u}r|rt) - (\bar{u}\bar{r}|\bar{r}t)\right\}$$

where $f_{tu} = 2F_{tu}$. For the Breit or Gaunt terms, the direct contribution vanishes due to the appearance of an x^{--} operator, and the only terms which survive are the exchange terms analogous to the ones above. Having defined the Fock matrix, the Dirac-Hartree-Fock (DHF) equation

$$\mathbf{F}\ \mathbf{c} = \mathbf{c}\ \boldsymbol{\epsilon},$$

may be solved by standard methods. These will be discussed in detail later.

Some results from DHF calculations are now presented, to give an estimate of the magnitude of relativistic effects.

Small component electron density

The percentage contribution to the electron density of each spinor from the small component is presented here for Pb.

Spinor	Value	Spinor	Value
$1s_{1/2}$	9.7819	$4d_{3/2}$	0.2017
$2s_{1/2}$	2.2586	$4d_{5/2}$	0.1940
$2p_{1/2}$	2.2635	$4f_{5/2}$	0.1572
$2p_{3/2}$	1.9923	$4f_{7/2}$	0.1541
$3s_{1/2}$	0.7617	$5s_{1/2}$	0.0663
$3p_{1/2}$	0.7548	$5p_{1/2}$	0.0582
$3p_{3/2}$	0.6771	$5p_{3/2}$	0.0496
$3d_{3/2}$	0.6627	$5d_{3/2}$	0.0324
$3d_{5/2}$	0.6417	$5d_{5/2}$	0.0301
$4s_{1/2}$	0.2595	$6s_{1/2}$	0.0091
$4p_{1/2}$	0.2499	$6p_{1/2}$	0.0052
$4p_{3/2}$	0.2225	$6p_{3/2}$	0.0038

Overlaps of radial functions

The percentage deviation from unit overlap is presented for the valence large component radial functions for the Group 14 elements. The overlap discrepancies between the nonrelativistic and relativistic functions, and between the spin-orbit components of the p functions are given.

	n	$\langle s\|s_{1/2}\rangle$	$\langle p\|p_{1/2}\rangle$	$\langle p\|p_{3/2}\rangle$	$\langle p_{1/2}\|p_{3/2}\rangle$
C	2	0.0428	0.0033	0.0033	0.0033
Si	3	0.0026	0.0015	0.0013	0.0013
Ge	4	0.0552	0.0193	0.0026	0.0262
Sn	5	0.2325	0.1442	0.0062	0.1680
Pb	6	2.4354	1.7052	0.0458	1.8503

Note the following

- The discrepancy for Si is smaller than for C, probably due to screening
- The $p_{3/2}$ is much closer to the NR function than the $p_{1/2}$
- There is less overlap between the spin-orbit components than with the NR function for the heavier elements.

Coupling scheme changes

The ground state J = 0 percentage eigenvector composition in LS and jj coupling is presented for Group 14 elements.

	n	$np^2\ {}^3P_0$	$np^2\ {}^1S_0$	$np^2_{1/2}$	$np^2_{3/2}$
C	2	99.9997	0.0003	66.8252	33.1748
Si	3	99.9903	0.0097	67.5908	32.4092
Ge	4	99.6877	0.3123	71.8230	28.1770
Sn	5	98.1042	1.8958	78.8923	21.1077
Pb	6	88.9736	11.0264	92.5217	7.4783

Dynamics vs screening

The orbital eigenvalues in eV for a single electron outside a closed shell are presented. For each system, the Dirac and Schrödinger equations are solved in either a relativistic or a nonrelativistic (fixed) core potential. This shows how much of the relativistic changes are due to dynamics and how much to screening.

Dynamics:	Dirac		Schrödinger	
Potential:	R	NR	R	NR
Au $6s$	7.94	7.97	6.18	6.01
Tl $6p_{1/2}$	5.81	6.79	4.58	5.24
Tl $6p_{3/2}$	4.79	5.63	4.46	5.24
Lu $5d_{3/2}$	5.25	7.32	4.74	6.63
Lu $5d_{5/2}$	5.01	6.90	4.81	6.63

Relativistic effects on atomic orbitals

Here some of the effects of relativity on atomic orbitals are summarized.

- Contraction due to relativistic dynamics.
- Expansion due to greater screening of nucleus.
- Spin-orbit splitting, increasing with Z.
- Change from LS to jj coupling as Z increases.

For valence orbitals these effects balance in different ways:

- s and $p_{1/2}$ orbitals contract. Both have s-character, either in the large or the small component. $6s$ subshell relatively inert, stabilizing oxidation state $q - 2$ (q is valence occupation).
- For $p_{3/2}$ orbitals the dynamics and screening approximately cancel. Thus the behaviour of the late p block elements shows little effect of relativity on many properties.
- For orbitals with $\ell > 1$ the screening effect dominates. The $5d$ expands, for example, and contributes much more to bonding.

Basis set expansion

There are several possible choices for the expansion of the molecular spinors in a basis set. One could expand each component in a basis, or one could take advantage of the angular coupling, and expand the radial large and small components in a basis set. The second can of course be reduced to the first, but it has some advantages which will be spelled out later. Thus we will start with the second option.

In the early days of basis set DHF calculations, some problems were encountered obtaining the correct nonrelativistic limit. Doubts have also been expressed over the variational safety of the DHF method in a finite basis set. We will therefore examine these questions using the one-electron Dirac equation as an example.

2-spinor Dirac equation

We expand the large and small components in a basis of 2-spinors,

$$\Psi_i^L = \sum_{\mu=1}^{n} c_{\mu i}^L \ \phi_\mu^L, \quad \Psi_i^S = \sum_{\mu=1}^{n} c_{\mu i}^S \ \phi_\mu^S,$$

and define the potential energy, overlap and kinetic energy matrices by

$$V_{\mu\nu}^{LL} = \langle \phi_\mu^L | V | \phi_\nu^L \rangle; \quad V_{\mu\nu}^{SS} = \langle \phi_\mu^S | V | \phi_\nu^S \rangle$$

$$S_{\mu\nu}^{LL} = \langle \phi_\mu^L | \phi_\nu^L \rangle; \quad S_{\mu\nu}^{SS} = \langle \phi_\mu^S | \phi_\nu^S \rangle$$

$$\Pi_{\mu\nu}^{LS} = \langle \phi_\mu^L | \sigma.\nabla | \phi_\nu^S \rangle; \quad \Pi_{\mu\nu}^{SL} = \langle \phi_\mu^S | \sigma.\nabla | \phi_\nu^L \rangle.$$

Note that $\Pi_{\mu\nu}^{SL} = -(\Pi_{\nu\mu}^{LS})^\dagger$, so that

$$\Pi = \begin{pmatrix} \mathbf{0} & \Pi^{LS} \\ -\Pi^{SL} & \mathbf{0} \end{pmatrix}$$

is Hermitian. The one-electron Dirac equation in a basis is

$$\begin{pmatrix} \mathbf{V}^{LL} - E \ \mathbf{S}^{LL} & c \ \Pi^{LS} \\ -c \ \Pi^{SL} & \mathbf{V}^{SS} - (2c^2 + E) \ \mathbf{S}^{SS} \end{pmatrix} \begin{pmatrix} \mathbf{c}^L \\ \mathbf{c}^S \end{pmatrix} = 0$$

Elimination of small component

Writing the 2-spinor Dirac equation as a pair of coupled equations,

$$(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\ \mathbf{c}^L + c\ \Pi^{LS}\ \mathbf{c}^S = 0$$

$$-c\ \Pi^{SL}\mathbf{c}^L + [\mathbf{V}^{SS} - (2c^2 + E)\mathbf{S}^{SS}]\ \mathbf{c}^S = 0$$

we invert the second equation (valid for $E > -2c^2$ because $\mathbf{V}$ is negative definite)

$$\mathbf{c}^S = c\ [\mathbf{V}^{SS} - (2c^2 + E)\mathbf{S}^{SS}]^{-1}\Pi^{SL}\mathbf{c}^L$$

and eliminate the small component from the first equation to obtain

$$(\mathbf{V}^{LL} - E\mathbf{S}^{LL})\mathbf{c}^L = -c\Pi^{LS}\left[\mathbf{V}^{SS} - (2c^2 + E)\mathbf{S}^{SS}\right]^{-1} c\Pi^{SL}\mathbf{c}^L$$

Using the matrix relation

$$(\mathbf{A} - \mathbf{B})^{-1} = -\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{A}(\mathbf{A} - \mathbf{B})^{-1}$$

with $2c^2\mathbf{S}^{SS} = \mathbf{B}$ and $(\mathbf{V}^{SS} - E\mathbf{S}^{SS}) = \mathbf{A}$, we obtain

$$\begin{aligned}\left\{\mathbf{V}^{LL} - E\mathbf{S}^{LL} - \frac{1}{2}\Pi^{LS}\left[\mathbf{S}^{SS}\right]^{-1}\Pi^{SL}\right\}\ \mathbf{c}^L \\ = \ -\frac{1}{2}\Pi^{LS}\left[\mathbf{S}^{SS}\right]^{-1}\left[\mathbf{V}^{SS} - E\mathbf{S}^{SS}\right] \\ \times\left[\mathbf{V}^{SS} - (2c^2 + E)\mathbf{S}^{SS}\right]^{-1}\Pi^{SL}\mathbf{c}^L\end{aligned}$$

Kinetic balance

In the nonrelativistic limit $c \rightarrow \infty$, the right hand side vanishes, leaving

$$\left\{\mathbf{V}^{LL} - E\mathbf{S}^{LL} - \frac{1}{2}\Pi^{LS}\left[\mathbf{S}^{SS}\right]^{-1}\Pi^{SL}\right\}\mathbf{c}^{L} = 0$$

The kinetic energy is therefore represented by a matrix product,

$$\mathbf{T}' = -\frac{1}{2}\Pi^{LS}\left[\mathbf{S}^{SS}\right]^{-1}\Pi^{SL}$$

This is the matrix equivalent of $(\sigma.\nabla)(\sigma.\nabla)$, which is equal to ∇^2. However, the matrix expression is *not* equal to the matrix of ∇^2 *unless* certain conditions are met.

Writing out the matrix product as

$$-2T'_{\mu\nu} = \sum_{\kappa\lambda} \langle\phi^L_\mu|\sigma.\nabla\underline{|\phi^S_\kappa\rangle\left[\mathbf{S}^{SS}\right]^{-1}_{\kappa\lambda}\langle\phi^S_\lambda|}\sigma.\nabla|\phi^L_\nu\rangle$$

it is clear that there is a projection operator onto the small component space between the $(\sigma.\nabla)$ operators. If we choose

$$\phi^S_\mu = (\sigma.\nabla)\phi^L_\mu$$

then $\Pi^{SL} = \mathbf{S}^{SS}$ and thus $\Pi^{LS} = -2\mathbf{T}' = -2\mathbf{T}$. This condition, which ensures that the kinetic energy is properly represented in the NR limit, is known as *kinetic balance*.

Lower bound

The second issue is that of a lower bound on the Dirac equation. If we transform the small component to an orthonormal basis, and define

$$\mathbf{P} = [\mathbf{S}^{SS}]^{-1/2}\Pi^{SL}; \quad 2\mathbf{T}' = \mathbf{P}^\dagger\mathbf{P}$$

$$\mathbf{W} = [\mathbf{S}^{SS}]^{-1/2}\mathbf{V}^{SS}[\mathbf{S}^{SS}]^{-1/2}/2c^2$$

the Rayleigh quotient defined using the large component equation is

$$E = \frac{\mathbf{c}^{L\dagger}\mathbf{V}^{LL}\mathbf{c}^L}{\mathbf{c}^{L\dagger}\mathbf{S}^{LL}\mathbf{c}^L} + \frac{1}{2}\,\frac{\mathbf{c}^{L\dagger}\mathbf{P}^\dagger[\mathbf{I}(1+E/2c^2)-\mathbf{W}]^{-1}\mathbf{P}\mathbf{c}}{\mathbf{c}^{L\dagger}\mathbf{S}^{LL}\mathbf{c}^L}$$

We are interested in solutions with $0 > E > -2c^2$, so that $\lambda = 1 + E/2c^2 > 0$. Since $\mathbf{W}$ is negative definite, the operator in square brackets is positive. The only non-diagonal matrix in this operator is $\mathbf{W}$. Now,

$$[\lambda\mathbf{I}-\mathbf{W}]^{-1} = [\mathbf{I}\lambda - w]^{-1} + [\mathbf{W}-w\mathbf{I}][(\mathbf{I}\lambda-\mathbf{W})(\lambda\mathbf{I}-w)]^{-1}$$

where w is a number, to be chosen so that $\mathbf{W} - w\mathbf{I}$ is a positive definite matrix. This will be the case if w is the largest negative eigenvalue of $\mathbf{W}$. Then the energy satisfies

$$E \;>\; \frac{\mathbf{c}^{L\dagger}\mathbf{V}^{LL}\mathbf{c}^L}{\mathbf{c}^{L\dagger}\mathbf{S}^{LL}\mathbf{c}^L} + \frac{1}{(\lambda-w)}\,\frac{\mathbf{c}^{L\dagger}\mathbf{T}'\mathbf{c}^L}{\mathbf{c}^{L\dagger}\mathbf{S}^{LL}\mathbf{c}^L}$$

Lower bound

The lower bound therefore depends on the value of $\lambda - w$. If this value is large, then the lower bound is the largest negative eigenvalue of the large component potential. The bound is therefore *not* the exact energy, but something more negative.

But if the virial theorem is approximately satisfied, then the value of $\lambda - w$ can be represented as $1 + (\bar{v}^L - 2v^S)/4c^2$. Provided $\bar{v}^L - 2v^S$ is not greater than $4c^2$, this can be expanded in a series, to give

$$E \;>\; \frac{\mathbf{c}^{L\dagger}\mathbf{V}^{LL}\mathbf{c}^L}{\mathbf{c}^{L\dagger}\mathbf{S}^{LL}\mathbf{c}^L} + \frac{\mathbf{c}^{L\dagger}\mathbf{T}'\mathbf{c}^L}{\mathbf{c}^{L\dagger}\mathbf{S}^{LL}\mathbf{c}^L}\left(1 - \frac{\bar{v}^L - 2v^S}{4c^2} - \ldots\right)$$

Thus the energy is bounded by a value which is somewhat smaller than the nonrelativistic energy expression. The first term is $\mathcal{O}(c^{-2})$. It can be shown that with a kinetically balanced basis, the maximum deviation below the exact energy is $\mathcal{O}(c^{-4})$.

There is therefore *no danger* of variational collapse in basis set DHF calculations. In numerical calculations on atoms, the variational safety is ensured by application of the boundary conditions at the nuclei and at infinity.

Matrix DHF equations

For a many-electron system, we first need to reduce the molecular spinor integrals into 2-component form. Thus, for the Coulomb interaction, the integrals reduce to

$$\begin{aligned}(ij|kl) &= (\mathbf{i}^L\mathbf{j}^L|\mathbf{k}^L\mathbf{l}^L) + (\mathbf{i}^L\mathbf{j}^L|\mathbf{k}^S\mathbf{l}^S) \\ &\quad + (\mathbf{i}^S\mathbf{j}^S|\mathbf{k}^L\mathbf{l}^L) + (\mathbf{i}^S\mathbf{j}^S|\mathbf{k}^S\mathbf{l}^S)\end{aligned}$$

and for the magnetic interactions,

$$\begin{aligned}(i\boldsymbol{\alpha}j|k\boldsymbol{\alpha}l) &= (\mathbf{i}^L\mathbf{j}^S|\mathbf{k}^S\mathbf{l}^L) + (\mathbf{i}^S\mathbf{j}^L|\mathbf{k}^L\mathbf{l}^S) \\ &\quad - (\mathbf{i}^L\mathbf{j}^S|\mathbf{k}^L\mathbf{l}^S) - (\mathbf{i}^S\mathbf{j}^L|\mathbf{k}^S\mathbf{l}^L),\end{aligned}$$

where the combinations of sigma matrices are now implied in the integrals by the fact that each density connects a large and a small component function. The negative signs result from the factors of i.

We define the spinor density matrices by

$$D^{LL}_{\mu\nu} = \sum_{i=1}^{N} c^{L\dagger}_{\mu i} c^{L}_{\nu i} \quad D^{LS}_{\mu\nu} = \sum_{i=1}^{N} c^{L\dagger}_{\mu i} c^{S}_{\nu i}$$

$$D^{SL}_{\mu\nu} = \sum_{i=1}^{N} c^{S\dagger}_{\mu i} c^{L}_{\nu i} \quad D^{SS}_{\mu\nu} = \sum_{i=1}^{N} c^{S\dagger}_{\mu i} c^{S}_{\nu i}$$

Note that $D^{LS}_{\mu\nu} = D^{SL\dagger}_{\nu\mu}$, so that

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}^{LL} & \mathbf{D}^{LS} \\ \mathbf{D}^{SL} & \mathbf{D}^{SS} \end{pmatrix}$$

is hermitian.

Matrix DHF equations

The 2-spinor matrix DHF equations including the magnetic interactions have the following form:

$$
\begin{aligned}
F^{LL}_{\mu\nu} = V^{LL}_{\mu\nu} &+ \sum_{\kappa\lambda} D^{LL}_{\kappa\lambda} \big\{ 2\,(\,\mu^L\nu^L \mid \kappa^L\lambda^L\,) - (\,\mu^L\lambda^L \mid \kappa^L\nu^L\,) \\
&\qquad - (\,\mu^L\bar{\kappa}^L \mid \bar{\lambda}^L\nu^L\,)\big\} \\
&+ \sum_{\kappa\lambda} D^{SS}_{\kappa\lambda} \big\{ 2\,(\,\mu^L\nu^L \mid \kappa^S\lambda^S\,) + (\,\mu^L\lambda^S \mid \kappa^S\nu^L\,) \\
&\qquad + (\,\mu^L\bar{\kappa}^S \mid \bar{\lambda}^S\nu^L\,)\big\}
\end{aligned}
$$

$$
\begin{aligned}
F^{LS}_{\mu\nu} = \Pi^{LS}_{\mu\nu} &- \sum_{\kappa\lambda} D^{LS}_{\kappa\lambda} \big\{ (\,\mu^L\lambda^S \mid \kappa^L\nu^S\,) + (\,\mu^L\bar{\kappa}^L \mid \bar{\lambda}^S\nu^S\,)\big\} \\
&- \sum_{\kappa\lambda} D^{SL}_{\kappa\lambda} \big\{ (\,\mu^L\lambda^L \mid \kappa^S\nu^S\,) + (\,\mu^L\bar{\kappa}^S \mid \bar{\lambda}^L\nu^S\,)\big\}
\end{aligned}
$$

$$
\begin{aligned}
F^{SS}_{\mu\nu} = V^{SS}_{\mu\nu} &+ \sum_{\kappa\lambda} D^{LL}_{\kappa\lambda} \big\{ 2\,(\,\mu^S\nu^S \mid \kappa^L\lambda^L\,) + (\,\mu^S\lambda^L \mid \kappa^L\nu^S\,) \\
&\qquad + (\,\mu^S\bar{\kappa}^L \mid \bar{\lambda}^L\nu^S\,)\big\} \\
&+ \sum_{\kappa\lambda} D^{SS}_{\kappa\lambda} \big\{ 2\,(\,\mu^S\nu^S \mid \kappa^S\lambda^S\,) - (\,\mu^S\lambda^S \mid \kappa^S\nu^S\,) \\
&\qquad - (\,\mu^S\bar{\kappa}^S \mid \bar{\lambda}^S\nu^S\,)\big\}
\end{aligned}
$$

Basis set size

At this point, is is worth considering some of the practical requirements of the solution of the DHF equations as compared to the nonrelativistic HF equations. The critical issue is the number of two-electron integrals. The following observations can be made.

- The spin can be factored out of the NR integrals, but in the R integrals, the spin is coupled in. Therefore one expects a factor of 2 increase for each basis function.
- The integrals and the basis is complex, so there is a factor of 2 increase for the complex integral values and 2 for the loss of one degree of permutational symmetry.
- The small component basis is the same size as the large component basis, giving a factor of 2 increase for each basis function.

The net result is that there would be an increase of a factor of 1024 more integrals in a DHF calculation than in a NRHF calculation. However:

- Time-reversal symmetry reduces the integral number by a factor of 4.
- Integrals with an odd number of large and small components never appear, giving a factor of 2 reduction.
- If only the Coulomb interaction is used, there is another factor of 2 reduction

Thus, for a Dirac-Coulomb SCF calculation there are 64 times as many integrals as in a NR SCF calculation.

Scalar basis sets

Since the available integral codes almost invariably work in a basis set of real spherical or Cartesian Gaussians, it is of value to decompose the 2-spinors into their scalar components and to represent the DHF equations in the scalar basis.

We start with the definition of the atomic jj-coupled 2-spinors,

$$\begin{aligned}\boldsymbol{\mu}_{An\kappa m} &= R_{n\kappa}(r)\ \chi_{\kappa,m}(\theta,\phi) \\ &= \frac{R_{n\kappa}(r)}{\sqrt{(2\ell+1)}} \begin{pmatrix} a\sqrt{(\ell+ma+\frac{1}{2})}\, Y_{\ell m-1/2}(\theta,\phi) \\ \sqrt{(\ell-ma+\frac{1}{2})}\, Y_{\ell m+1/2}(\theta,\phi) \end{pmatrix}.\end{aligned}$$

where $a = (-1)^{\ell-j+1/2} = -\kappa/|\kappa|$. Operating on the atomic 2-spinor with the time-reversal operator produces the spinor with opposite m value:

$$\hat{\mathcal{K}}\,\boldsymbol{\mu}_{An\kappa m} = (-1)^{\ell-j+m}\boldsymbol{\mu}_{An\kappa -m}.$$

The 2-spinor basis is chosen to incorporate the phase factor for the time-reversed spinor,

$$\begin{aligned}\bar{\boldsymbol{\mu}}^{L}_{An\kappa m} &= (-1)^{\ell-j+m}\boldsymbol{\mu}^{L}_{An\kappa -m}, \\ \bar{\boldsymbol{\mu}}^{S}_{An\kappa m} &= (-1)^{\ell-j+m+1}\boldsymbol{\mu}^{S}_{An\kappa -m},\end{aligned}$$

where $m > 0$. The sign for the small component comes from the factor of i. This defines the phases for the 2-spinor basis.

Scalar basis sets

We now consider a more general expansion of the 2-spinor functions. First, the 2-spinor functions are expressed in terms of complex scalar functions and spin functions

$$\boldsymbol{\phi}_\mu^X = \begin{pmatrix} \phi_{\mu\alpha}^X \\ \phi_{\mu\beta}^X \end{pmatrix} \equiv \phi_{\mu\alpha}^X \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \phi_{\mu\beta}^X \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \sum_\tau \phi_{\mu\tau}^X \, \boldsymbol{\eta}(\tau), \quad X = L, S.$$

The complex scalar functions are further reduced to real functions:

$$\phi_{\mu\tau}^X = \phi_{\mu\tau}^{XR} + i\phi_{\mu\tau}^{XI} \equiv \sum_{C=0,1} i^{\,C} \phi_{\mu\tau}^{XC}, \quad \tau = \alpha, \beta,$$

where $C = 0$ and 1 correspond to labels R and I. The real functions are expressed in terms of spherical harmonic or Cartesian Gaussian functions ϕ_a^X:

$$\phi_{\mu\tau}^{XC} = \sum_a T_{a\mu}^{\tau XC} \phi_a^X, \quad C = R, I,$$

so that finally, the 2-spinors may be written in terms of Gaussian basis functions as

$$\boldsymbol{\phi}_\mu^X = \sum_{\tau C a} i^{\,C} \, T_{a\mu}^{\tau XC} \, \phi_a^X \, \boldsymbol{\eta}(\tau), \quad X = L, S.$$

Scalar basis sets

We may make the following observations about the basis definition.

- The superscript X attached to the scalar basis functions indicates that in general different scalar functions will be used for large and small components and the basis function is indexed accordingly.
- The sum over a is usually limited to a few terms. For real spherical Gaussians, there is only one term except for the positive kappa small component spinors which consist of two terms.
- The coefficients $T_{a\mu}^{\tau XC}$ include coupling coefficients and normalization factors.
- Each real function ϕ_a contributes to either the real or the imaginary part of a spinor, not to both: the sum over C and a is actually only a single sum, since C is determined by the properties of ϕ_a.

Integrals in scalar basis

For one-electron operators, the 2-spinor and scalar matrix elements are

$$A_{\mu\nu}^{XY} = \sum_{\tau\upsilon}\sum_{CD}\sum_{ab} i^{\,(D-C)}\, T_{a\mu}^{\tau XC} T_{b\nu}^{\upsilon YD}\, A_{ab}^{XY\tau\upsilon}, \quad X, Y = L, S.$$

The expression for two-electron matrix elements follows by extension. The scalar matrix elements may be real or complex, but the effect of time-reversal goes beyond simple conjugation because of the implicit spin-dependence and the component type. By applying $\hat{\mathcal{K}}$ to the above expression, the following results:

$$\hat{\mathcal{K}} A_{ab}^{XY\tau\upsilon} = (A_{ab}^{XY\tau\upsilon})^* = (-1)^{X+Y+\tau+\upsilon+1} A_{ab}^{XY-\tau-\upsilon}$$

The one-electron matrix elements from the Dirac operators are

$$V_{ab}^{XY\tau\upsilon} = \delta_{\tau\upsilon}\delta_{XY}\langle\, a^X \mid V^{nuc} \mid b^X \,\rangle,$$

$$S_{ab}^{XY\tau\upsilon} = \delta_{\tau\upsilon}\delta_{XY}\langle\, a^X \mid b^X \,\rangle,$$

$$\Pi_{ab}^{XY\tau\tau} = (1-\delta_{XY})\langle\, a^X \mid 2\tau\,\frac{\partial}{\partial z} \mid b^Y \,\rangle,$$

$$\Pi_{ab}^{XY\tau\upsilon} = (1-\delta_{XY})\langle\, a^X \mid \frac{\partial}{\partial x} + 2\upsilon\, i\frac{\partial}{\partial y} \mid b^Y \,\rangle.$$

Integrals in scalar basis

No comment need be made on the Coulomb integrals but the reduction of the magnetic integrals to the scalar basis is not as straightforward. Inserting the spin matrices, we may write for the Gaunt interaction

$$
\begin{aligned}
(\,\boldsymbol{\mu}^L\boldsymbol{\sigma}\boldsymbol{\nu}^S \mid \boldsymbol{\kappa}^L\boldsymbol{\sigma}\boldsymbol{\lambda}^S\,) = 2\big[(\,\mu^L_\alpha\nu^S_\beta \mid \kappa^L_\beta\lambda^S_\alpha\,) + (\,\mu^L_\beta\nu^S_\alpha \mid \kappa^L_\alpha\lambda^S_\beta\,)\big] \\
+ \sum_{\upsilon,\tau}(-1)^{\upsilon+\tau+1}(\,\mu^L_\upsilon\nu^S_\upsilon \mid \kappa^L_\tau\lambda^S_\tau\,)
\end{aligned}
$$

The remainder of the reduction follows from this. For the Breit interaction, the form of the integrals needs also to be considered, because of the structure of the operator. These will be considerably more complicated in the scalar basis.

For the SCF equations in the scalar basis, we also need the scalar density matrices,

$$
D^{XY\,\upsilon\tau}_{cd} = \sum_{CD}\sum_{\kappa\lambda} i^{\,D-C}\, T^{\upsilon XC}_{c\kappa}\, D^{XY}_{\kappa\lambda}\, T^{\tau YD}_{d\lambda}
$$

$$
D^{XY\,\upsilon\tau}_{\bar{c}\bar{d}} = \sum_{CD}\sum_{\kappa\lambda} i^{\,D-C}\, T^{\upsilon XC}_{c\bar{\kappa}}\, D^{XY}_{\bar{\kappa}\bar{\lambda}}\, T^{\tau YD}_{d\bar{\lambda}},
$$

$$
\mathcal{D}^{XY\,\tau\upsilon}_{cd} = D^{XY\,\tau\upsilon}_{cd} + D^{XY\,\tau\upsilon}_{\bar{c}\bar{d}}.
$$

These have the same relations under time-reversal as the matrix elements.

Scalar Fock matrix

The (Coulomb) Fock matrix elements in the scalar basis are

$$
\begin{aligned}
F_{ab}^{XX\tau\tau} &= V_{ab}^{XX\tau\tau} - 2c^2\delta_{XS}S_{ab}^{XX\tau\tau} \\
&\quad + \sum_{cd,Y} (\, a^X b^X \mid c^Y d^Y \,) \left[D_{cd}^{YY\alpha\alpha} + D_{cd}^{YY\beta\beta} \right] \\
&\quad - \sum_{cd} (\, a^X d^X \mid c^X b^X \,) \, D_{cd}^{XX\tau\tau} \\
F_{ab}^{XX\tau\upsilon} &= - \sum_{cd} (\, a^X d^X \mid c^X b^X \,) \, D_{cd}^{XX\upsilon\tau}. \\
F_{ab}^{LS\tau\tau} &= (2\tau)\, c\Pi_{ab}^{\tau\tau} - \sum_{cd} (\, a^L d^L \mid c^S b^S \,) \, D_{cd}^{SL\tau\tau} \\
F_{ab}^{LS\tau\upsilon} &= c\Pi_{ab}^{\tau\upsilon} - \sum_{cd} (\, a^L d^L \mid c^S b^S \,) \, D_{cd}^{SL\upsilon\tau}
\end{aligned}
$$

Addition of the magnetic integrals is a little more complicated, and does not lead to expressions which are as symmetric as are seen in the 2-spinor basis, but in general, though, the construction of the Fock matrix may be performed efficiently in either the scalar or the 2-spinor basis.

Scalar vs. 2-spinor basis.

The question of whether to work in the 2-spinor or in the scalar basis involves several considerations:

- The length of the scalar basis, and therefore the number of integrals in the scalar basis
- Whether to use an uncontracted or contracted basis, and what to do with the spin-orbit splitting in a contracted basis.
- The scheme used to implement symmetry — the Dacre/Elder scheme with a petite list of integrals, or the use of a symmetry-adapted basis.
- Whether to use a direct or a conventional SCF scheme.
- Whether to construct supermatrices or work from ordered integrals.

Scalar basis sets

The first of these considerations is the length of the scalar basis set. We have already seen that in a 2-spinor (primitive) basis, there are 64 times as many integrals as in a non-relativistic calculation. The large component scalar basis is clearly the same as the NR basis. It is the small component basis size which must be determined.

- Using the principle of kinetic balance to determine the small component basis, it can be seen that the small component basis must contain the *derivatives* of the large component basis.
- This means that there are roughly 2.5 times as many functions in the small component basis.
- The number of $(LL|SS)$ integrals is then 13 times the number of NR integrals.
- The number of $(SS|SS)$ integrals is then 39 times the number of NR integrals.
- The total increase for a Dirac-Coulomb calculation over a NR calculation is a factor of 53 in the scalar basis.

Thus the scalar basis involves fewer integrals than the 2-spinor basis, and therefore less work in the Fock matrix construction. This applies to an *uncontracted* basis set.

Scalar basis sets

In a contracted basis set, the core spin-orbit splitting must be accounted for, and the calculations become a little more difficult to do directly.

- The worst case for basis set size would be an ANO-type basis set, where all functions are contracted. Then the large component basis set would double in size, apart from the s functions.
- Allowing for a 40% increase in the basis size, the $(LL|LL)$ set would increase by a factor of 4 and the $(LL|SS)$ set by a factor of 2, giving $4 + 25 + 39 = 68$ as the overall factor, which is a little more than for the 2-spinor basis.

Thus, the scalar and 2-spinor integral sets come out approximately the same length in a contracted basis.

There is one further point to be made about the small component basis. It was assumed that the derivative functions would be stored separately. However, for the functions belonging to the $\kappa > 0$ spinors, it would be possible to combine the two constituents, which have $n = \ell + 1$ and $n = \ell + 3$, into a single function, and thereby reduce the size of the small component basis. This combination is called *minimal* or *restricted* kinetic balance. In some codes, the transformation to the spinor basis retains both constitutents as separate functions, giving *extended* or *unrestricted* kinetic balance.

Scalar basis sets

Before leaving the topic of basis sets, the question of whether nonrelativistic basis sets can be carried over to relativistic calculations should be addressed, in the context of the more general issue of basis set design.

A *balanced* basis set is one which as far as possible

- has the same number of primitive functions to represent each maximum in each atomic function, especially the outermost maximum, and
- the energy change on addition of a function in any angular symmetry is the same for all symmetries.

For a relativistic basis, one must decide whether the angular symmetry is the nonrelativistic symmetry or the relativistic symmetry.

If a nonrelativistic (primitive) basis set is used in a relativistic calculation *without* modification, for light elements it performs well but for heavy elements there are serious deficiencies in the core region. The main deficiencies are in the $p_{1/2}$ space, due to the s character of the small component. Addition of at least two extra functions is needed for the $6p$ block to reduce this deficiency to less than 0.5 E_{h}. These functions are not needed for the $p_{3/2}$. It has been found that a balanced basis set for Rn requires as many functions for the $p_{1/2}$ as for the s space — thus the core is in the relativistic extreme where the good quantum number is j, not ℓ.

Group theory

The use of symmetry is a powerful means of reducing the amount of work in a calculation. The symmetry operations — rotations, reflections and inversions —which transform a molecule into an identical structure form a group. The concepts of group theory are very briefly summarized here.

- A group consists of a set of elements which may be combined by an operation, denoted multiplication, $\otimes$.
- A group is closed under multiplication, i.e. each product of elements is also an element of the group.
- The elements obey the distributive and associative laws, i.e. $a(b+c) = ab + ac$ and $a(bc) = (ab)c$.
- The elements however do not obey the commutative law, $ab = ba$; if they do the group is called Abelian.
- Each group has an identity element, and each element has an inverse. This defines an abstract group. There are many realizations of each group, but an important one is in terms of a set of matrices.
- The matrix representation of the elements of a group may in general be reduced to block diagonal form. This is called the *irreducible representation*, or *irrep*.
- The trace of the irreducible matrix is called the *character* of the irrep.

Rotation of atomic functions

Since reflection is a combination of a rotation and an inversion, we consider first the rotation of atomic functions as a precursor to the consideration of molecular rotations and other operations. An atomic function of a particular angular momentum is transformed into a linear combination of functions of the same angular momentum by a finite rotation

$$\phi_{jm}(R) = \sum_n \mathcal{D}^j_{mn}(R)^* \ \phi_{jn}$$

where $R = (\alpha, \beta, \gamma)$ is the set of Euler angles defining the rotation. The *rotation matrices* $\mathcal{D}^j_{mn}(R)$ are given by

$$\mathcal{D}^j_{mn}(R) = e^{-i(\alpha m + \gamma n)} d^j_{mn}(\beta)$$

where the reduced matrices $d^j_{mn}(\beta)$ are real and are functions of $\cos\beta/2$ and $\sin\beta/2$. In particular,

$$d^j_{mn}(\pi) = (-1)^{j+m}\delta_{m,-n}; \qquad d^j_{mn}(2\pi) = (-1)^{2j}\delta_{m,n}.$$

Double groups

If the function has integer angular momentum, a rotation of 2π in any direction returns the function to its original state. If the function has half-integer angular momentum, a rotation of 2π in any direction returns the function to its original state but with a change of sign. Therefore,

- For bosons, the unit operation is a rotation by 2π.
- For fermions, the unit operation is not a rotation by 2π but a rotation by 4π.

The implications of this for molecular point groups are

- The operation of rotation by 2π, given the label $\bar{E}$, must be added to the group.
- This doubles the order of the group - hence the name *double groups*.
- The extra irreps in the group are the ones that describe fermions — the original irreps are for bosons. The irreps will therefore be called *boson* and *fermion* irreps.

Fermion irreps

Though $\bar{E}$ is *not* the time-reversal operator, there is a connection with time reversal in the double groups. The fermion functions must be doubly-degenerate, and each member of the Kramers pair must belong to one of the fermion irreps. We may divide the fermion irreps into three classes for the simpler groups, according to the distribution of the members of the Kramers pairs.

Real irreps

- A Kramers pair belongs to a doubly degenerate irrep. Spinor p belongs to one row and its time-reversed partner, $\bar{p}$, belongs to the other row.
- Spinor matrix elements are real, hence the name.
- One-electron matrices are block-diagonal in the Kramers components.

Complex irreps

- The members of a Kramers pair belong to different, singly degenerate irreps, which together form a doubly degenerate rep.
- Spinor matrix elements are complex, hence the name.
- One-electron matrices are block-diagonal in the Kramers components.

Note: This is *not* the usual group-theoretical designation, and a more proper name for the "real" irreps would be real-matrix irreps.

Fermion irreps

Quaternionic irreps

- The members of a Kramers pair belong to the same irrep which is singly degenerate.
- Spinor matrix elements are complex
- One-electron matrices are *not* block-diagonal in the Kramers components.

However, it is possible to make the one-electron matrix elements diagonal by a *quaternionic* transformation, provided that the operator is symmetric under time-reversal.

Quaternions form a Clifford algebra, which consists of the real unit (i.e. 1) and a series of units whose square is -1. A quaternionic number is given in terms of two complex numbers a and b by

$$q = a + bj = \operatorname{Re} a + i \operatorname{Im} a + j \operatorname{Re} b + k \operatorname{Im} b,$$

where

$$i^2 = j^2 = k^2 = -1, \quad ij = k, \quad jk = i, \quad ki = j.$$

The quaternion units anticommute. The quaternionic conjugate of q is

$$q^* = a^* - bj = \operatorname{Re} a - i \operatorname{Im} a - j \operatorname{Re} b - k \operatorname{Im} b.$$

Because the units do not commute, quaternionic numbers do not commute under multiplication: $q_1 q_2 \neq q_2 q_1$.

Fermion irreps

The matrix of an operator $\hat{\mathcal{F}}$ in a Kramers basis can be represented as

$$\mathbf{F} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ -t\mathbf{B}^* & t\mathbf{A}^* \end{pmatrix}$$

where $A_{pq} = F_{pq}$ and $B_{pq} = F_{p\bar{q}}$. The quaternionic transformation which diagonalizes this operator is

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{I} & j\mathbf{I} \\ j\mathbf{I} & \mathbf{I} \end{pmatrix}$$

giving the result

$$\mathbf{U}^\dagger \mathbf{F} \mathbf{U} = \begin{pmatrix} \mathbf{A} + \mathbf{B}j & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^* - \mathbf{B}^* j \end{pmatrix} = \begin{pmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & -k\mathbf{G}k \end{pmatrix}.$$

$\mathbf{G}$ is a quaternionic Hermitian matrix, and the eigenvalues of $\mathbf{F}$ are clearly doubly degenerate. Thus, we may restate the features of quaternionic irreps:

- The members of a Kramers pair belong to the same irrep which is singly degenerate.
- One-electron matrices may be block-diagonalized into new Kramers components by a quaternionic transformation.
- Spinor matrix elements are quaternionic.

More on quaternions

It may be thought that quaternions could be used in the field operators to define a new basis in which all matrices were block-diagonal in the Kramers pairs. However, because of the non-commutative algebra, the step in which the creation operator is permuted over the matrix element to separate the two does not produce the desired result. Therefore, quaternions will be used only at the matrix algebra stage, and not in the formalism.

An important realization of quaternions is to be found in the Pauli matrices. The set of matrices $\{\mathbf{I}_2, i\sigma_z, i\sigma_y, i\sigma_x\}$ is isomorphic to the set of quaternion units, $\{1, i, j, k\}$. This will be used later in the construction of symmetry spinors.

Fermion irreps

For the binary groups, D_{2h} and subgroups, whose operations are all two-fold, the extra irreps all fall into one of the three categories above.

- The groups D_{2h}, D_2 and C_{2v} whose fermion irreps are all real will be called *real* groups.
- The groups C_{2h}, C_2 and C_s whose fermion irreps are all complex will be called *complex* groups.
- The groups C_i and C_1 whose fermion irreps are all quaternionic will be called *quaternionic* groups.

Among the higher groups, many have fermion irreps which fall into these categories. C_{4v}, for example, has two real fermion irreps, C_{3v} has one real and one complex fermion irrep. Others have irreps which display higher degeneracies, and these fall into two categories:

- Genuine $2m$-fold degenerate irreps, where the members of Kramers pairs can be assigned to different rows. An example is T_d, which has two real irreps of degeneracy 2 and one of degeneracy 4.
- $2m$-fold degenerate irreps which occur in conjugate pairs, so that the pair together forms a rep. The members of a Kramers pair belong to the different irreps. An example is T, which has 6 fermion irreps, 2 doubly degenerate irreps which together form a rep, and 4 singly degenerate irreps which form 2 reps.

Group chains

In order to classify the irreps of the higher groups — and also to gain insight into the properties of the lower groups — it is useful to make use of a *group chain*.

- The higher group is decomposed into a product of two groups, one of which is of order 2.
- A correspondence is set up between the fermion irreps in the higher group and the lower group
- This is used to define the rows of the irreps in the higher group.
- It is particularly useful to make this definition so that real irreps reduce to complex irreps, where the rows and the multiplication table is well-defined.

A simple example of this is the decomposition $C_{2v} = C_2 \otimes C_s$. The reduction $C_{2v} \to C_2$ gives the correspondence $A_1, A_2 \to A$, $B_1, B_2 \to B$ for the boson irreps, and $('E,''E) \to (E, \bar{E})$ for the fermion irreps, where the primes indicate the rows of the irrep in C_{2v} and the bar indicates the conjugate irrep in C_2. A more extensive example is the reduction $T_d \to T$, which follows the same pattern.

Spin-orbitals and spinors

The spin functions α and β must transform according to the fermion irreps, and the spatial functions according to the boson irreps. Therefore we ought to be able to classify the real and imaginary, alpha and beta components of a spinor in terms of the boson irreps.

We consider the simplest case where we have a clear division of the spin functions, and that is the group C_2, for which the multiplication table is given, and take α in E and β in $\bar{E}$.

- Since $A \otimes E = E$ and $B \otimes \bar{E} = E$, a 2-spinor belonging to E must have functions of A symmetry for α spin and B symmetry for β spin.
- Similarly, since $A \otimes \bar{E} = \bar{E}$ and $B \otimes E = \bar{E}$, a 2-spinor belonging to $\bar{E}$ must have functions of B symmetry for α spin and A symmetry for β spin.

For a larger group such as C_{2v}, the group chain to C_2 gives some information, but the division between A_1 and A_2, and B_1 and B_2, needs further input. We therefore need a more general approach.

Multiplication table for C_2

Symmetry of spinor components

The symmetry of the small components is related to the symmetry of the large components by the operator

$$\hat{\mathcal{G}} = i\boldsymbol{\sigma} \cdot \Gamma_{\nabla} \equiv i\boldsymbol{\sigma} \cdot \Gamma_{\mathbf{r}} = \begin{pmatrix} [0, \Gamma_z] & [\Gamma_y, \Gamma_x] \\ [\Gamma_y, \Gamma_x] & [0, \Gamma_z] \end{pmatrix}$$

where $\Gamma_{\mathbf{r}} = (\Gamma_x, \Gamma_y, \Gamma_z)$, and the characters in parentheses are the characters of the real and imaginary parts. Applying this to a totally symmetric large-component alpha spin-orbital, we have

$$\hat{\mathcal{G}} \begin{pmatrix} [\Gamma_0, 0] \\ [0, 0] \end{pmatrix} = \begin{pmatrix} [0, \Gamma_z] \\ [\Gamma_y, \Gamma_x] \end{pmatrix}$$

This is not the most general form, but then neither is the large component. Applying the operator to this small component symmetry spinor, we should get the large component again:

$$\hat{\mathcal{G}} \begin{pmatrix} [0, \Gamma_z] \\ [\Gamma_y, \Gamma_x] \end{pmatrix} = \begin{pmatrix} [\Gamma_{\mathbf{r}}^2, \Gamma_x \otimes \Gamma_y] \\ [\Gamma_x \otimes \Gamma_z, \Gamma_y \otimes \Gamma_z] \end{pmatrix} = \begin{pmatrix} [\Gamma_0, \Gamma_{R_z}] \\ [\Gamma_{R_y}, \Gamma_{R_x}] \end{pmatrix}$$

This is now the most general form of a large component spinor in terms of symmetry types. The most general form for the small component results from another application of the operator:

$$\hat{\mathcal{G}} \begin{pmatrix} [\Gamma_0, \Gamma_{R_z}] \\ [\Gamma_{R_y}, \Gamma_{R_x}] \end{pmatrix} = \begin{pmatrix} [\Gamma_{xyz}, \Gamma_z] \\ [\Gamma_y, \Gamma_x] \end{pmatrix} = \Gamma_{xyz} \begin{pmatrix} [\Gamma_0, \Gamma_{R_z}] \\ [\Gamma_{R_y}, \Gamma_{R_x}] \end{pmatrix}$$

Symmetry of spinor components

The basic structure of the spinors is established, and to generate all possible spinors, these spinor symmetry vectors can be multiplied by any of the characters of the boson irreps of the group. However, it should also be noted that time-reversal symmetry will invert the vectors,

$$\hat{\mathcal{K}}\begin{pmatrix}[\Gamma_0,\Gamma_{R_z}]\\ [\Gamma_{R_y},\Gamma_{R_x}]\end{pmatrix}=\begin{pmatrix}[\Gamma_{R_y},\Gamma_{R_x}]\\ [\Gamma_0,\Gamma_{R_z}]\end{pmatrix},$$

and therefore it is not necessary to work through all boson irreps to obtain all fermion symmetry functions, but only those for operations which commute with the Dirac Hamiltonian (when presented in proper form). For D_{2h} and subgroups, it is only necessary to consider Γ_0 and Γ_{xyz}.

The basic spinor symmetry vectors for D_{2h} and subgroups are

$$C_{2v}\ :\ \begin{pmatrix}[A_1,A_2]\\ [B_1,B_2]\end{pmatrix}\qquad D_2\ :\ \begin{pmatrix}[A,B_1]\\ [B_2,B_3]\end{pmatrix}$$

$$C_2\ :\ \begin{pmatrix}A\\ B\end{pmatrix}\qquad C_s\ :\ \begin{pmatrix}A'\\ A''\end{pmatrix}$$

$$C_1\ :\ \begin{pmatrix}A\\ A\end{pmatrix}$$

from which the Kramers partner symmetries may be obtained using the above expression, and the symmetries for the groups with inversion by applying the character for the inversion operation.

Symmetries of 2-particle states

Having considered the symmetry of the spinors, we now consider the symmetries of 2-particle states. These must of course transform as the boson irreps. We expect that the product of a spinor and its Kramers partner should give the totally symmetric irrep, because it represents a closed shell. From the example of C_2, we see that indeed $E \otimes \bar{E} = A$, and also that $E \otimes E = \bar{E} \otimes \bar{E} = B$. Thus for the simple two-fold groups where the irreps are all singly degenerate, the symmetries of the 2-particle states are easily determined from the multiplication table.

To proceed further, we may take the direct product of the symmetry spinors to form symmetry bisinors. Thus for two spinors from the same row,

$$\begin{pmatrix} [\Gamma_0, \Gamma_{R_z}] \\ [\Gamma_{R_y}, \Gamma_{R_x}] \end{pmatrix} \otimes \begin{pmatrix} [\Gamma_0, \Gamma_{R_z}] \\ [\Gamma_{R_y}, \Gamma_{R_x}] \end{pmatrix} = \begin{pmatrix} [\Gamma_0, \Gamma_{R_z}] & [\Gamma_{R_y}, \Gamma_{R_x}] \\ [\Gamma_{R_y}, \Gamma_{R_x}] & [\Gamma_0, \Gamma_{R_z}] \end{pmatrix},$$

and for two spinors from different rows,

$$\begin{pmatrix} [\Gamma_0, \Gamma_{R_z}] \\ [\Gamma_{R_y}, \Gamma_{R_x}] \end{pmatrix} \otimes \begin{pmatrix} [\Gamma_{R_y}, \Gamma_{R_x}] \\ [\Gamma_0, \Gamma_{R_z}] \end{pmatrix} = \begin{pmatrix} [\Gamma_{R_y}, \Gamma_{R_x}] & [\Gamma_0, \Gamma_{R_z}] \\ [\Gamma_0, \Gamma_{R_z}] & [\Gamma_{R_y}, \Gamma_{R_x}] \end{pmatrix},$$

where the rows and columns of the bispinor are labelled with α and β.

Symmetries of 2-particle states

To obtain the bispinor symmetry we must also consider the products of spin functions, which transform like components of spherical tensors of rank 0 and 1:

$$\begin{pmatrix} \alpha\beta - \beta\alpha & \alpha\beta + \beta\alpha \\ \alpha\alpha - \beta\beta & \alpha\alpha + \beta\beta \end{pmatrix} \sim \begin{pmatrix} [\Gamma_0, 0] & [\Gamma_{R_z}, 0] \\ [0, \Gamma_{R_y}] & [\Gamma_{R_x}, 0] \end{pmatrix}$$

To form these spin functions, we take linear combinations of the bispinor and its time-reversed partner, which also have the same symmetry structure. Then we have four cases:

$$ee' + \bar{e}\bar{e}' \;:\; \begin{pmatrix} [\Gamma_{R_y}, 0] \otimes [\Gamma_0, 0] & [0, \Gamma_{R_x}] \otimes [\Gamma_{R_z}, 0] \\ [\Gamma_0, 0] \otimes [0, \Gamma_{R_y}] & [0, \Gamma_{R_z}] \otimes [\Gamma_{R_x}, 0] \end{pmatrix}$$

$$ee' - \bar{e}\bar{e}' \;:\; \begin{pmatrix} [0, \Gamma_{R_x}] \otimes [\Gamma_0, 0] & [\Gamma_{R_y}, 0] \otimes [\Gamma_{R_z}, 0] \\ [0, \Gamma_{R_z}] \otimes [0, \Gamma_{R_y}] & [\Gamma_0, 0] \otimes [\Gamma_{R_x}, 0] \end{pmatrix}$$

$$e\bar{e}' + \bar{e}e' \;:\; \begin{pmatrix} [0, \Gamma_{R_z}] \otimes [\Gamma_0, 0] & [\Gamma_0, 0] \otimes [\Gamma_{R_z}, 0] \\ [0, \Gamma_{R_x}] \otimes [0, \Gamma_{R_y}] & [\Gamma_{R_y}, 0] \otimes [\Gamma_{R_x}, 0] \end{pmatrix}$$

$$e\bar{e}' - \bar{e}e' \;:\; \begin{pmatrix} [\Gamma_0, 0] \otimes [\Gamma_0, 0] & [0, \Gamma_{R_z}] \otimes [\Gamma_{R_z}, 0] \\ [\Gamma_{R_y}, 0] \otimes [0, \Gamma_{R_y}] & [0, \Gamma_{R_x}] \otimes [\Gamma_{R_x}, 0] \end{pmatrix}$$

Symmetries of 2-particle states

The four functions clearly transform as

$$\begin{aligned} ee' + \bar{e}\bar{e}' &\sim \Gamma_{R_y} \\ i(ee' - \bar{e}\bar{e}') &\sim \Gamma_{R_x} \\ i(e\bar{e}' + \bar{e}e') &\sim \Gamma_{R_z} \\ e\bar{e}' - \bar{e}e' &\sim \Gamma_0. \end{aligned}$$

They are in fact examples of the real N-particle basis introduced earlier. To obtain other symmetries, where there are other symmetry operators which commute with the Dirac Hamiltonian, this result is simply multiplied by the character factors in front of the symmetry spinors.

The spin functions, which are symmetric under time-reversal, are

$$\begin{aligned} S_0 &= \alpha\beta - \beta\alpha \\ S_{1z} &= i(\alpha\beta + \beta\alpha) \\ S_{1y} &= i(\alpha\alpha - \beta\beta) \\ S_{1x} &= \alpha\alpha + \beta\beta \end{aligned}$$

Matrix elements and symmetry

Having discussed the symmetry of the spinors and products of spinors, we are now in a position to discuss the symmetries of the one- and two-electron integrals. In nonrelativistic programs, the MO integrals are often divided into symmetry classes for convenient handling of symmetry. We restrict the discussion here to the binary groups.

The nonrelativistic case is relatively straighforward because the integrals are real. The one-electron integrals f_{pq} must have both p and q belonging to the same irrep because the integral itself must belong to the totally symmetric irrep. The two electron integrals fall into one of four symmetry classes, $(II|II)$, $(II|JJ)$, $(IJ|JI)$ and $(IJ|KL)$. In the last class, not all combinations will give the totally symmetric irrep, so this class is restricted to only those combinations which do. Note that the product of boson irreps II belongs to the totally symmetric irrep, and that the product IJ must belong to another irrep.

One-electron matrix elements

The KR relativistic case is a little more complicated. First, note that if the ket in an integral transforms as one row of the (ir)rep, the bra must transform as the opposite row. Thus for the one-electron integrals, p and q must belong to the same row of the same irrep. (A reference to the C_2 multiplication table will establish this point.) The symmetry reduction for the one-electron integrals results from the following considerations

- For the quaternionic groups, both f_{pq} and $f_{p\bar{q}}$ are non-zero and complex.
- For the complex groups, the second of these is zero, giving a reduction of a factor of 2.
- For the real groups, the first matrix element is also real, giving another factor of 2.
- The remaining symmetry reductions arise from the remaining symmetry operations, such as inversion, which provide good fermion quantum numbers.

The same arguments may be applied to the two-electron integrals. Thus the implementation of time-reversal symmetry on the integrals is equivalent to the implementation of some elements of boson symmetry.

Two-electron matrix elements

For the two-electron integrals it is useful again to consider the groups separately.

For the quaternionic groups, the only symmetry operations which give a reduction may be treated in exactly the same way as for the NR case.

For the complex groups,

- The products of fermion irreps e^*e and $\bar{e}^*\bar{e}$ belong to the totally symmetric irrep, and the products $e^*\bar{e}$ and $\bar{e}^*e$ belong to another irrep.
- Therefore the integrals $(pq|rs)$ belong to the class $(II|II)$. The integrals $(pq|\bar{r}\bar{s})$ belong to the class $(II|JJ)$, but since they are equivalent to the $(pq|rs)$ integrals, the distinction is not important.
- The integrals $(p\bar{q}|\bar{r}s)$ and $(p\bar{q}|r\bar{s})$ both belong to the class $(IJ|IJ)$ if p and r belong to the same irrep, otherwise they belong to the class $(IJ|KL)$.

Thus the classification according to time-reversal symmetry into $(II|II)$ and $(I\bar{I}|I\bar{I})$ is equivalent to a symmetry classification.

For the real groups, it is necessary to take linear combinations of the products of fermion irreps to form the boson irreps, but in the end the conclusion is the same as for the complex groups.

LCAS diatomic molecular spinors

Having considered the general aspects of symmetry, we now turn to some specific examples which give insight into relativistic effects on bonding. First, we consider the simple linear combination of atomic spinors (LCAS) in diatomic molecules.

The simplest combinations are those between spinors with $j = m_j$. Half-integral subscripts here refer to m_j, not j, and the bar will represent the $j = \ell - 1/2$ spin-orbit component.

$$s_{1/2} \pm s_{1/2} = \begin{pmatrix} s \\ \\ 0 \end{pmatrix} \pm \begin{pmatrix} s \\ \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} s\sigma \\ \\ 0 \end{pmatrix}, \begin{pmatrix} s\sigma^* \\ \\ 0 \end{pmatrix}$$

$$p_{3/2} \pm p_{3/2} = \begin{pmatrix} p_1 \\ \\ 0 \end{pmatrix} \pm \begin{pmatrix} p_1 \\ \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} p\pi^* \\ \\ 0 \end{pmatrix}, \begin{pmatrix} p\pi \\ \\ 0 \end{pmatrix}$$

$$d_{5/2} \pm d_{5/2} = \begin{pmatrix} d_2 \\ \\ 0 \end{pmatrix} \pm \begin{pmatrix} d_2 \\ \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} d\delta \\ \\ 0 \end{pmatrix}, \begin{pmatrix} d\delta^* \\ \\ 0 \end{pmatrix}$$

and so on. These therefore give the pure bonding and anti-bonding diatomic MSs.

LCAS diatomic molecular spinors

The next example is the set of spinors with $\omega \equiv m_j = \frac{1}{2}$ for the p shell.

$$\bar{e}^{+}_{1/2} = \bar{p}_{1/2} + \bar{p}_{1/2} = \begin{pmatrix} -p_0 \\ \\ \sqrt{2}p_1 \end{pmatrix} + \begin{pmatrix} -p_0 \\ \\ \sqrt{2}p_1 \end{pmatrix} \rightarrow \begin{pmatrix} -p\sigma^* \\ \\ \sqrt{2}p\pi \end{pmatrix}$$

$$e^{+}_{1/2} = p_{1/2} + p_{1/2} = \begin{pmatrix} \sqrt{2}p_0 \\ \\ p_1 \end{pmatrix} + \begin{pmatrix} \sqrt{2}p_0 \\ \\ p_1 \end{pmatrix} \rightarrow \begin{pmatrix} \sqrt{2}p\sigma^* \\ \\ p\pi \end{pmatrix}$$

Since these two have the same ω, we may take a linear combination of them, e.g. to get pure spin spinors:

$$\sqrt{2}\begin{pmatrix} -p\sigma^* \\ \\ \sqrt{2}p\pi \end{pmatrix} + \begin{pmatrix} \sqrt{2}p\sigma^* \\ \\ p\pi \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ \\ p\pi \end{pmatrix}$$

$$-\begin{pmatrix} -p\sigma^* \\ \\ \sqrt{2}p\pi \end{pmatrix} + \sqrt{2}\begin{pmatrix} \sqrt{2}p\sigma^* \\ \\ p\pi \end{pmatrix} \rightarrow \begin{pmatrix} p\sigma^* \\ \\ 0 \end{pmatrix}$$

The negative combination gives the same results with bonding and antibonding character reversed. Note that for homonuclear diatomics, these are the only allowed combinations, giving spinors of g and u symmetry: $e^{+}_{1/2} \equiv e_{1/2g}$, $e^{-}_{1/2} \equiv e_{1/2u}$.

LCAS diatomic molecular spinors

For heteronuclear diatomics, we may take the combinations

$$\bar{p}_{1/2} + p_{1/2} = \begin{pmatrix} -p_0 \\ \\ \sqrt{2}p_1 \end{pmatrix} + \begin{pmatrix} \sqrt{2}p_0 \\ \\ p_1 \end{pmatrix} \rightarrow \begin{pmatrix} -p\sigma \\ \\ p\pi \end{pmatrix}$$

This spinor is not symmetric: the σ part is skewed towards one atom, and the π towards the other. Both components are *bonding*, in contrast to the previous cases. The combination with the functions on the opposite centres gives

$$p_{1/2} + \bar{p}_{1/2} = \begin{pmatrix} \sqrt{2}p_0 \\ \\ p_1 \end{pmatrix} + \begin{pmatrix} -p_0 \\ \\ \sqrt{2}p_1 \end{pmatrix} \rightarrow \begin{pmatrix} p\sigma \\ \\ p\pi \end{pmatrix}$$

Note the change of sign for the $m_s = 1/2$ component. Linear combinations of these will return pure-spin spinors as before.

Combinations such as these will be favoured for the early p-block heavy elements such as Tl and Pb where the spin-orbit splitting is large and there is an energetic cost to promote from $\bar{p}$ to p.

LCAS diatomic molecular spinors

The next example is of s-p bonding. The spinors, which have $\omega = 1/2$, are

$$s_{1/2} + \bar{p}_{1/2} = \begin{pmatrix} s \\ \\ 0 \end{pmatrix} + \begin{pmatrix} -p_0 \\ \\ \sqrt{2}p_1 \end{pmatrix} \rightarrow \begin{pmatrix} sp\sigma \\ \\ n_p \end{pmatrix}$$

$$s_{1/2} - p_{1/2} = \begin{pmatrix} s \\ \\ 0 \end{pmatrix} - \begin{pmatrix} \sqrt{2}p_0 \\ \\ p_1 \end{pmatrix} \rightarrow \begin{pmatrix} sp\sigma \\ \\ -n_p \end{pmatrix}$$

We may also take linear combinations of these two. Here, lone-pair character is introduced; the amount depends on the amount of $\bar{p}_{1/2}$ and $p_{1/2}$ mixing in the MS.

LCAS diatomic molecular spinors

We may summarize the effects of relativity on bonding for s and p shells as follows:

- In the NR limit, pure spin and pure bonding type are returned.
- $s\sigma$ bonds retain pure spin type. They are *stronger* and *shorter* due to the s-orbital contraction effect.
- $p\pi^*$ character is introduced into $p\sigma$ bonds (and $p\sigma^*$ into $p\pi$), making for *weaker* bonds. This is especially true in the homonuclear diatomics.
- The degree of mixing of opposite bonding character is a function of the spin-orbit splitting of the p shell.
- Strong bonds are retained where a $\bar{p}_{1/2}$ spinor can combine with a $p_{1/2}$ spinor: here $p\sigma$ and $p\pi$ character are mixed. This can be the case for heteronuclear diatomics.
- $sp\sigma$ bonds are slightly weaker due to mixing of lone-pair character from the atom contributing the p orbital.

LCAS diatomic molecular spinors

As an example of the p-block bonding, we examine the dissociation energies in eV for homonuclear group 13 diatomics and the heteronuclear group 13 – group 15 diatomics.

Dimer	D_0	$D_0 \times 1.5$	Dimer	D_0
Al_2	1.55	2.32	AlP	2.20
Ga_2	1.40	2.10	GaAs	2.18
In_2	1.01	1.51	InSb	1.54
Tl_2	$\sim$0.1	$\sim$0.15	TlBi	1.21

In the relativistic limit:

- Group 13 has a single $\bar{p}$ electron outside a closed s shell.
- Group 15 has a single p electron outside a closed $\bar{p}$ shell.
- 13-13 bonds approach the weaker $\bar{p} - \bar{p}$ bonding limit.
- 13-15 bonds approach the stronger $\bar{p} - p$ bonding limit.

LCAS diatomic molecular spinors

Extending these ideas to higher angular momenta, we find that combining p- and d-type spinors with opposite sign of κ gives totally bonding or antibonding results:

$$p_{1/2} - \bar{d}_{1/2} = \begin{pmatrix} \sqrt{2}p_0 \\ \\ p_1 \end{pmatrix} - \begin{pmatrix} -\sqrt{2}d_0 \\ \\ \sqrt{3}d_1 \end{pmatrix} \rightarrow \begin{pmatrix} pd\sigma \\ \\ pd\pi \end{pmatrix}$$

$$d_{1/2} - \bar{p}_{1/2} = \begin{pmatrix} \sqrt{3}d_0 \\ \\ \sqrt{2}d_1 \end{pmatrix} - \begin{pmatrix} -p_0 \\ \\ \sqrt{2}p_1 \end{pmatrix} \rightarrow \begin{pmatrix} pd\sigma \\ \\ -pd\pi \end{pmatrix}$$

while combining p- and d-type spinors with the same sign of κ gives mixed bonding and antibonding character:

$$p_{1/2} + d_{1/2} = \begin{pmatrix} \sqrt{2}p_0 \\ \\ p_1 \end{pmatrix} + \begin{pmatrix} \sqrt{3}d_0 \\ \\ \sqrt{2}d_1 \end{pmatrix} \rightarrow \begin{pmatrix} pd\sigma \\ \\ pd\pi^* \end{pmatrix}$$

$$\bar{p}_{1/2} + \bar{d}_{1/2} = \begin{pmatrix} -\sqrt{2}d_0 \\ \\ \sqrt{3}d_1 \end{pmatrix} + \begin{pmatrix} -p_0 \\ \\ \sqrt{2}p_1 \end{pmatrix} \rightarrow \begin{pmatrix} -pd\sigma \\ \\ pd\pi^* \end{pmatrix}$$

LCAS diatomic molecular spinors

In summary, the effects of relativity on bonding in diatomics are:

- Normal strong bonds where contributing spinors are of *opposite* sign of κ *or* are of pure spin type.
- Weaker bonds where contributing spinors are of the *same* sign of κ *or* where a pure-spin and mixed-spin spinor combine.
- Mixing of different simple LCAS spinors occurs due to (double group) symmetry.
- The degree of mixing is a function of the spin-orbit splitting of the shells involved.

Extension to linear polyatomics shows that the strong bonding combinations occur for alternating signs of κ as one moves along the molecule.

Examples of relativistic bonding

Two good examples of relativistic effect on bonding are PbO and PbO_2. The valence atomic orbitals are $6s$ and $6p$ on Pb, $2p$ on O. Mulliken populations are given on the next page.

- PbO has the valence electron configuration $\sigma^2\sigma^2\pi^4$ or $e^2_{1/2}e^2_{1/2}e^2_{1/2}e^2_{3/2}$.
- The lowest σ orbital is Pb $6s$ – O $2p$. Since the spin-orbit splitting on O is not large, this is a strong bonding orbital.
- The highest two $e_{1/2}$ spinors display $\bar{p}_{1/2} + p_{1/2}$ bonding. The spin purity is 56.8% β spin for $25e_{1/2}$, 53.8% α spin for $26e_{1/2}$.

- PbO_2 has the valence electron configuration $\sigma^2_g\pi^4_u\sigma^2_u\pi^4_g$ or $e^2_{1/2g}e^2_{1/2u}e^2_{3/2u}e^2_{3/2g}e^2_{1/2g}e^2_{1/2u}$.
- $e_{1/2g}$ spinors retain spin purity. $e_{1/2u}$ spinors mix, due to proximity; these are mostly O spinors.
- The spinors are not as close to totally bonding as for PbO, but the $14e_{1/2u}$ shows some tendency to the strong bonding combination with p on O and $\bar{p}$ on Pb.
- The spin purity is 96.5% β spin for $13e_{1/2u}$, 94.9% α spin for $14e_{1/2u}$.

Example of p-block bonding.

Mulliken gross populations of PbO and PbO_2 at r_e.

		Pb		O	
		$\bar{p}_{1/2}$	$p_{1/2}$	$\bar{p}_{1/2}$	$p_{1/2}$
PbO					
	$25e_{1/2}$	0.00	0.25	1.45	0.01
	$26e_{1/2}$	0.72	0.03	0.04	1.04
PbO_2					
	$13e_{1/2u}$	0.13	0.11	1.46	0.30
	$14e_{1/2u}$	0.34	0.13	0.30	1.13

Correlation methods

It is apparent from a moment's consideration that, when it comes to electron correlation, there is in principle no reason why nonrelativistic methods based on spin-orbitals should not be transferred directly to the relativistic case. The Hamiltonian expressed in terms of spinors (not Kramers pairs) has an identical representation in second quantization. Provided the following two considerations are taken into account, the methods are transferable.

- The relativistic Hamiltonian is projected onto the positive energy space. That is, we work within a *no-pair* approximation.
- Account is taken of the fact that the integrals are complex and the permutational symmetry is thereby reduced.

For states that are well-described by a single closed-shell determinant, the usual single-reference methods may be employed. We consider the MP2 method and the CCSD method, and give some notes on the Kramers-restricted theory.

Kramers-retricted MP2

The expression for the second-order energy in spinor form is

$$E^{(2)} = \frac{1}{4} \sum_{I,J}^{2n_o} \sum_{A,B}^{2n_v} \frac{\left|\langle IJ||AB\rangle\right|^2}{\epsilon_I + \epsilon_J - \epsilon_A - \epsilon_B}$$

where the ϵ are eigenvalues of the Fock matrix, and $\langle IJ||AB\rangle$ are the antisymmetrized two-electron integrals in Dirac notation, expressed in terms of integrals in Mulliken notation as

$$\langle IJ||AB\rangle = (IA|JB) - (IB|JA).$$

Upper case is used for general spinor indices, and lower case will be used for Kramers pair indices. Occupied indices will be selected from I, J, K, etc. and virtual indices from A, B, C, etc. The Kramers-restricted zeroth-order Hamiltonian is

$$\hat{\mathcal{H}}_0 = \sum_{p=1}^{n} \hat{n}_p \epsilon_p,$$

where $n = n_o + n_v$ is the number of Kramers pairs, and the number operator is

$$\hat{n}_p = {a_p}^\dagger a_p + {a_{\bar{p}}}^\dagger a_{\bar{p}}.$$

Kramers-retricted MP2

The implementation of Kramers restriction in the second-order energy follows directly from the expansion of the integrals. Some care must be taken with this because of the antisymmetry:

$$
\begin{aligned}
\langle ij||ab\rangle &= (ia|jb) - (ib|ja) \\
\langle i\bar{j}||a\bar{b}\rangle &= (ia|\bar{j}\bar{b}) - (i\bar{b}|\bar{j}a) \\
\langle \bar{i}j||\bar{a}b\rangle &= (\bar{i}\bar{a}|jb) - (\bar{i}b|j\bar{a}) = \langle a\bar{b}||i\bar{j}\rangle \\
\langle \bar{i}\bar{j}||\bar{a}\bar{b}\rangle &= (\bar{i}\bar{a}|\bar{j}\bar{b}) - (\bar{i}\bar{b}|\bar{j}\bar{a}) = \langle ab||ij\rangle
\end{aligned}
$$

The Kramers-restricted expression is

$$
\begin{aligned}
E^{(2)} = \frac{1}{2} \sum_{i,j}^{n_o} \sum_{a,b}^{n_v} & \frac{\left|\langle ij||ab\rangle\right|^2 + 2\left|\langle i\bar{j}||a\bar{b}\rangle\right|^2 + \left|\langle ij||\bar{a}\bar{b}\rangle\right|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \\
+ \sum_{i,j}^{n_o} \sum_{a,b}^{n_v} & \frac{\left|\langle i\bar{j}||ab\rangle\right|^2 + \left|\langle ij||a\bar{b}\rangle\right|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}
\end{aligned}
$$

This is not quite a reduction of a factor of 4 which would be expected on the basis of the reduction in the number of integrals. There are however some connections between the integrals which might be exploited, but the integrals would have to be broken down into pieces which display the full permutational symmetry.

Kramers-retricted CCSD

The wave function in the coupled-cluster method is derived by operating with a wave operator in exponential form on the reference function,

$$\Psi = e^{\hat{T}}\,\Psi_0$$

where the operator $\hat{T}$ is a sum of products of excitation operators and amplitudes,

$$\begin{aligned}\hat{T} &= \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots \\ &= T_I^A\,\hat{E}_I^A + T_{IJ}^{AB}\,\hat{E}_{IJ}^{AB} + T_{IJK}^{ABC}\,\hat{E}_{IJK}^{ABC} +\end{aligned}$$

The Einstein summation convention has been employed here. The excitation operators are given by

$$\hat{E}_I^A = a_A{}^\dagger\,a_I\,; \qquad \hat{E}_{IJ}^{AB} = a_A{}^\dagger\,a_B{}^\dagger\,a_J\,a_I\,; \qquad \ldots\ .$$

The CCSD method truncates the expansion of $\hat{T}$ after the double excitation term. The t_2 amplitudes obey the relations

$$T_{IJ}^{AB} = -T_{JI}^{AB} = -T_{IJ}^{BA} = T_{JI}^{BA},$$

i.e. are antisymmetric to index interchange.

Kramers-retricted CCSD

Since the reference is a closed-shell determinant, it is symmetric under time-reversal. We require that the CC wave function is also symmetric under time-reversal, i.e.

$$\hat{\mathcal{K}}\,\Psi = \hat{\mathcal{K}}\,e^{\hat{T}}\,\Psi_0 = \Psi = e^{\hat{T}}\,\hat{\mathcal{K}}\,\Psi_0.$$

Thus $\hat{\mathcal{K}}$ commutes with the wave operator, which must be symmetric under time-reversal. Application to the $\hat{T}_1$ term gives

$$\begin{aligned}\hat{\mathcal{K}}\,\hat{T}_1 &= \hat{\mathcal{K}}\left\{t_i^a\,\hat{e}_i^a + t_{\bar{\imath}}^a\,\hat{e}_{\bar{\imath}}^a + t_i^{\bar{a}}\,\hat{e}_i^{\bar{a}} + t_{\bar{\imath}}^{\bar{a}}\,\hat{e}_{\bar{\imath}}^{\bar{a}}\right\}\\ &= t_i^{a*}\,\hat{e}_{\bar{\imath}}^{\bar{a}} - t_{\bar{\imath}}^{a*}\,\hat{e}_i^{\bar{a}} - t_i^{\bar{a}*}\,\hat{e}_{\bar{\imath}}^{a} + t_{\bar{\imath}}^{\bar{a}*}\,\hat{e}_i^{a}\end{aligned}$$

which gives the relations

$$t_i^a = t_{\bar{\imath}}^{\bar{a}*}; \qquad t_{\bar{\imath}}^a = -t_i^{\bar{a}*}.$$

These are the normal relations expected from time-reversal of a one-electron operator. Similarly, for the two-electron amplitudes,

$$t_{ij}^{ab} = t_{\bar{\imath}\bar{\jmath}}^{\bar{a}\bar{b}*}; \qquad t_{ij}^{\bar{a}\bar{b}} = t_{\bar{\imath}\bar{\jmath}}^{ab*}; \qquad t_{i\bar{\jmath}}^{a\bar{b}} = t_{\bar{\imath}j}^{\bar{a}b*}; \qquad t_{ij}^{a\bar{b}} = -t_{\bar{\imath}\bar{\jmath}}^{\bar{a}b*}.$$

The $\hat{T}$ operator can therefore be expressed in terms of excitation operators, but here the E operators must be used, not the X operators. For example,

$$\hat{T}_1 = \mathrm{Re}\,t_i^a\;E_{ai}^{+} + \mathrm{Im}\,t_i^a\;E_{ai}^{-} + \mathrm{Re}\,t_i^{\bar{a}}\;E_{\bar{a}i}^{+} + \mathrm{Im}\,t_i^{\bar{a}}\;E_{\bar{a}i}^{-}$$

Kramers-retricted CCSD

An assessment of the number of operations required to solve the CCSD equations will show the relative efficiency of KR-CCSD and NR-CCSD.

- The fact that the integrals are complex in the relativistic version gives a factor of 4 increase in the number of operations.
- The scaling of the CCSD equations is $\mathcal{O}(n^6)$. Thus the reduction due to spin symmetry should be a factor of $2^6 = 64$.
- The reduction due the the optimal implementation of time-reversal symmetry is a factor of 8.
- The KR-CCSD method is therefore a factor of 32 times more expensive than the NR-CCSD method.

This should be compared with the SCF stage of the calculation in the spinor basis.

- There are 32 times more KR Coulomb integrals than NR integrals.
- Complex arithmetic gives an extra factor of 4.
- A KR-SCF calculation is therefore 128 times more expensive than a NR-SCF calculation.

Integral transformation

We saw that the efficiency of the KR-CCSD correlation method relative to the NR method was greater than that of the KR-SCF method. The SCF scales as $\mathcal{O}(n^4)$; the CCSD as $\mathcal{O}(n^6)$. The integral transformation, which scales as $\mathcal{O}(n^5)$, also needs to be considered, and for that purpose we compare the NR and 2-spinor transformations.

- The integral transformation is performed with one triangular pair index and one square pair index to exploit the efficiency of matrix multiplications. This is possible for both NR and KR integrals.
- The NR operation count (multiply and add) is $2n^5$.
- The relativistic transformation must include integral classes $(LL|LL)$, $(LL|SS)$, $(SS|LL)$ and $(SS|SS)$. Therefore for the $(pq|rs)$ integrals one would expect a factor of 4, giving $8n^5$.
- However, the half-transformed $(LL|LL)$ and $(SS|LL)$ may be added together, as may the $(LL|SS)$ and $(SS|SS)$. There are therefore only 2 second half transformation steps, and the operation count is $6n^5$.
- There are 4 classes of MS integrals, $(pq|rs)$, $(p\bar{q}|\bar{r}s)$, $(p\bar{q}|r\bar{s})$ and $(pq|r\bar{s})$, each of which scales the same.
- The integrals are complex, introducing another factor of 4.
- The KR integral transformation is therefore 96 times more expensive than the NR integral transformation.

Multi-reference expansions

In nonrelativistic theory, there is a wide range of molecules which may be treated with a single-determinant reference: closed-shell molecules and high-spin open-shell molecules. In relativistic theory, this is no longer the case. Only the cases of a closed shell and a single electron outside a closed shell can be represented as a single determinant. As seen from the last section, for two electrons outside a closed shell, the minimum number of determinants in the reference is two. Therefore it is necessary to resort to multi-reference wave functions, more frequently than in nonrelativistic theory. Here we develop some of the ideas necessary for a multi-reference CI expansion.

We start with the determinant basis introduced earlier given in terms of A and B strings, and consider all possible determinants which may be constructed from a given set of Kramers pairs. We define subsets of determinants with a given value of N_α and N_β, defining

$$M_K = \tfrac{1}{2}(N_A - N_B)$$

- In the NR case, the CI is limited to a given $M_K \equiv M_S$ value due to spin factorization. The length of a full CI expansion is $\binom{n}{N_\alpha}\binom{n}{N_\beta}$ where n is the number of active orbitals.
- In the KR case, the CI includes all possible M_K values because of the spin-orbit interaction. The length of a full CI is $\binom{2n}{N}$.

Length of N-particle basis

For an example of the relative lengths of the N-particle basis for a full CI, the following table for 12 electrons in 12 orbitals is instructive. Symmetry is not used. The second column is the number of determinants in the NR expansion. The number of determinants in the R expansion is the same for each M_S at 2704156.

M_S	Determinants	Ratio R/NR
0	853776	3
1	627264	4
2	245025	11
3	48400	55
4	4356	629
5	144	18778
6	1	2704156

Thus for low spin, the relativistic full CI vector is only 3 times the length of the nonrelativistic vector. The work increases by a factor of 36 — 9 from the length of the vector and 4 from the complex arithmetic.

Hamiltonian matrix structure

We divide the determinants into subsets defined by the M_K value, and order the subsets by their M_K values. Then the Hamiltonian matrix is partitioned into blocks defined by these subsets. Because the electron-electron interaction is a 2-particle interaction, H is block pentadiagonal. This structure is shown in the diagrams for an even and an odd number of electrons.

The composition of the matrix elements on each diagonal is as follows.

- The main diagonal has $|M_K^\mu - M_K^\nu| = 0$. Matrix elements are composed of integral classes h_{tu}, $(tu|vw)$ and $(t\bar{u}|\bar{v}w)$
- The first off-diagonal has $|M_K^\mu - M_K^\nu| = 1$. Matrix elements are composed of integral classes $h_{t\bar{u}}$ and $(tu|v\bar{w})$.
- The second off-diagonal has $|M_K^\mu - M_K^\nu| = 2$. Matrix elements are composed of integral class $(t\bar{u}|v\bar{w})$

The classes of two-electron integrals required for the construction of the matrix elements are different for each diagonal, a feature which may be exploited in the construction of the sigma vector in a direct CI scheme.

Hamiltonian matrix structure

Diagram of block structure for N even. The blocks are ordered by M_K.

Hamiltonian matrix structure

Diagram of block structure for N odd. The blocks are ordered by M_K.

Time-reversal and symmetry

The integrals on the first off-diagonal are only non-zero for the quaternionic groups. These are represented by the grey shading in the diagrams. Then for the real and complex groups, the diagonal and second off-diagonal blocks partition into two disjoint, interleaving groups, represented by the striped white and grey shading. The determinants partition into two sets as a consequence.

- For N odd, time-reversal of one set productes the other set. The two sets of determinants therefore form a basis for the doubly-degenerate Kramers pairs of the N-particle states. The only other element of symmetry in the binary groups is inversion, which is implemented in the same way as in NR CI theory.
- For N even, the application of time-reversal maps each set onto itself. The two sets form bases for states of different boson symmetry. The set with M_K even contains the basis for the totally symmetric irrep. For the complex binary groups this is the only element of boson symmetry, except for the inversion symmetry in C_{2h}. Thus selection of the parity of M_K determines the state symmetry.

Note that time-reversal creates a symmetry about both diagonals of $\mathbf{H}$, and only 1/4 of $\mathbf{H}$ is unique, and 1/2 the CI coefficients, since $c_{\bar{P}} = c_P^*$, where P is the index of a determinant. Exploiting this feature reduces the work by a factor of 4.

Time-reversal and symmetry

For the real groups, the remaining element of symmetry is introduced by forming a real basis. From the consideration of the Hamiltonian matrix in the real basis, it may be seen that the elements of $\mathbf{H}$ in the real basis which connect the positive and negative combinations come from the imaginary part of the elements in the determinant basis, which must be zero. Thus the Hamiltonian blocks into 4, defined by the parity of M_K and the sign of the combination of determinants in the real basis. The symmetry of each set of CSF may be deduced from the discussion of the symmetry of 2-particle states above.

The blocks of $\mathbf{H}$ for which both H_{PQ} and $H_{P\bar{Q}}$ are non-zero occur only for $|M_K^P| + |M_K^Q| \leq 2$:

- For odd M_K the implementation of symmetry involves only the phase of the block with $M_K^P = 1$ and $M_K^Q = -1$.
- For even M_K, the triangularity for $M_K^P = 0$ and $M_K^Q = 0$ must be accounted for with its associated phase and the relative phases for $M_K^Q = \pm 2$ with $M_K^P = 0$ in the implementation of symmetry.

The only remaining case to consider is that of the quaternionic groups with N odd. The Hamiltonian matrix may be blocked by the quaternionic transformation given above.

$$\begin{pmatrix} \Phi^{'} \\ \Phi^{''} \end{pmatrix} = \begin{pmatrix} \Phi \\ \bar{\Phi} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & j \\ j & 1 \end{pmatrix}.$$

producing two quaternionic matrices of half the rank, with matrix elements

$$\begin{aligned} H_{P'Q'} &= H_{PQ} - H_{P\bar{Q}} j \\ H_{P''Q''} &= H_{\bar{P}\bar{Q}} - H_{\bar{P}Q} j \\ H_{P'Q''} &= H_{P''Q'} = 0 \end{aligned}$$

As for the real groups for N even, the blocks for which both H_{PQ} and $H_{P\bar{Q}}$ are non-zero occur only for $|M_K^P| + |M_K^Q| \leq 2$. Although the amount of work accumulating the sigma vector would be the same without the transformation, the indexing work is halved, and therefore it may be useful to transform.

Other approaches

It is clear from the discusssion so far that 4-component calculations are expensive, due to two factors: the small component of the wave function and the spin-dependence of the Hamiltonian. Therefore there has been much work put into methods which seek to eliminate one or other, or both, of these factors. We will examine both, starting with the standard elimination of the small component (ESC), and proceeding to other methods.

The Dirac equation with the small component eliminated is

$$(V-E)\Psi^L + \frac{1}{2m}\,(\boldsymbol{\sigma}\cdot\mathbf{p})\left[1-\frac{V-E}{2mc^2}\right]^{-1}(\boldsymbol{\sigma}\cdot\mathbf{p})\Psi^L = 0.$$

The spin-dependence in this equation may be separated by use of the Dirac relation

$$(\boldsymbol{\sigma}\cdot\mathbf{u})(\boldsymbol{\sigma}\cdot\mathbf{v}) = \mathbf{u}\cdot\mathbf{v} + i(\boldsymbol{\sigma}\cdot\mathbf{u}\times\mathbf{v}).$$

giving the equation

$$(V-E)\Psi^L + \frac{1}{2m}\,[\mathbf{p}R\cdot\mathbf{p} + i\boldsymbol{\sigma}\cdot\mathbf{p}R\times\mathbf{p}]\Psi^L = 0$$

where $R = \left[1-(V-E)/2mc^2\right]^{-1}$. This equation is exact, and it has the small component removed, and the spin-free and spin-dependent terms separated. But it is not very useful because both the energy and the potential are in the denominator.

The Pauli Hamiltonian

One approach to this problem is to expand the inverse operator in powers of $(V-E)/2mc^2$:

$$R = 1 + \frac{(V-E)}{2mc^2} + \frac{(V-E)^2}{4m^2c^4} + \ldots$$

Then the first term in the series gives $\mathbf{p}\cdot\mathbf{p}/2m = \hat{T}$, the nonrelativistic kinetic energy operator. The second term gives

$$(\boldsymbol{\sigma}\cdot\mathbf{p})(V-E)(\boldsymbol{\sigma}\cdot\mathbf{p}) = -(\nabla V)\cdot\nabla + (V-E)\mathbf{p}^2 + \boldsymbol{\sigma}\cdot(\nabla V)\times\mathbf{p}$$

Treating the series as a perturbation series in $1/c^2$, the first term is equivalent in expectation to $\nabla^2 V/2$; for a Coulomb potential (in atomic units), $V = -Z/r$, $\nabla^2 V = Z\delta(r)$. In the second term, $(E-V)$ may be replaced with $\hat{T}$, to give $-\mathbf{p}^4/2m$. For the third term in a Coulomb potential, $\nabla V = Z\mathbf{r}/r^3$, and with $\mathbf{r}\times\mathbf{p} = \boldsymbol{\ell}$ and $\boldsymbol{\sigma} = 2\mathbf{s}$, the expanded Hamiltonian is

$$\hat{\mathcal{H}} = \hat{T} + V - \frac{\mathbf{p}^4}{8m^3c^2} + \frac{Z\delta(r)}{8m^2c^2} + \frac{Z\mathbf{s}\cdot\boldsymbol{\ell}}{2m^2c^2r^3} + \ldots$$

The first added term is the mass-velocity term; the second term is the Darwin term and the third term is the spin-orbit term. The added terms are those of the Pauli Hamiltonian in a Coulomb potential.

As pointed out earlier, this expansion is not valid for small r.

Expansion in the potential

An alternative expansion makes use of the fact that $2mc^2 - V$ is always positive definite. Therefore we may write

$$R = \frac{2mc^2}{2mc^2 - V}\left[1 + \frac{E}{2mc^2 - V}\right]^{-1}$$

and expand the series to give

$$R = \frac{2mc^2}{2mc^2 - V} - \frac{2mc^2 E}{(2mc^2 - V)^2} + \dots .$$

With this expansion, the first term in the expansion of the Hamiltonian is dependent on the potential, and gives

$$\begin{aligned}(\boldsymbol{\sigma}\cdot\mathbf{p})\frac{2mc^2}{2mc^2 - V}(\boldsymbol{\sigma}\cdot\mathbf{p}) &= \frac{2mc^2}{2mc^2 - V}\mathbf{p}^2 \\ &\quad - \frac{2mc^2}{(2mc^2 - V)^2}[(\nabla V)\cdot\nabla - \boldsymbol{\sigma}\cdot(\nabla V)\times\mathbf{p}]\end{aligned}$$

Note that we get a corrected kinetic energy term, which cuts off as V becomes large, a Darwin-type term and a spin-orbit term. The second term is very similar and may be formed by multiplying the first by $-E/(2mc^2 - V)$. Thus the Hamiltonian is

$$\begin{aligned}\hat{\mathcal{H}} = V + \frac{2mc^2}{2mc^2 - V}&\left[1 - \frac{2E}{(2mc^2 - V)} + \frac{3E^2}{(2mc^2 - V)^2} + \dots\right] \\ &\times \left[T + \{2m(2mc^2 - V)\}^{-1}\{(\nabla V)\cdot\nabla - \boldsymbol{\sigma}\cdot(\nabla V)\times\mathbf{p}\}\right]\end{aligned}$$

Expansion in the potential

This expansion has the following features:

- The first term in the expansion contains a modified kinetic energy operator, which cuts off at small r, a Darwin-type term and a spin-orbit term.
- The zeroth-order Hamiltonian is therefore *relativistic*, not nonrelativistic.
- We may *not* replace E with $T + V$ in this expression to remove it from the first-order term, as we did in the expansion in $1/2mc^2$.
- The expansion is valid for all r, and converges.
- However, although it now has the energy in the numerator, the zeroth-order term still has the potential in the denominator.
- This approach is therefore more suited to a method which has a local potential, such as the density functional approach.

Direct perturbation theory

A third approach makes the partitioning in the Dirac equation itself, without eliminating the small component. First, the small component is replaced,

$$\Phi^S = c\Psi^S,$$

which defines a new metric,

$$\hat{\mathcal{G}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & c^{-2}\mathbf{I} \end{pmatrix}.$$

Then the Dirac Hamiltonian is partitioned according to

$$\hat{\mathcal{H}} = \begin{pmatrix} V & (\boldsymbol{\sigma}\cdot\mathbf{p}) \\ (\boldsymbol{\sigma}\cdot\mathbf{p}) & -2m \end{pmatrix} + \frac{1}{c^2}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & V \end{pmatrix} = \hat{\mathcal{H}}_0 + \frac{1}{c^2}\hat{\mathcal{H}}_2.$$

The zeroth-order Hamiltonian is the Lévy-Leblond Hamiltonian. For small components chosen to obey

$$\Phi^S = (\boldsymbol{\sigma}\cdot\mathbf{p})\Psi^L/2m$$

the Lévy-Leblond equation becomes bounded below, with the same solutions as the Schrödinger equation.

We now have an apparently well-behaved operator, in which no expansion has been made which is invalid at small r.

Direct perturbation theory

The perturbation $\hat{\mathcal{H}}'$ may be written in the form

$$\hat{\mathcal{H}}_2 = \hat{\mathcal{P}}^S V \hat{\mathcal{P}}^S$$

where the projector onto the small component is

$$\hat{\mathcal{P}}^S = (\mathbf{I}_4 - \beta)/2.$$

The metric can be written

$$\hat{\mathcal{G}} = \hat{\mathcal{P}}^L + c^{-2}\hat{\mathcal{P}}^S = \hat{\mathcal{G}}_0 + \frac{1}{c^2}\hat{\mathcal{G}}_2.$$

The perturbation parameter is chosen to be $1/c$, and the expansion performed in even powers, in the usual way. The lowest-order energy is

$$\begin{aligned} E_2 &= \langle \Psi_0 | \hat{\mathcal{H}}_2 - E_0 \hat{\mathcal{G}}_2 | \Psi_0 \rangle / \langle \Psi_0 | \hat{\mathcal{G}}_0 | \Psi_0 \rangle \\ &= \langle \Phi^S | (V - E_0) | \Phi^S \rangle \\ &= \frac{1}{4m^2} \langle \Psi^L | (\boldsymbol{\sigma} \cdot \mathbf{p})(V - E_0))(\boldsymbol{\sigma} \cdot \mathbf{p}) | \Psi^L \rangle \end{aligned}$$

where intermediate normalization for the large component has been chosen. This is the same energy as was obtained from the elimination of the small component. Thus, direct perturbation theory *with kinetic balance* gives the same results as the ESC method.

Direct perturbation theory

The problem of the validity of the ESC expansion is apparently not a serious one. There are divergences in the higher-order terms, but a proper formulation in terms of direct perturbation theory can be made which is free of divergences, and is properly bounded below. Therefore, an iterative, or variational, solution may be made.

However, although this provides a clean perturbation theory, the small component is still retained in the equations (albeit with kinetic balance). Moreover, the spin-dependence has not been factored out. Therefore, in the end, the problem of the number of integrals ought to be the same as for the Dirac equation.

The Foldy-Wouthuysen transformation

If it is desirable to remove the small component from consideration, then perhaps one ought to seek a transformation which will block-diagonalize the 4×4 matrix of the Dirac Hamiltonian.

To do this, we first note that the operators in the Dirac equation can be divided into even and odd operators, $\hat{\mathcal{E}}$ and $\hat{\mathcal{O}}$, according to whether they connect components of the same type or of different type. Thus the potential and βmc^2 are even operators, and $\boldsymbol{\alpha} \cdot \mathbf{p}$ is an odd operator. We write symbolically,

$$\hat{\mathcal{H}} = \beta mc^2 + \hat{\mathcal{E}} + \hat{\mathcal{O}}$$

where the rest mass term has been retained explicitly for later convenience. The commutation properties of these operators are:

$$[\hat{\mathcal{E}}, \beta] = 0; \qquad [\hat{\mathcal{O}}, \beta]_+ = 0,$$

i.e. even operators commute with β, odd operators anticommute. Any operator can be written as a sum of even and odd operators as

$$\hat{\mathcal{G}} = \frac{1}{2}(\hat{\mathcal{G}} + \beta\hat{\mathcal{G}}\beta) + \frac{1}{2}(\hat{\mathcal{G}} - \beta\hat{\mathcal{G}}\beta)$$

The Foldy-Wouthuysen transformation

We now make a unitary transformation on the wave function

$$\Psi_1 = e^{i\hat{\mathcal{S}}}\Psi$$

where $\hat{\mathcal{S}}$ is Hermitian, and use the Baker-Campbell-Hausdorff expansion to write the new Hamiltonian in terms of the commutator series,

$$\hat{\mathcal{H}}_1 = \hat{\mathcal{H}} + i[\hat{\mathcal{S}}, \hat{\mathcal{H}}] + \ldots .$$

Here $\hat{\mathcal{H}}$ is the time-dependent Hamiltonian. Choosing

$$\hat{\mathcal{S}} = -i\beta\hat{\mathcal{O}}/2mc^2$$

produces the result

$$i[\hat{\mathcal{S}}, \beta mc^2] = -\hat{\mathcal{O}}$$

which cancels the odd operator in the Dirac Hamiltonian. However, insertion of this operator into the expansion shows that the new Hamiltonian still has odd terms, but they are of higher order in $1/c^2$:

$$\hat{\mathcal{H}}_1 = \beta mc^2 + \hat{\mathcal{E}}_1 + \hat{\mathcal{O}}_1$$

where

$$\hat{\mathcal{O}}_1 = \beta[i\hbar\partial/\partial t - \hat{\mathcal{E}}, \hat{\mathcal{O}}]/2mc^2 - \hat{\mathcal{O}}^3/3m^2c^4 + \mathcal{O}(c^{-3})$$

and

$$\hat{\mathcal{E}}_1 = \hat{\mathcal{E}} + \beta\hat{\mathcal{O}}^2/2mc^2 - [\hat{\mathcal{O}}, [i\hbar\partial/\partial t - \hat{\mathcal{E}}, \hat{\mathcal{O}}]]/8m^2c^4 - \beta\hat{\mathcal{O}}^4/8m^3c^6 + \mathcal{O}(c^{-4})$$

The Foldy-Wouthuysen transformation

Thus the Foldy-Wouthuysen (FW) transformation decouples the large and small components to a particluar order in $1/c$. The lowest-order transformation produces the Pauli Hamiltonian, but the higher-order transformations give highly singluar operators, due to the presence of the potential.

For a free particle, however, the transformation can be performed exactly with

$$\hat{\mathcal{U}} = e^{i\hat{\mathcal{S}}} = \hat{\mathcal{A}}(1 + \beta\hat{\mathcal{R}})$$

and the operators are

$$\hat{\mathcal{A}} = \left(\frac{E_p + mc^2}{2E_p}\right)^{1/2}, \qquad \hat{\mathcal{R}} = \frac{c\boldsymbol{\alpha}\cdot\mathbf{p}}{E_p + mc^2}$$

where the energy operator is

$$E_p = (c^2\mathbf{p}^2 + m^2c^4)^{1/2}$$

The Douglas-Kroll transformation

Applying the free-particle FW transformation to the Dirac operator in momentum space with the external potential

$$V_{ext}\Psi(\mathbf{p}) = \int V_{ext}(\mathbf{p},\mathbf{p}')\mathrm{d}^3p'$$

gives the transformed Hamiltonian as above, with the even and odd operators in a more compact form:

$$\hat{\mathcal{E}}_1 = \hat{\mathcal{A}}(V_{ext} + \hat{\mathcal{R}}V_{ext}\hat{\mathcal{R}})\hat{\mathcal{A}}$$

$$\hat{\mathcal{O}}_1 = \beta\hat{\mathcal{A}}(\hat{\mathcal{R}}V_{ext} - V_{ext}\hat{\mathcal{R}})\hat{\mathcal{A}}.$$

Instead of expanding in powers of $1/c$, another unitary operator is defined,

$$\hat{\mathcal{U}}_1 = (1 + \hat{\mathcal{W}}_1^2)^{1/2} + \hat{\mathcal{W}}_1$$

where $\hat{\mathcal{W}}_1$ is anti-Hermitian. This transformation is applied to $\hat{\mathcal{H}}_1$ and the result expressed in terms of a commutator expansion by expanding the square root operator in powers of $\hat{\mathcal{W}}_1$.

$$\hat{\mathcal{H}}_2 = \hat{\mathcal{U}}_1\hat{\mathcal{H}}_1\hat{\mathcal{U}}_1^{-1} = \beta E_p - [\beta E_p, \hat{\mathcal{W}}_1] + \hat{\mathcal{E}}_1 + \hat{\mathcal{O}}_1 + \ldots$$

Then $\hat{\mathcal{O}}_1$ disappears if

$$\hat{\mathcal{W}}_1 E_p + E_p\hat{\mathcal{W}}_1 = \beta\hat{\mathcal{O}}_1.$$

It can be shown that $\hat{\mathcal{W}}_1$ is of order V_{ext}, and thus the Douglas-Kroll (DK) transformation is expanded in powers of the external potential rather than $1/c$. The above transformation gives an operator correct to second order in the external potential which is non-singular.

Summary of approximate methods

The conclusions of our consideration of the approximate methods are:

- Elimination of the small component with expansion in powers of $1/c$, either directly or via the FW transformation, gives rise to singular operators, but ones which have a generally simple form. These must be used in perturbation theory.
- Elimination of the small component with expansion in powers of V, either directly or via the DK transformation gives rise to non-singular operators which may then be used in variational calculations, but are complicated in form.
- The use of perturbation theory in $1/c$ without elimination of the small component gives an equation which is useful, but has the same integral size problems as in the Dirac equation. Spin separation may be achieved at the matrix level.

We therefore turn to another method, in which the operators have simple form, the small component is retained and spin separation is achieved.

Modified Dirac equation

To achieve the spin separation, we note that in the perturbation theories presented above, the spin-dependence could be factored because of the elimination of the small component. In the nonrelativistic limit,

$$\lim_{c\to\infty} 2mc\ \Psi^S = (\boldsymbol{\sigma}\cdot\mathbf{p})\Psi^L.$$

Using this directly to define an auxiliary large component at finite c

$$2mc\ \Psi^S \equiv (\boldsymbol{\sigma}\cdot\mathbf{p})\Phi^L$$

and substituting into the Dirac equation, premultiplying the second equation by $(\boldsymbol{\sigma}\cdot\mathbf{p})/2mc$ gives

$$(V-E)\Psi^L + T\Phi^L = 0$$

$$T\Psi^L + [(\boldsymbol{\sigma}\cdot\mathbf{p})(V-E)(\boldsymbol{\sigma}\cdot\mathbf{p})/4m^2c^2 - T]\Phi^L = 0.$$

Eliminating $T\Phi^L$ from the first equation using the second,

$$(T+V-E)\Psi^L + \frac{1}{4m^2c^2}\big[(\boldsymbol{\sigma}\cdot\mathbf{p})(V-E)(\boldsymbol{\sigma}\cdot\mathbf{p})\big]\Phi^L = 0,$$

which is suspiciously like the Pauli equation. But this equation has no approximations, and since the same highly singular operators appear, why should it be variationally stable? The difference is in the presence of Φ^L rather than Ψ^L.

Modified small component

We therefore make an expansion of the wave function in powers of r, and find the following.

For a point nucleus:

- For small r, $\Phi^L \sim r\Psi^L$
- Integrals over the singular operators are therefore well-behaved.
- $\Phi^L \to \Psi^L$ in the limit $c \to \infty$ but only as an asymptotic expansion in r.

For a finite nucleus:

- For small r, $\Phi^L \sim \Psi^L$
- Operators are non-singular and integrals are well-behaved.
- $\Phi^L \to \Psi^L$ smoothly in the limit $c \to \infty$.

Thus, even though there are singular operators present, these are regularized by the wave function, in particular by the retention of the small component. This should be compared with the regularization of the operators in the potential expansion perturbation theory.

Modified Dirac operators

Spin separation is achieved in exactly the same manner as in the Pauli equation, giving the modified Dirac Hamiltonian (in atomic units)

$$\tilde{h}_D = \begin{pmatrix} V & T \\ T & \frac{\alpha^2}{4}\mathbf{p}\cdot V\mathbf{p} - T \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \frac{\alpha^2}{4} i\boldsymbol{\sigma}\cdot(\mathbf{p}V)\times\mathbf{p} \end{pmatrix},$$

which has a new metric

$$\tilde{\mathbf{G}} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{\alpha^2}{2}T \end{pmatrix}$$

The modified wave function is

$$\tilde{\Psi} = \begin{pmatrix} \Psi^L \\ \Phi^L \end{pmatrix}.$$

- Φ^L has the same symmetry properties as Ψ^L and may therefore be expanded in the same (primitive) basis.
- The modified equation has the same spectrum as the original equation: bound (exponentially decaying) solutions between 0 and $-2mc^2$ and continuum (oscillatory) solutions above 0 and below $-2mc^2$, whether or not the spin-dependent terms are neglected.
- The spin-free Hamiltonian is now real, so that spin factorization may be used and the wave function expanded in a real scalar basis.

Comparison with the Pauli and DK Hamiltonians

The energy for the modified Dirac equation may be written

$$\begin{aligned} E\langle\Psi|\Psi\rangle =& \langle\Psi|T+V|\Psi\rangle - \langle\Psi-\Phi|T|\Psi-\Phi\rangle \\ &+\frac{\alpha^2}{4}\langle\Phi|(\boldsymbol{\sigma}\cdot\mathbf{p})(V-E)(\boldsymbol{\sigma}\cdot\mathbf{p})|\Phi\rangle \end{aligned}$$

The second term is the modification of the kinetic energy. If we replace Φ with $\Psi \equiv \Psi^{NR}$,

- The second term vanishes.
- The third term gives the Pauli energy.

As noted before, this makes the Pauli Hamiltonian non-variational because of the incorrect small r behaviour.

Comparison with the Pauli and DK Hamiltonians

The DK energy for a single electron is

$$
\begin{aligned}
E\langle\Psi|\Psi\rangle &= \langle\Psi|E_p - mc^2 - \mathcal{W}|\Psi\rangle + \langle\Psi|\mathcal{A}\,V\mathcal{A} + \mathcal{A}\mathcal{R}\,V\mathcal{R}\mathcal{A}|\Psi\rangle \\
&= \langle\Psi|E_p - mc^2 - \mathcal{W}|\Psi\rangle + \langle\Psi|\mathcal{A}\,V\mathcal{A}|\Psi\rangle \\
&+ \langle\Psi|\frac{c\mathcal{A}}{E_p + mc^2}\,(\boldsymbol{\sigma}\cdot\mathbf{p})V(\boldsymbol{\sigma}\cdot\mathbf{p})\,\frac{c\mathcal{A}}{E_p + mc^2}|\Psi\rangle
\end{aligned}
$$

- The first term is essentially the relativistic kinetic energy, the second term is the nuclear Coulomb potential term, the third term contributes a Darwin-type term and a spin-orbit term.
- The perturbation in the DK method is the potential, the coupling constant is e instead of α, the zeroth order Hamiltonian that of a relativistic free particle.
- The DK transformation introduces momentum-dependent kinematic factors $\mathcal{A}$ and $\mathcal{R}$ which account for the renormalization of Ψ and the difference between Φ and Ψ. The integrals involving V are thereby regularized and can be used in variational calculations.
- Setting $\mathbf{p} = 0$ in the kinematic factors makes $\mathcal{A} = 1$ and $c\mathcal{A}/(E_p + mc^2) = 1/2mc$, and the terms from the Pauli Hamiltonian are regained.

Two-electron terms

Expanding the two-electron integrals, we have for the Coulomb interaction

$$(\Psi_i\Psi_i|\Psi_j\Psi_j) = (\Psi_i^L\Psi_i^L|\Psi_j^L\Psi_j^L) + (\Psi_i^L\Psi_i^L|\Psi_j^S\Psi_j^S) + (\Psi_i^S\Psi_i^S|\Psi_j^L\Psi_j^L) + (\Psi_i^S\Psi_i^S|\Psi_j^S\Psi_j^S).$$

Substituting for the small components, we end up with the operator

$$\tilde{g}^C = \hat{\mathcal{P}}_i^+\hat{\mathcal{P}}_j^+ \; g_{ij}^{00} \; \hat{\mathcal{P}}_i^+\hat{\mathcal{P}}_j^+ + \frac{\alpha^2}{4}\hat{\mathcal{P}}_i^+\hat{\mathcal{P}}_j^- \; g_{ij}^{02} \; \hat{\mathcal{P}}_i^+\hat{\mathcal{P}}_j^- + \frac{\alpha^2}{4}\hat{\mathcal{P}}_i^-\hat{\mathcal{P}}_j^+ \; g_{ij}^{20} \; \hat{\mathcal{P}}_i^-\hat{\mathcal{P}}_j^+ + \frac{\alpha^4}{16}\hat{\mathcal{P}}_i^-\hat{\mathcal{P}}_j^- \; g_{ij}^{22} \; \hat{\mathcal{P}}_i^-\hat{\mathcal{P}}_j^-$$

where the projectors onto the large and small component space are

$$\hat{\mathcal{P}}^{\pm} = \frac{1}{2}[1 \pm \beta].$$

The operators are

$$g_{ij}^{00} = \frac{1}{r_{ij}}$$

$$g_{ij}^{02} = (\boldsymbol{\sigma}_j \cdot \mathbf{p}_j) \; \frac{1}{r_{ij}} \; (\boldsymbol{\sigma}_j \cdot \mathbf{p}_j)$$

$$g_{ij}^{20} = (\boldsymbol{\sigma}_i \cdot \mathbf{p}_i) \; \frac{1}{r_{ij}} \; (\boldsymbol{\sigma}_i \cdot \mathbf{p}_i)$$

$$g_{ij}^{22} = (\boldsymbol{\sigma}_i \cdot \mathbf{p}_i)(\boldsymbol{\sigma}_j \cdot \mathbf{p}_j) \; \frac{1}{r_{ij}} \; (\boldsymbol{\sigma}_j \cdot \mathbf{p}_j)(\boldsymbol{\sigma}_i \cdot \mathbf{p}_i).$$

Two-electron terms

The first of the four operators is the normal Coulomb operator. The second and third are related by index interchange, and have their analogies in the one-electron terms:

$$g_{ij}^{20} = -\nabla_i \frac{1}{r_{ij}} \cdot \nabla_i + \boldsymbol{\sigma}_i \cdot \nabla_i \frac{1}{r_{ij}} \times \mathbf{p}_i .$$

These two terms contribute to the *two-electron Darwin* and *two-electron spin-(own-)orbit* terms. The fourth term is of $\mathcal{O}(\alpha^4)$, and contributes a "double Darwin", a mixed spin-orbit/Darwin term and a "double spin-orbit" term.

If we consider using the spin-free Hamiltonian variationally and the spin-dependent terms as a perturbation, we may work in a scalar basis. Then there are only twice as many Darwin integrals as Coulomb integrals, and as many double-Darwin integrals as Coulomb integrals Thus, the spin-free calculation involves only *four* times the number of integrals required in a nonrelativistic calculation.

Now, considering the spin-dependent integrals, there are 6 times as many spin-orbit and mixed Darwin-spin-orbit integrals as Coulomb integrals, and 9 times as many double-spin-orbit integrals, giving a total factor of 21 times the number of Coulomb integrals. A full relativistic calculation with the modified Dirac equation would then be 25 times the cost of a NR calculation, compared with a factor of 64 for the unmodified Dirac equation, of which a factor of 2 is due to the complex arithmetic.

Two-electron terms

The Gaunt interaction contributes operators which arise from the product

$$(\boldsymbol{\sigma}_i \cdot \mathbf{p}_i)(\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j)(\boldsymbol{\sigma}_j \cdot \mathbf{p}_j) = \mathbf{p}_i \cdot \mathbf{p}_j + i(\boldsymbol{\sigma}_i + \boldsymbol{\sigma}_j) \cdot (\mathbf{p}_i \times \mathbf{p}_j) + (\boldsymbol{\sigma}_i \times \mathbf{p}_i) \cdot (\boldsymbol{\sigma}_j \times \mathbf{p}_j)$$

The first of these is a Darwin-type term, but *not* the same as that which arises from the Coulomb interaction. The second contributes the *spin-other-orbit* interaction. The third contributes the *spin-spin* and *Fermi contact* interactions.

Factors for integral numbers from these terms are: 4 for the Darwin-type term, 12 for the spin-other-orbit term and 24 for the spin-spin terms. However many of the spin-spin terms are counted in the other two; the total factor is 36.

A similar reduction may be performed for the gauge term of the Breit interaction, but it is more involved, and results only in the addition of the *orbit-orbit* interaction to the Gaunt interaction.

Summary: modified Dirac equation

The features of the modified Dirac equation may be summarized:

- The modified equation is exactly equivalent to the original.
- Spin separation is achieved without any approximations.
- The spin-free Hamiltonian is real, leading to real 2-component orbitals.
- "Exact" spin-free Dirac-Coulomb SCF calculations are only 4 times as expensive as a nonrelativistic calculation.
- Correlated calculations in the no-pair approach are no more expensive than nonrelativistic calculations.
- The spin factorization means that all the nonrelativistic machinery may be used, both for spin-free and spin-dependent calculations.
- The integrals are the same as those over the Breit-Pauli operators.
- If contraction is employed, a dual basis must be defined, and care taken in the integral evaluation.
- In this case, it is better to use the original Dirac equation if a variational spin-dependent calculation is to be performed.