SPIN

Mathematics of Spin

		page
$\mathbf{A}$	Example matrix	2
$\vec{\mathrm{A}}$	Magnetic vector potential	11
Av()	Average value of	16
$\vec{\mathrm{B}}$	Magnetic field	11
c	Velocity of light	10
$\vec{\mathrm{E}}$	Electric field	11
$\mathcal{E}$	Total energy	10
$\vec{\mathrm{F}}$	Force	11, 26
$\mathbf{H}$	Hamiltonian	
$\mathbf{I}$	Unit matrix	4
m	Mass	11
m_0	Rest mass	11
$\mathbf{O_{op}}$	Quantum mechanical operator	16
p	Momentum	11
$\hbar$	Planck's constant	
$\mathcal{V}$	Potential energy	
$\dagger$	Hermetian conjugate	4
μ	Magnetic moment	25
$\|\psi>$	Two component wave function	6
β	Fraction light velocity	10
γ	Relativity correction factor	9
σ	Pauli matrix	6, 26
ς	Value of matrix	9, 32

Matrix Algebra

Although probably familiar, here are the elementary matrix operations. [1]

For a matrix $\mathbf{A}$ as shown below,

$$\mathbf{A} = \begin{vmatrix} A_{0,0} & A_{0,1} & A_{0,2} & \cdots & A_{0,n} \\ A_{1,0} & A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,0} & A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{m,0} & A_{m,1} & \cdots & \cdots & A_{m,n} \end{vmatrix}$$

To add another such matrix $\mathbf{B}$ one adds corresponding terms so that the terms of the sum matrix $\mathbf{S} = \mathbf{A} + \mathbf{B}$ are:

$$S_{i,j} = A_{i,j} + B_{i,j}.$$

or

$$\begin{vmatrix} A_{0,0} & A_{0,1} & A_{0,2} & \cdots & A_{0,n} \\ A_{1,0} & A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,0} & A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{m,0} & A_{m,1} & \cdots & \cdots & A_{m,n} \end{vmatrix} + \begin{vmatrix} B_{0,0} & B_{0,1} & B_{0,2} & \cdots & B_{0,n} \\ B_{1,0} & B_{1,1} & B_{1,2} & \cdots & B_{1,n} \\ B_{2,0} & B_{2,1} & B_{2,2} & \cdots & B_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ B_{m,0} & B_{m,1} & \cdots & \cdots & B_{m,n} \end{vmatrix} =$$

$$\begin{vmatrix} A_{0,0} + B_{0,0} & A_{0,1} + B_{0,1} & A_{0,2} + B_{0,2} & \cdots & A_{0,n} + B_{0,n} \\ A_{1,0} + B_{1,0} & A_{1,1} + B_{1,1} & A_{1,2} + B_{1,2} & \cdots & A_{1,n} + B_{1,n} \\ A_{2,0} + B_{2,0} & A_{2,1} + B_{2,1} & A_{2,2} + B_{2,2} & \cdots & A_{2,n} + B_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{m,0} + B_{m,0} & A_{m,1} + B_{m,1} & \cdots & \cdots & A_{m,n} + B_{m,n} \end{vmatrix}$$

It is assumed in the above that $\mathbf{A}$ has the same number of rows and the same number of columns as $\mathbf{B}$. If this is not the case, the missing terms are, on occasion, filled with zeros.[2]

[1] Excellent compact reference:Margenau, H. and Murphy

[2] The need to do this often suggests an error.

The product **P** of **A** and **B** is slightly more complicated and requires that the left term, **A** have the same number of columns , in this case n + 1, as B does rows. In this case the matrices are referred to as conformable. The product **P** = **AB** is then given by:

$$P_{i,j} = \sum_{k=0}^{k=n} A_{i,k} B_{k,j}$$

.

$$\begin{vmatrix} A_{0,0} & A_{0,1} & A_{0,2} & \cdots & A_{0,n} \\ A_{1,0} & A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,0} & A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{m,0} & A_{m,1} & \cdots & \cdots & A_{m,n} \end{vmatrix} \times \begin{vmatrix} B_{0,0} & B_{0,1} & B_{0,2} & \cdots & B_{0,m} \\ B_{1,0} & B_{1,1} & B_{1,2} & \cdots & B_{1,m} \\ B_{2,0} & B_{2,1} & B_{2,2} & \cdots & B_{2,m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ B_{n,0} & B_{n,1} & \cdots & \cdots & B_{n,m} \end{vmatrix} =$$

$$\begin{vmatrix} \sum A_{0,k}B_{k,0} & \sum A_{0,k}B_{0,1} & \sum A_{0,k}B_{k,2} & \sum \cdots & \sum A_{0,k}B_{k,m} \\ \sum A_{1,k}B_{k,0} & \sum A_{1,k}B_{k,1} & \sum A_{1,k}B_{k,2} & \sum \cdots & \sum A_{1,k}B_{k,m} \\ \sum A_{2,k}B_{k,0} & \sum A_{2,k}B_{k,1} & \sum A_{2,k}B_{k,2} & \sum \cdots & \sum A_{2,k}B_{k,m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum A_{m,k}B_{k,0} & \sum A_{m,k}B_{k,1} & \sum \cdots & \sum \cdots & \sum A_{m,k}B_{k,m} \end{vmatrix}$$

In the above, the top left element is given the subscripts $0, 0$. For many purposes it is more convienent to have the top left begin with $1, 1$. It makes no difference in the results except that for a three by three matrix the terms are numbered 1 to 3 rather than 0 to 2.

One should also note that the product of two matrices depends upon the order of the factors.

Thus except in special cases

$$\mathbf{AB} \neq \mathbf{BA}$$

For our purposes we need to define only a few of the familiar matrix operations. The inverse of a matrix **M** (defined only for n=m) indicated by

a superscript $\mathbf{M^{-1}}$ is the matrix such that the product $\mathbf{M^{-1}M}$ is a matrix consisting of ones along the diagonal and all off diagonal terms zero.

$$\mathbf{M^{-1}M = I} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

and the transpose of a matrix (n need not equal m) which is obtained by interchanging rows and columns. Thus:

$$\begin{vmatrix} A_{0,0} & A_{0,1} & A_{0,2} & \cdots & A_{0,n} \\ A_{1,0} & A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,0} & A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{m,0} & A_{m,1} & \cdots & \cdots & A_{m,n} \end{vmatrix}^{T} = \begin{vmatrix} A_{0,0} & A_{1,0} & A_{2,0} & \cdots & A_{m,0} \\ A_{0,1} & A_{1,1} & A_{2,1} & \cdots & A_{m,1} \\ A_{0,2} & A_{1,2} & A_{2,2} & \cdots & A_{m,2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{0,n} & A_{1,n} & \cdots & \cdots & A_{m,n} \end{vmatrix} .$$

One more called the Hermitian conjugate indicated by † is the transpose with each term replaced by its complex congugate.

$$\begin{vmatrix} A_{0,0} & A_{0,1} & A_{0,2} & \cdots & A_{0,n} \\ A_{1,0} & A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,0} & A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{m,0} & A_{m,1} & \cdots & \cdots & A_{m,n} \end{vmatrix}^{\dagger} = \begin{vmatrix} A^*_{0,0} & A^*_{1,0} & A^*_{2,0} & \cdots & A^*_{m,0} \\ A^*_{0,1} & A^*_{1,1} & A^*_{2,1} & \cdots & A^*_{m,1} \\ A^*_{0,2} & A^*_{1,2} & A^*_{2,2} & \cdots & A^*_{m,2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A^*_{0,n} & A^*_{1,n} & \cdots & \cdots & A^*_{m,n} \end{vmatrix} .$$

There is one special case of the above that we will use in the section *Angular Momentum*. This is the matrix consisting of one column and indicated by

$$|\psi> = \begin{vmatrix} \psi_\alpha \\ \psi_\beta \end{vmatrix} .$$

Its Hermitian conjugate is written as

$$< \psi| = \left| \begin{array}{cc} \psi_\alpha^* & \psi_\beta^* \end{array} \right| .$$

[3]

This particular notation is used to simplify the appearance of the product

$$< \psi| \ |\psi >= \left| \begin{array}{cc} \psi_\alpha^* & \psi_\beta^* \end{array} \right| \left| \begin{array}{c} \psi_\alpha \\ \psi_\beta \end{array} \right| = \psi_\alpha^* \psi_\alpha + \psi_\beta^* \psi_\beta = |\psi_\alpha|^2 + |\psi_\beta|^2$$

In pre Dirac quantum mechanics, the ψ function is assumed to be normalized so that the integral of its square, defined as $\psi^* \psi$, over all space is equal to one.

$$\int_{\text{all space}} \psi^* \psi \ \ \text{d space} = 1.$$

The average value of a measurable quantity represented by an operator $\mathbf{O}$ is given by

$$\text{Average of O} = \int_{\text{all space}} \psi^* \mathbf{O} \psi \ \ \text{d space}$$

This is often presented as one of the assumptions of axiom based quantum mechanics, but is in fact largly derivable from very physical considerations. Much as the author would like to discuss these, they are far better done in Bohm's *Quantum Mechanics* and would lead one far astray.

The critical point here is that we will introduce two element ψ functions and the operations corresponding to the above are then

$$\int_{\text{all space}} < \psi| \ |\psi > \quad \text{d space} = 1$$

[3] It is customary to omit the asterisk on the $< \psi|$ term even though its components are complex conjugates of those of $|\psi >$.

and

$$\text{Average of O} = \int_{\text{all space}} < \psi|\mathbf{O}|\psi > \quad \text{d space.}$$

The Pauli Matrices.

In order to make the cross product terms of Eq. 22 cancel each other, the Pauli matrices must satisfy the following conditions:

Condition 1:

$$\sigma_{\mathbf{x}}\sigma_{\mathbf{x}} = \mathbf{I}, \qquad \sigma_{\mathbf{y}}\sigma_{\mathbf{y}} = \mathbf{I}, \qquad \sigma_{\mathbf{z}}\sigma_{\mathbf{z}} = \mathbf{I}$$

and Condition 2:

$$\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} = -\sigma_{\mathrm{y}}\sigma_{\mathrm{x}}, \qquad \sigma_{\mathrm{x}}\sigma_{\mathrm{z}} = -\sigma_{\mathrm{z}}\sigma_{\mathrm{x}}, \qquad \sigma_{\mathrm{y}}\sigma_{\mathrm{z}} = -\sigma_{\mathrm{z}}\sigma_{\mathrm{y}}.$$

Strictly speaking, in each of the products above, the first term should be shown as a Hermitian conjugate (†) . Thus the last line should be

$$\sigma_{\mathrm{y}}^{\dagger}\sigma_{\mathrm{z}} = -\sigma_{\mathrm{z}}^{\dagger}\sigma_{\mathrm{y}}.$$

But, we will find out that each of the σ happens to be its own Hermitian conjugate, (transposed complex conjugate) so that one need not indicate σ or $\sigma^{\dagger}$ in the products above. Its just a little simplier this way.

However, in the derivation we cannot assume this result so that when we multiply two matrices we require the first to be the transposed complex conjugate (indicated by †) of the second.

The first thing we observe is that the unit matrix cannot be one of the σ_{s}. If it were, the second condition above would lead to the unpleasant result that the others σ are all zero. For example if we try

$$\sigma_{\mathbf{x}} = \mathrm{I}, \qquad \text{As a temporary test only.}$$

Then,

$$\sigma_{\mathbf{y}} = -\sigma_{\mathbf{y}} \quad \text{which leads to } \sigma_{\mathbf{y}} = 0 \qquad \text{An unacceptable result.}$$

Having excluded the unit matrix, we try as a starting point

$$\mathbf{M_{trial}} = \begin{vmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{vmatrix}$$

the square of which is found by multiplying on the left by its complex conjugate transpose:

$$\mathbf{M^{\dagger}_{trial}} = \begin{vmatrix} \mathrm{A}^* & \mathrm{C}^* \\ \mathrm{B}^* & \mathrm{D}^* \end{vmatrix}.$$

This product must now be the unit matrix **I**.

$$\mathbf{M^2_{trial}} = \begin{vmatrix} \mathrm{A}^* & \mathrm{C}^* \\ \mathrm{B}^* & \mathrm{D}^* \end{vmatrix} \begin{vmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{vmatrix} = \begin{vmatrix} \mathrm{AA}^* + \mathrm{CC}^* & \mathrm{A}^*\mathrm{B} + \mathrm{C}^*\mathrm{D} \\ \mathrm{AB}^* + \mathrm{CD}^* & \mathrm{BB}^* + \mathrm{DD}^* \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

There are many ways of satisfying the above of which the most obvious is $\mathrm{AA}^* = 1$ and $\mathrm{DD}^* = 1$ with B and C both zero. Thus we can have $\mathrm{A} = \pm 1$ and $\mathrm{D} = \pm 1$ but they must be chosen with different signs to avoid the unit matrix exclusion noted above. Pauli's choice was $\mathrm{A} = 1,\ \mathrm{D} = -1$.

Clearly other choices are possible. For example $\mathrm{A} = \mathrm{D} = 0,\ \mathrm{BB}^* = \mathrm{CC}^* = 1$ and in fact we will use these as the other two σ_{s} if we can make them satisfy the second set of conditions.

There are still other possibilities but these are used because they have two of the four elements equal to zero which makes manipulation easier but nothing in the logic requires this.

In any case we follow Pauli's lead and use

$$\sigma_{\mathbf{z}} = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}$$

and see if we can find the others from this.

To find $\sigma_{\mathbf{x}}$ and $\sigma_{\mathbf{y}}$ we try the second set of possibilities from above. Let $\mathrm{A} = \mathrm{D} = 0$ and $\mathrm{BB}^* = \mathrm{CC}^* = 1$

From Condition 2 using $\sigma_{\mathbf{z}}$ as one of the matrices we have

$$\sigma_{\mathbf{z}}{}^{\dagger}\mathrm{M}_{\mathrm{trial}} = -\mathrm{M}^{\dagger}_{\mathrm{trial}}\sigma_{\mathbf{z}} = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} \begin{vmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{vmatrix} = - \begin{vmatrix} \mathrm{A}^* & \mathrm{C}^* \\ \mathrm{B}^* & \mathrm{D}^* \end{vmatrix} \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}$$

which results in

$$\begin{vmatrix} \mathrm{A} & \mathrm{B} \\ -\mathrm{C} & -\mathrm{D} \end{vmatrix} = \begin{vmatrix} -\mathrm{A}^* & \mathrm{C}^* \\ -\mathrm{B}^* & \mathrm{D}^* \end{vmatrix}$$

Now Condition 2 can be met by either $\mathrm{B} = 1$ $\mathrm{C} = 1$ or $\mathrm{B} = -\mathrm{i}$ $\mathrm{C} = \mathrm{i}$.

$$\sigma_{\mathbf{x}} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

$$\sigma_{\mathbf{y}} = \begin{vmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{vmatrix}$$

These are now the Pauli matrices as used in the text.

Of course these could be interchanged or the signs interchanged in σ_{y}, but these particular choices lead to the easily remembered relations:

$$\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} = \mathrm{i}\sigma_{\mathrm{z}}$$
$$\sigma_{\mathrm{y}}\sigma_{\mathrm{z}} = \mathrm{i}\sigma_{\mathrm{x}}$$
$$\sigma_{\mathrm{z}}\sigma_{\mathrm{x}} = \mathrm{i}\sigma_{\mathrm{y}}$$

Value of a Pauli Matrix

To assign a value to a Pauli matrix, consider the following. Let

$$|\mathrm{vv}>=\left|\begin{array}{c}\mathrm{AA}\\ \mathrm{BB}\end{array}\right|$$

be an arbitrary but non zero vector. Its length squared is found by multiplying it by its Hermitian conjugate, its transposed complex congugate.

$$\mathrm{Length}^2\mathrm{vv}=\left|\begin{array}{cc}\mathrm{AA}^*, & \mathrm{BB}^*\end{array}\right|\left|\begin{array}{c}\mathrm{AA}\\ \mathrm{BB}\end{array}\right|=(\mathrm{AA})^2+(\mathrm{BB})^2$$

Now suppose one multiplys $|\mathrm{vv}>$ by a constant ς. Each component of the vector is mutiplied by the same constant which multiplys the Length^2 by ς^2.

Now define a new vector $|\mathrm{ww}>$ by mulitplying $|\mathrm{vv}>$ on the left by one of the Pauli matrices σ. Its Hermitian conjugate is $<\mathrm{ww}|$ and the Hermitian conjugate product would be $<\mathrm{vv}|\sigma$. Then the square of the length of $|\mathrm{ww}>$ would be

$$\mathrm{Length}^2\mathrm{ww}=<\mathrm{vv}|\sigma|\ \ |\sigma|\mathrm{vv}>=<\mathrm{vv}|\mathbf{I}|\mathrm{vv}>=<\mathrm{vv}||\mathrm{vv}>$$

Once again if we multiply each component of $|\mathrm{vv}>$ by ς this product must be multiplied ς^2. This forces us to assign the value of ± 1 to ς.

Relativity

Here is the difficulty that relativity introduces into the Schrödiger form of quantum mechanics. In one dimension Schrödiger's the equation is

$$\left(-\frac{\hbar^2}{2\mathrm{m}}\frac{\partial^2}{\partial \mathrm{x}^2}+\mathrm{V}(\mathrm{x})\right)\psi=-\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial \mathrm{t}}\psi.$$

Let us subject this to a Lorentz transformation.

$$x' = \gamma(x - vt)$$

$$t' = \gamma(t - vx/c^2)$$

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

$$\frac{\partial}{\partial x} = \frac{\partial x'}{\partial x}\frac{\partial}{\partial x'} = \gamma\frac{\partial}{\partial x'}$$

$$\frac{\partial}{\partial x^2} = \gamma^2\frac{\partial}{\partial x'^2}$$

In the primed coordinate system we would have

$$\left(-\frac{\hbar^2}{2m}\gamma^2\frac{\partial^2}{\partial x'^2} + V(x')\right)\psi = -\frac{\hbar}{i}\gamma\frac{\partial}{\partial t'}\psi$$

Even putting aside the effect of the potential energy, the new equation is not the same as Schrödiger's. The γ appears in different powers in different terms. This is a serious problem since it showed that Schrödiger's equation was fundamentally flawed.

A moving body of rest mass m_0 has a true, physically measurable, mass m related to its velocity by

$$m = \frac{m_0}{\sqrt{1 - \beta^2}} \quad \text{where} \quad \beta = \frac{v}{c} \quad \text{with} \quad v^2 = v_x^2 + v_y^2 + v_z^2$$

This is the mass that occurs in the famous $\mathcal{E} = mc^2$. It is also the mass used to calculate momentum:

$$p_x = mv_x = \frac{m_0 v_x}{\sqrt{1 - \beta^2}}, \qquad p_y = mv_y, \qquad p_z = mv_z.$$

However, it has the nasty property that it depends upon the coordinate system since a body stationary in one coordinate system can be moving in another. The quantity that has the same value in any inertial coordinate system is the rest mass m_0.

In order to express physical laws in a way that will remain correct under Lorentz transformations, one expresses mass in term of the invariant quantity m_0. Then

$$\mathcal{E} = \frac{1}{\sqrt{1-\beta^2}} m_0 c^2 \quad \text{or} \quad \mathcal{E}^2 = \frac{1}{1-\beta^2} m_0^2 c^4 \tag{1}$$

Solving for β in terms of the momenta.

$$p^2 = p_x^2 + p_y^2 + p_z^2 = \frac{m_0^2 v^2}{1-\beta^2} = m_0^2 c^2 \frac{\beta^2}{1-\beta^2}$$

$$(1-\beta^2)p^2 = m_0^2 c^2 \beta^2 \quad \text{or} \quad \beta^2(p^2 + m_0^2 c^2) = p^2$$

Thus

$$1 - \beta^2 = \frac{m_0^2 c^2}{p^2 + m_0^2 c^2}$$

which becomes when inserted into Eq. 1

$$\mathcal{E}^2 = m_0^2 c^4 + c^2 p^2$$

which is the form used to derive Dirac's equation.

Equation of Motion in a Magnetic Field .

In addition to electrostatic forces, moving electric charges are subject to a force from a magnetic field. For a charge q moving with velocity $\vec{v}$ the so called Lorentz force is:

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})$$

Because $\vec{E}, \vec{v}$ and $\vec{B}$ are all in different directions, except in simple cases, this form is difficult to use in the solution of mechanics problems.

It turns out to be better to introduce the magnetic vector potential $\vec{A}$ defined as

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

It may seem strange that introducing a quantity defined only by its curl makes things simpler but it does. Unfortunately all this simplicity begins with a theorem needed to calculate the $\vec{v} \times \vec{B}$ above, in terms of $\vec{A}$ which is particularly tedious to derive.

First let us find $\vec{v} \times \vec{B}$.

$$\begin{aligned}
\vec{v} \times \vec{B} = \vec{v} \times \vec{\nabla} \times \vec{A} = \\
\tilde{x}(v_y\ (\vec{\nabla} \times \vec{A})_z - v_z\ (\vec{\nabla} \times \vec{A})_y) + \\
\tilde{y}(v_z\ (\vec{\nabla} \times \vec{A})_x - v_x\ (\vec{\nabla} \times \vec{A})_z) + \\
\tilde{z}(v_x\ (\vec{\nabla} \times \vec{A})_y - v_y\ (\vec{\nabla} \times \vec{A})_x)
\end{aligned}$$

$$\begin{aligned}
(\vec{\nabla} \times \vec{A})_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \\
(\vec{\nabla} \times \vec{A})_y &= \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \\
(\vec{\nabla} \times \vec{A})_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}
\end{aligned}$$

So that

$$\begin{aligned}
\vec{v} \times \vec{B} = \vec{v} \times \vec{\nabla} \times \vec{A} = \\
\tilde{x}\left(v_y\ (\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}) - v_z\ (\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x})\right) +
\end{aligned}$$

$$\tilde{y}\left(v_z\,(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z})-v_x\,(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y})\right)+$$
$$\tilde{z}\left(v_x\,(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x})-v_y\,(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z})\right)$$

Next we add and subtract a term from each component to allow the terms to group in a familiar fashion:

$$\vec{v}\times\vec{B}=\vec{v}\times\vec{\nabla}\times\vec{A}=$$
$$\tilde{x}\left(v_y\,(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y})-v_z\,(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x})+v_x\frac{\partial A_x}{\partial x}-v_x\frac{\partial A_x}{\partial x}\right)$$
$$\tilde{y}\left(v_z\,(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z})-v_x\,(\frac{\partial A_y}{\partial x}-\frac{\partial A_x}{\partial y})+v_y\frac{\partial A_y}{\partial y}-v_y\frac{\partial A_y}{\partial y}\right)$$
$$\tilde{z}\left(v_x\,(\frac{\partial A_x}{\partial z}-\frac{\partial A_z}{\partial x})-v_y\,(\frac{\partial A_z}{\partial y}-\frac{\partial A_y}{\partial z})+v_z\frac{\partial A_z}{\partial z}-v_z\frac{\partial A_z}{\partial z}\right)$$

In the top line collect all the terms that have $\frac{\partial}{\partial x}$ in one group and all the terms that have A_x in an other group so that that line becomes

$$\tilde{x}\left(v_y\,\frac{\partial A_y}{\partial x}+v_z\frac{\partial A_z}{\partial x}+v_x\frac{\partial A_x}{\partial x}\;-v_y\,\frac{\partial A_x}{\partial y}-v_z\,\frac{\partial A_x}{\partial z}-v_x\frac{\partial A_x}{\partial x}\right)=$$

$$\tilde{x}\left(\vec{v}\cdot\frac{\partial}{\partial x}\vec{A}-\vec{v}\cdot\nabla A_x\right)$$

Where the terms are interpreted as follows:

$$\vec{v}\cdot\frac{\partial}{\partial x}\vec{A}=v_x\frac{\partial A_x}{\partial x}+v_y\frac{\partial A_y}{\partial x}+v_z\frac{\partial A_z}{\partial x}$$

$$\vec{v}\cdot\nabla A_x=v_x\frac{\partial A_x}{\partial x}+v_y\frac{\partial A_x}{\partial y}+v_z\frac{\partial A_x}{\partial z}$$

Inserting the other two terms one finds

$$\vec{\mathrm{v}} \times \vec{\mathrm{B}} = \vec{\mathrm{v}} \times \vec{\nabla} \times \vec{\mathrm{A}} = \nabla(\vec{\mathrm{v}} \cdot \vec{\mathrm{A}}) - (\vec{\mathrm{v}} \cdot \nabla)\vec{\mathrm{A}} \tag{2}$$

This is the theorem we need below to build magnetic forces into the equations of motion of a charged particle.

Now we can calculate the rate of change of velocity of a charged electron under the influence of an magnetic field.

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{m}\vec{\mathrm{v}} = \mathrm{q}(\vec{\mathcal{E}} + \vec{\mathrm{v}} \times \vec{\mathrm{B}}) \tag{3}$$

Making the usual substitutions

$$\vec{\mathrm{B}} = \nabla \times \vec{\mathrm{A}} \qquad\qquad \vec{\mathcal{E}} = -\nabla\phi - \frac{\partial \vec{\mathrm{A}}}{\partial \mathrm{t}}$$

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{m}\vec{\mathrm{v}} = \mathrm{q}(-\nabla\phi - \frac{\partial \vec{\mathrm{A}}}{\partial \mathrm{t}} + \vec{\mathrm{v}} \times \nabla \times \vec{\mathrm{A}})$$

From Eqs. 3 and 2

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{m}\vec{\mathrm{v}} = \mathrm{q}\left(-\nabla\phi - \frac{\partial \vec{\mathrm{A}}}{\partial \mathrm{t}} + \nabla(\vec{\mathrm{v}} \cdot \ \vec{\mathrm{A}}) - (\vec{\mathrm{v}} \cdot \nabla)\vec{\mathrm{A}}\right)$$

Now

$$\frac{\mathrm{d}\vec{\mathrm{A}}}{\mathrm{dt}} = \frac{\partial \vec{\mathrm{A}}}{\partial \mathrm{t}} + \vec{\mathrm{v}} \cdot \ \nabla\vec{\mathrm{A}}$$

The first term is due to $\vec{\mathrm{A}}$ changing at a point and the second term to the motion of the charged particle.

Two terms can be moved to the left as the total time derivative of $\vec{A}$.

$$\frac{d}{dt}(m\vec{v} + q\vec{A}) = -q\nabla(\phi - v \cdot A)$$

This yields the rate of change of momentum of a charged body in terms of electric and magnetic fields where the magnetic field is in terms of the magnetic vector potential.

The right hand side is still not as simple as one would like it to be. In the section *The well according to Schrödiger* page 38 we assume a magnetic field having only a z component and derived from a vector potential

$$A_x = \frac{-yB_0}{2}$$

$$A_y = \frac{xB_0}{2}$$

In our specfic case, $\nabla(\vec{v} \cdot \vec{A})$, can be simplified by noting that

$$\vec{v} \cdot \vec{A} = v_x A_x + v_y A_y = v_0 B_0 \left((\frac{-y}{r})(\frac{-y}{2}) + \frac{x}{r}\frac{x}{2} \right) = v_0 B_0 \frac{r}{2}$$

$$\nabla(\vec{v} \cdot \vec{A}) = \hat{r}\frac{v_0 B_0}{2}$$

Thus $\nabla(\vec{v} \cdot \vec{A})$ has only an r component. This of course is what one expects from $\vec{v} \times \vec{B}$. Its effect would be to change the radius of rotation. Thus for our purposes here the rule by which momentum is replaced by momentum plus $q\vec{A}$ is simply

$$\vec{p} - - - > \vec{p} + q\vec{A}.$$

Conventionally this replacement is indicated by the notation $\vec{p}---> \vec{p}+q\vec{A}$. This is a recipe for confusion for which we apologize. One really ought to use a separate symbol for $\vec{p} + q\vec{A}$.[4]

Quantum Equation of Motion

In the familer version of quantum mechanics, one finds the average value of an observable from

$$\mathrm{Av(O)} = \int_{\mathrm{Allspace}} \psi^* \mathbf{O_{op}} \psi \quad \mathrm{dxdydz} \tag{4}$$

[5] where we are assuming that ψ has been normalized so that

$$\int_{\mathrm{Allspace}} \psi^* \psi \quad \mathrm{dxdydz} = 1$$

and $\mathbf{O_{op}}$ is the operator associated with the observable O.

Sometimes we need the rate of change of the average value of the operator. For example if the operator is the x directed momentum, we might want to know the rate of change of momentum.

We begin by differentiating Eq. 4. We are assuming here that time does not appear inside of $\mathbf{O_{op}}$.

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{Av(O)} = \int_{\mathrm{Allspace}} \psi^* \mathbf{O_{op}} \frac{\partial \psi}{\partial \mathrm{t}} + \frac{\partial \psi^*}{\partial \mathrm{t}} \mathbf{O_{op}} \psi \quad \mathrm{dxdydz}$$

We replace the time derivatives by means of Schrödiger's equation and its complex conjugate transpose, recalling that the transpose of a product is the

[4] Even worse, many texts call $\vec{p} + q\vec{A}$ a momentum. At best it should be called a generalized momentum. Using the same word for two different quantities seems to this author to be a bad practice but a common one.

[5] This need not be in rectangular coordinates.

product of the transposes in reverse order and that the complex conjugate replaces i by $-$i. Thus ψ complex conjugate becomes ψ^*.

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial \mathrm{t}} = \mathcal{E}\psi = \mathrm{H}\psi \quad \text{So that } \frac{\partial\psi}{\partial \mathrm{t}} = -\frac{\mathrm{i}}{\hbar}\mathrm{H}\psi \tag{5}$$

$$-\mathrm{i}\hbar\frac{\partial\psi^*}{\partial \mathrm{t}} = \mathcal{E}\psi^* = \psi^*\mathrm{H}^\dagger \quad \text{and } \frac{\partial\psi^*}{\partial \mathrm{t}} = \frac{\mathrm{i}}{\hbar}\psi^*\mathrm{H}^\dagger. \tag{6}$$

Then we replace $\mathrm{H}^\dagger$ by H. This last step is allowed because all quantum mechanical operators have the property that they are equal to their own complex congugate transpose. This is required to insure that average value calculations yield real numbers. [6]

Thus

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{Av(O)} = \int_{\mathrm{Allspace}} \psi^*(\mathbf{O_{op}}\mathrm{H} - \mathrm{H}\mathbf{O_{op}})\psi \quad \mathrm{dxdydz} \tag{7}$$

and the operator for the time derivative of O is

$$\dot{\mathbf{O}}_{\mathbf{op}} = -\frac{\mathbf{i}}{\hbar}(\mathbf{O_{op}}\mathrm{H} - \mathrm{H}\mathbf{O_{op}})\mathrm{..} \tag{8}$$

This last step may appear questionable since in general, knowing the value of an integral does not enable one to determine the integrand. However, Eq. 7, must hold for arbitrary ψ and this can only occur when the operator is as given above in Eq. 8. For example, let ψ to be a single narrow spike, located at a point where it is alleged that Eq. 8 is not true. Then Eq. 7 would also not be true.

The analogous operation for a two component ψ and a 2×2 operator would be

$$\mathrm{Av(O)} = \int_{\mathrm{Allspace}} < \psi|\mathbf{O_{op}}|\psi > \quad \mathrm{dxdydz}$$

[6] See any elementary text.

where $|\psi>$ is a two component ψ_α, ψ_β column vector, and $<\psi|$ is a two component row vector with each component the complex conjugate of the corresponding component of $|\psi>$. $\mathbf{O_{op}}$ is a two by two matrix.

Once again the condition applys that

$$\int_{\text{Allspace}} <\psi||\psi> \quad \text{dxdydz} = 1$$

The result of Av(O) is a single number.

The rate of change of the average of an observable is

$$\frac{\text{d}}{\text{dt}}\text{Av}(\mathbf{O}) = \int_{\text{Allspace}} <\psi|\dot{\mathbf{O}}_{\mathbf{op}}|\psi> \quad \text{dxdydz} \tag{9}$$

where the operator $\dot{\mathbf{O}}_{\mathbf{op}}$ is given by the previous logic as

$$\dot{\mathbf{O}}_{\mathbf{op}} = -\frac{\text{i}}{\hbar}(\mathbf{O_{op}H - HO_{op}}) \tag{10}$$

and $\mathbf{H}$ is the appropriate Hamiltonian.[7] In the case of the section *Angular Momentum of the Electron* page 50 $\mathbf{H}$ is the square bracket of Eq. 49 solved for $\mathcal{E}$

$$\mathbf{H} = [\text{V}\mathbf{I} \pm (\text{c}\sigma_\text{x}\tilde{\text{p}_\text{x}} + \text{c}\sigma_\text{y}\tilde{\text{p}_\text{y}} + \sigma_\text{z}\text{m}_0\text{c}^2)] \tag{11}$$

which is then simplified by noting that a number of terms in Eq.10 will cancel in the subtraction.

Two Easy Theorems.

Both of these are easily demonstrated by trial since there are only a small number of cases for each.

$\text{p}_\text{a}\sigma_\text{b} = \sigma_\text{b}\text{p}_\text{a}$ for any a,b or c

[7] This result is sometimes called the equation of motion.

Example:

$$\frac{\partial}{\partial \mathrm{x}}\ \sigma_{\mathrm{y}} = \frac{\partial}{\partial \mathrm{x}}\ \begin{vmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{vmatrix} = \begin{vmatrix} 0 & -\mathrm{i}\frac{\partial}{\partial \mathrm{x}} \\ \mathrm{i}\frac{\partial}{\partial \mathrm{x}} & 0 \end{vmatrix} = \begin{vmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{vmatrix}\ \frac{\partial}{\partial \mathrm{x}}$$

The second theorem is:

$$\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} = \mathrm{i}\sigma_{\mathrm{z}} \tag{12}$$
$$\sigma_{\mathrm{y}}\sigma_{\mathrm{z}} = \mathrm{i}\sigma_{\mathrm{x}} \tag{13}$$
$$\sigma_{\mathrm{z}}\sigma_{\mathrm{x}} = \mathrm{i}\sigma_{\mathrm{y}} \tag{14}$$

Most easily shown by trial: Example :

$$\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}\ \begin{vmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{vmatrix} = \mathrm{i}\ \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}$$

Converting to Polar Coordinates.

For

$$\mathrm{x} = \mathrm{r}\cos\theta, \qquad \mathrm{y} = \mathrm{r}\sin\theta$$

$$\frac{\partial}{\partial \mathrm{x}} = \frac{\partial \mathrm{r}}{\partial \mathrm{x}}\frac{\partial}{\partial \mathrm{r}} + \frac{\partial \theta}{\partial \mathrm{x}}\frac{\partial}{\partial \theta} = \cos\theta\frac{\partial}{\partial \mathrm{r}} - \frac{\sin\theta}{\mathrm{r}}\frac{\partial}{\partial \theta}$$

$$\frac{\partial}{\partial \mathrm{y}} = \frac{\partial \mathrm{r}}{\partial \mathrm{y}}\frac{\partial}{\partial \mathrm{r}} + \frac{\partial \theta}{\partial \mathrm{y}}\frac{\partial}{\partial \theta} = \sin\theta\frac{\partial}{\partial \mathrm{r}} + \frac{\cos\theta}{\mathrm{r}}\frac{\partial}{\partial \theta}$$

$$\frac{\partial}{\partial \mathrm{x}} + \mathrm{i}\frac{\partial}{\partial \mathrm{y}} = \mathrm{e}^{\mathrm{i}\theta}\left(\frac{\partial}{\partial \mathrm{r}} + \mathrm{i}\frac{1}{\mathrm{r}}\frac{\partial}{\partial \theta}\right) \qquad \frac{\partial}{\partial \mathrm{x}} - \mathrm{i}\frac{\partial}{\partial \mathrm{y}} = \mathrm{e}^{-\mathrm{i}\theta}\left(\frac{\partial}{\partial \mathrm{r}} - \mathrm{i}\frac{1}{\mathrm{r}}\frac{\partial}{\partial \theta}\right)$$

So that

$$\tilde{p_x} + i\tilde{p_y} = \frac{\hbar}{i}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) = e^{i\theta}\frac{\hbar}{i}\left(\frac{\partial}{\partial r} + i\frac{1}{r}\frac{\partial}{\partial \theta}\right) \tag{15}$$

$$\tilde{p_x} - i\tilde{p_y} = \frac{\hbar}{i}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) = e^{-i\theta}\frac{\hbar}{i}\left(\frac{\partial}{\partial r} - i\frac{1}{r}\frac{\partial}{\partial \theta}\right)$$

We are going to be assuming circular orbits. Thus momentum in the radial direction must be zero. Since

$$p_r = \frac{\hbar}{i}\frac{\partial}{\partial r},$$

we can set

$$\frac{\partial}{\partial r} = 0.$$

So that along the circle,

$$\tilde{p_x} + i\tilde{p_y} = \hbar e^{+i\theta}\frac{1}{r}\frac{\partial}{\partial \theta} \qquad \tilde{p_x} - i\tilde{p_y} = -\hbar e^{-i\theta}\frac{1}{r}\frac{\partial}{\partial \theta} \tag{16}$$

Which yields a simple form for the following quantities that we will need:

$$(\tilde{p_x} + i\tilde{p_y})(\tilde{p_x} - i\tilde{p_y}) = \frac{-\hbar^2}{r^2}e^{i\theta}\frac{\partial}{\partial \theta}\left(e^{-i\theta}\frac{\partial}{\partial \theta}\right) = \frac{-\hbar^2}{r^2}\left(\frac{\partial^2}{\partial \theta^2} - i\frac{\partial}{\partial \theta}\right)$$
$$(\tilde{p_x} - i\tilde{p_y})(\tilde{p_x} + i\tilde{p_y}) = \frac{-\hbar^2}{r^2}e^{-i\theta}\frac{\partial}{\partial \theta}\left(e^{i\theta}\frac{\partial}{\partial \theta}\right) = \frac{-\hbar^2}{r^2}\left(\frac{\partial^2}{\partial \theta^2} + i\frac{\partial}{\partial \theta}\right).$$

In the section *Angular Momentum of the Electron* page 50 we encounter the term

$$H' = \frac{c\hbar}{i}\left(\sigma_x\frac{\partial}{\partial x} + \sigma_y\frac{\partial}{\partial y}\right) = \tag{17}$$

The parenthesis evaluates as follows:

$$\begin{vmatrix} 0 & -\frac{\sin(\theta)}{r}\frac{\partial}{\partial\theta} \\ -\frac{\sin(\theta)}{r}\frac{\partial}{\partial\theta} & 0 \end{vmatrix} + \begin{vmatrix} 0 & -\frac{i\cos(\theta)}{r}\frac{\partial}{\partial\theta} \\ \frac{i\cos(\theta)}{r}\frac{\partial}{\partial\theta} & 0 \end{vmatrix} = \begin{vmatrix} 0 & -ie^{-i\theta}\frac{1}{r}\frac{\partial}{\partial\theta} \\ ie^{i\theta}\frac{1}{r}\frac{\partial}{\partial\theta} & 0 \end{vmatrix}$$

So that

$$H' = \begin{vmatrix} 0 & -e^{-i\theta} \\ e^{i\theta} & 0 \end{vmatrix} \frac{c\hbar}{r}\frac{\partial}{\partial\theta}$$

The Need For Spin

One of the central puzzles of Physics in the early 20th century was the existence of discontinuous but sharply defined spectral lines, emitted by electrically or thermally excited atoms.

In 1913 Niels Bohr published a remarkably simple explanation that came very close to accounting for the spectrum of the hydrogen atom. (The Bohr atom is described in the section on Matrix Mechanics.) At least a very good gross description of the spectrum, but his model could not account for the fact that some of the spectral lines were actually closely spaced pairs of lines: The so called fine structure.[8]

It was then suggested by Sommerfeld (1916) that this could be explained by assuming that there existed elliptical orbits as well as the circular ones and that because of the relativistic change in mass resulting from of the speed differences in different parts of the orbit these would result in slightly shifted energy levels and therefore different wavelengths of about the correct separation.

This explanation was accepted for a couple of years but it ran into difficulty when applied to other univalent atoms such as alkali metals. Their orbits would have been too close to circular to produce enough relativistic correction. [9]

It was then suggested by Goudsmit and Uhlenbeck (1925) that one could explain the necessary split energy levels by assigning a magnetic dipole to the electron which would interact with the *electric field* field of the nucleus, part of which appears as a magnetic field to the electron because of the electrons motion.[10]

This explanation was right on the mark. Nevertheless, it too, ran into apparent paradoxes. It not only required that the electrons possess a mag-

[8]One such pair is easily observed in the bright yellow spectral lines of Sodium. These are of wavelengths 588.99 and 589.59 nanometers or 0.1 percent apart.

[9]Leighton p 141. Also page 142 for good explanation of Thomas precession.

[10]Looked at from the electron, one can say that the nuclear charge appears to be rotating around the electron and acts as if it were a circle of current.

netic dipole but that this dipole always be aligned either parallel to, or anti parallel to the plane of motion around the nucleus. Not unlike the orbits of most of the planets but difficult to understand in the context of atoms.

Stern-Gerlach

It had been assumed that an orbiting electron would act like a circulating current and as a result an atom would have a magnetic moment perpendicular to the plane of rotation. Since there was no reason to expect that the planes of rotation had any particular orientation, one would expect the magnetic moments to have random orientations.

However when Stern and Gerlach (1922) measured this orbital magnetic moment in silver atoms they found the astonishing result that it was always aligned as above, parallel to, or anti parallel to, the gradient of the magnetic field used to measure it. If this were true, then it was also possible that electrons had a similar property.

The Stern and Gerlach experiment was similar to the one outlined below and is well described in almost any advanced Physics text.

We are going to use the Stern and Gerlach experiment in a thought experiment as a means of introducing the Pauli matrices. This will make them less ethereal than does the usual approach. Then we can use them to show that, in fact, electrons do have a magnetic moment not associated with their orbital motion.

For our purposes we will describe the experiment as if it were done with isolated electrons since this is the topic of interest here. [11]

The drawing in fig. 1 shows a magnet made with one sharp pole piece and one flat pole piece so that it produces a magnetic field gradient in the z direction. When a beam of electrons is sent between the pole pieces parallel to the x axis, the beam is split into two parts. Approximately half of the

[11] See Liboff for a discussion of the limitations due to the uncertainty principle that make this simple version difficult.

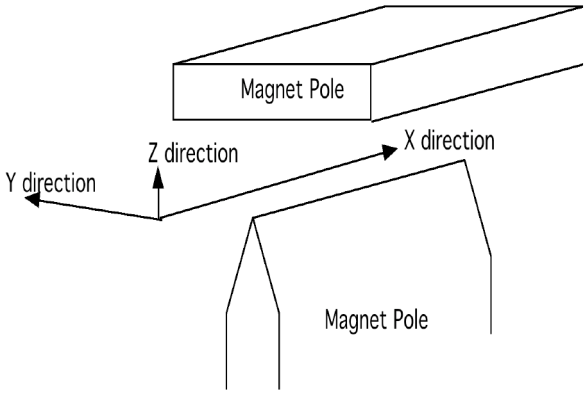

Figure 1: Stern-Gerlach experiment

electrons are deflected downward and the other half deflected upward. [12]

This effect does not occur in a uniform field but depends upon the field gradient in the z direction. A uniform field would produce a net force on a magnetic monopole just as a uniform electric field produces a force on a single charge. Since no such effect occurs in a uniform magnetic field, the observed effect must be due to the electron having equal and opposite magnetic poles, a dipole.

The key and surprising aspect of this experiment is that the beam is sharply divided. Each electron gains a velocity component either full up or full down. Intermediate values never occur. At first, the only reasonable explanation of this seems to be that the electron's magnetic dipole is oriented either parallel to the z axis or anti parallel to it but never at any other angle.

This is, to say the least, puzzling if no prior attempt has been made to sort or select the electrons to have their dipoles in one of the two directions. But the problem is much deeper than that. If the magnets are rotated around the x axis so that the field gradient is in the y direction instead of the z direction, one finds the electron beam is now split into left or right beams instead of up and down beams. Thus the possibility that the effect is due to

[12] We are neglecting curvature of the entire beam in the y direction which, by a small change in the experiment, can be made negligible. We also neglect any effect of a y component of B which would be zero along the axis.

a prior orientation of the electron's dipole moment is excluded.

In the macroscopic world, the energy stored by a magnetic dipole in a magnetic field is given by the dot product of the magnetic moment $\vec{\mu}$ with the applied field $\vec{\mathrm{B}}$. (The product of the magnitude of each multiplied by the cosine of the angle between them.)

Assuming that the angle remains constant, if the electron moves through a field gradient, a distance $\Delta\vec{\mathrm{r}}$ to where the magnetic field is smaller, the stored energy $\mathcal{E}$ is reduced by

$$\Delta\mathcal{E} = -\Delta\vec{\mathrm{r}} \cdot \nabla(\vec{\mu} \cdot \vec{\mathrm{B}})$$

which produces a force

$$\vec{\mathrm{F}} = -\nabla(\vec{\mu} \cdot \vec{\mathrm{B}})$$

One might visualize the magnetic dipole as being generated by the current due to rotating charge in the electron except that this raises the issue of how the charge knows which way it should be rotating before it encounters the experimental apparatus. Should it be rotating around the z axis to account for an up down split or around the y axis to account for a left right split. Perhaps it is rotating about both axis! A bit of reflection will convince the reader that this won't do the job either. It would be equivalent to rotation around a diagonal axis.

In short, the magnetic dipole associated with an electron is not due to charge rotation in the sense of a spinning baseball. It is something new that has the property of having a dipole moment in *any direction* such that when one attempts to measure it by a magnetic field gradient, yields *a force in the direction of the magnetic field gradient or in the reverse direction and is such that the magnitude of the force does not depend on the orientation of the field gradient with respect to the dipole orientation.*

Pauli's Matrices

No simple three dimensional vector for the magnetic moment $\vec{\mu}$ has this property.

Not only must the correct interaction not depend upon the angle between $\vec{\mu}$ and $\vec{\mathrm{B}}$, but it must also have the property that for some electrons *it has its sign reversed.* It is analogous to suggesting that the same end of a compass needle can point either toward or away from the north pole.

Thus we can have a force

$$\vec{\mathrm{F}} = -\nabla(\vec{\mu}\cdot\vec{\mathrm{B}}) \quad \text{OR} \quad \vec{\mathrm{F}} = +\nabla(\vec{\mu}\cdot\vec{\mathrm{B}}) \tag{18}$$

This peculiar property must be due to the dot product above not really being a conventional vector dot product. To indicate that we are dealing with some other kind of interaction between μ and B we introduce the symbol $\underline{(\mu\mathrm{B})}$.

The components of $\vec{\mathrm{B}}$ here are classical quantities. Therefore as a first trial we should leave them undisturbed. The trick is going to be to redefine $\vec{\mu}$ so that its components do not form a conventional vector. One way of doing this is to define them as multi dimensional quantities, specifically as 2 by 2 matrices. The $\vec{\mu}$ is then made up of three 2 by 2 matrices each multiplied by a unit vector. This new kind of quantity will have some of the properties of a vector but not all of them and in particular when dotted into another vector will have exactly the properties we need.

Therefore we replace $\vec{\mu}$ by

$$\vec{\mu} - - > \mu(\hat{\mathrm{x}}\sigma_{\mathbf{x}} + \hat{\mathrm{y}}\sigma_{\mathbf{y}} + \hat{\mathrm{z}}\sigma_{\mathbf{z}}) \tag{19}$$

where $\hat{\mathrm{x}}$ indicates a unit vector in the x direction.

Now for reasons given below, we are going to define the σ as:

$$\sigma_{\mathbf{x}} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \qquad \sigma_{\mathbf{y}} = \begin{vmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{vmatrix}, \qquad \sigma_{\mathbf{z}} = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}. \tag{20}$$

If one has an equation having one root and needs to replace it with one having two roots, the natural thing to do is to square it and square root the result. This may seem an artificial maneuver unless the squared intermediate allows some operation that cannot be performed in the root form. In this case we calculate the square of the energy due to the "dipole" in a magnetic field where the term "dipole" is in quotation marks to indicate that it is not the usual dipole. The square of the energy is given by

$$\mathcal{E}^2 = \underline{(\mu \mathrm{B})(\mu \mathrm{B})} \tag{21}$$

where if we use the σ defined in Eq. 20 we can make the cross product terms in B^2 vanish as follows:

$$\begin{aligned}
&\mu^2(\mathrm{B_x}\sigma_{\mathbf{x}} + \mathrm{B_y}\sigma_{\mathbf{y}} + \mathrm{B_z}\sigma_{\mathbf{z}})(\mathrm{B_x}\sigma_{\mathbf{x}} + \mathrm{B_y}\sigma_{\mathbf{y}} + \mathrm{B_z}\sigma_{\mathbf{z}}) = \\
&\\
&\mu^2(\mathrm{B_x}\mathrm{B_x}\sigma_{\mathbf{x}}\sigma_{\mathbf{x}} + \mathrm{B_x}\mathrm{B_y}\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} + \mathrm{B_x}\mathrm{B_z}\sigma_{\mathbf{x}}\sigma_{\mathbf{z}} + \\
&\mathrm{B_y}\mathrm{B_x}\sigma_{\mathbf{y}}\sigma_{\mathbf{x}} + \mathrm{B_y}\mathrm{B_y}\sigma_{\mathbf{y}}\sigma_{\mathbf{y}} + \mathrm{B_y}\mathrm{B_z}\sigma_{\mathbf{y}}\sigma_{\mathbf{z}} + \\
&\mathrm{B_z}\mathrm{B_x}\sigma_{\mathbf{z}}\sigma_{\mathbf{x}} + \mathrm{B_z}\mathrm{B_y}\sigma_{\mathbf{z}}\sigma_{\mathbf{y}} + \mathrm{B_z}\mathrm{B_z}\sigma_{\mathbf{z}}\sigma_{\mathbf{z}}) = \\
&\\
&\mu^2(\mathrm{B_x}\mathrm{B_x} + \mathrm{B_y}\mathrm{B_y} + \mathrm{B_z}\mathrm{B_z})\mathbf{I} = \\
&\mu^2\mathrm{B}^2 \quad \text{See below.}
\end{aligned} \tag{22}$$

To make the cross product terms cancel, we require that the product two different σ reverse sign when the order of multiplication is reversed. This is of course possible with matrices but not with scaler numbers. Appropriate matrices would have to have the following proprieties.

$$\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} = -\sigma_{\mathbf{y}}\sigma_{\mathbf{x}}, \qquad \sigma_{\mathbf{x}}\sigma_{\mathbf{z}} = -\sigma_{\mathbf{z}}\sigma_{\mathbf{x}}, \qquad \sigma_{\mathbf{y}}\sigma_{\mathbf{z}} = -\sigma_{\mathbf{z}}\sigma_{\mathbf{y}}. \tag{23}$$

Then to insure that each component of B is equally weighted we require that the square of each matrix (The matrix multiplied by its transposed complex conjugate.) be the unit matrix $\mathbf{I}$.

$$\sigma_{\mathbf{x}}\sigma_{\mathbf{x}} = \mathbf{I}, \qquad \sigma_{\mathbf{y}}\sigma_{\mathbf{y}} = \mathbf{I}, \qquad \sigma_{\mathbf{z}}\sigma_{\mathbf{z}} = \mathbf{I}. \tag{24}$$

This is why the σ were chosen as indicated in Eq.20. The algebra leading from Eq.20 to the above is given in *Pauli Matrices* in the *Mathematics of Spin.* It is quite straightforward, just out of place here. [13]

Thus Eq. 21 becomes as indicated in Eq. 22.

$$\mathcal{E}^2 = \mu^2 \mathrm{B}^2$$

The energy now has two possible values:

$$\mathcal{E} = \pm \mu \mathrm{B}$$

$$\vec{\mathrm{F}} = \pm \mu \bigtriangledown \mathrm{B} \quad \text{where} \quad \mathrm{B} = \sqrt{\mathrm{B}_\mathrm{x}^2 + \mathrm{B}_\mathrm{y}^2 + \mathrm{B}_\mathrm{z}^2} \tag{25}$$

Force is always in the direction of decreasing field energy.[14] If the positive sign is used above, that is in the direction of decreasing B. However if the minus sign is used above it is in the direction of increasing B. Thus the force can be either in the direction of the gradient of B or the reverse but in either case is independent of the relative direction of the electron's "spin" to that of the applied magnetic gradient.

The reader may wonder why it is necessary to define μ in all three directions. Why does it not suffice to define only x and z. The reason is that one needs to be able to rotate the coordinates around the z axis so that for a 90 degree rotation, for example, the same experiment is described as if the beam were along the y axis. Similarly a 45 degree rotation must represent the beam as lying along the diagonal. For this to work, μ must contain all three directions.

[13] The matrices could have been defined in a number of equivalent ways. This particular set of matrices was chosen here and probably by Pauli because, except for interchanging within the set, it is the set of three in which each one has two zero terms.

[14] If a particle is accelerated by a field, its increase in kinetic energy is at the cost of decreasing field energy. If the force is $\vec{\mathrm{F}}$, the work it does on the particle in moving from x to $\mathrm{x} + \Delta\mathrm{x}$ is $\vec{\mathrm{F}} \cdot \Delta\mathrm{x}$. From this it is easily seen, component by component that

$$\vec{\mathrm{F}} = -\nabla \cdot \mathcal{E}$$

There is a simpler way to think about this result. In the section *Value of a Pauli Matrix* in *Mathematics of Spin* we show how one can assign a numerical value to a Pauli matrix. This value is represented by ς and is equal to ± 1.

Returning to Eq. 21 we can write the energy as

$$\mathcal{E} = \mu(\mathrm{B_x}\varsigma_\mathrm{x} + \mathrm{B_y}\varsigma_\mathrm{y} + \mathrm{B_z}\varsigma_\mathrm{z}) \tag{26}$$

.

Pauli Properties

The simple requirements that define the Pauli matrices in the previous section, result in an important and possibly unexpected property.

In the previous section we defined a vector $\vec{\mu}$ that has the property we need to account for the splitting of an electron beam in a Stern-Gerlach experiment.

This quantity $\vec{\mu}$ is a real vector. It is peculiar in that its coefficients are matrices rather than simple numbers but it is still a vector. Therefore we can find its projection in any direction by taking the dot product with a unit vector in that direction. For example the z projection (up-down) is

$$\boldsymbol{\mu_z} = \hat{\mathrm{z}} \cdot \vec{\mu} = \hat{\mathrm{z}} \cdot \mu(\hat{\mathrm{x}}\sigma_\mathrm{x} + \hat{\mathrm{y}}\sigma_\mathrm{y} + \hat{\mathrm{z}}\sigma_\mathrm{z}) = \mu\sigma_\mathrm{z}$$

This is not a number, it is a matrix. A measurement of spin yields a single number. Therefore we need to assign a numerical value to each matrix, which we will call ς, and who's value we find by squaring the matrix

$$\sigma_\mathrm{z}\sigma_\mathrm{z} = \varsigma_\mathrm{z}^2\,\mathbf{I} = \varsigma_\mathrm{z}^2 \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}.$$

If the numerical value ς_z is to make sense, it obviously has to be ± 1. [15] We have chosen the σ_z matrix to illustrate this point but obviously a similar result applies to σ_x and σ_y.[16]

Now we have a number,

$$\mu_\mathrm{z} = \mu\varsigma_\mathrm{z}$$

It is this quantity that is acted upon by the z component of the magnetic field gradient. It can be μ or $-\mu$ which accounts for the splitting of the electron beam.

[15] See *Value of a Pauli Matrix* in *Mathematics of Spin* at the beginning of this chapter.

[16] Or to certain linear combinations of these that correspond to rotations of the coordinate system.

Now which is it, 1 or -1? The Stern-Gerlach experiment tells us. Approximately half of the electrons are deflected up and half are deflected down. Therefore the average value of ς_z over a number of electrons, is much less than one. For a particular electron once it is measured, it is known to be 1 or -1 but the absolute value of the average is much less than that of a single electron.

Empirically it is found that in a beam that *has not been previously sorted or aligned*, about half the electrons go up and half go down. We could express this as being the result of

$$\vec{\mu} = \mu(\hat{\mathrm{x}}\varsigma_{\mathrm{x}} + \hat{\mathrm{y}}\varsigma_{\mathrm{y}} + \hat{\mathrm{z}}\varsigma_{\mathrm{z}})$$

Suppose we had rotated the magnet so that the field gradient is in the y direction. Then according to the above we would get the beam split in the left-right direction and that is what is observed.

Note the italicized line above. Suppose the beam has been previously sorted by running it through two such Stern-Gerlach magnets both of which have the magnetic gradient in the z direction. We find the first magnet causes a 50-50 up-down split but if only the z-up beam enters the second magnet, although a further deflection occurs, it is all in the up direction. It appears as if the first magnetic gradient has changed the ς_z in the last term of the previous equation which can be ± 1 to $+1$ only. We do not know what it has done to the other terms but we do know this much.

$$\text{For a z sorted beam,} \quad \vec{\mu} = \mu(\hat{\mathrm{x}}\varsigma_{\mathrm{x}} + \hat{\mathrm{y}}\varsigma_{\mathrm{y}} + 1\hat{\mathrm{z}}) \tag{27}$$

We can find out what happens to the y term by rotating the second magnet so that its gradient is in the y direction. Not too surprisingly we find that the beam that has been sorted for z-up is still random in y. The beam is split left-right.

Now here is where weirdness begins! If the y-right beam is now sorted a third time in the z direction, one finds that the electrons having been sorted in y in the second magnet are once again random in the z direction. They have "forgotten" that they were already sorted and selected to be all z-up coming out of the first magnet.

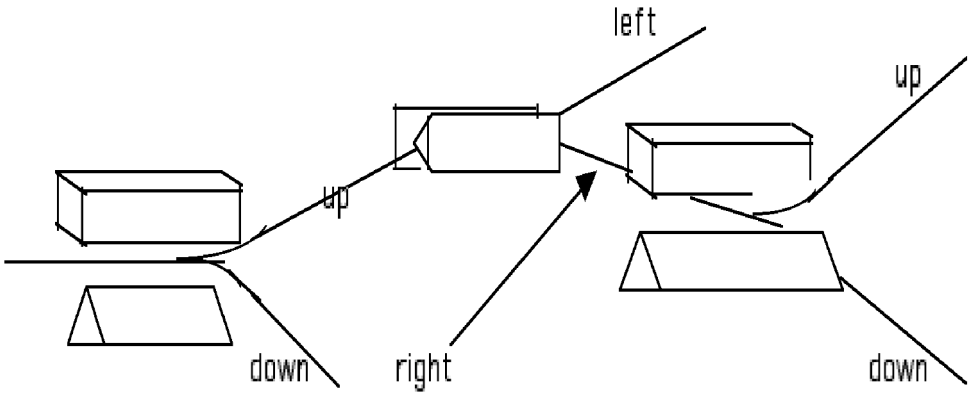

Figure 2: Triple Stern-Gerlach

It would appear that nothing we have said before can account for this but in fact the weirdness is already built into the Pauli matrices. Here is how it occurs.

Suppose we want to know the y projection of $\vec{\mu}$. We find it by taking the dot product of $\vec{\mu}$ with a unit vector in the y direction. Thus

$$\mu_y = \hat{y} \cdot \vec{\mu} = \mu \varsigma_y \tag{28}$$

where for any one electron the value of ς_y is ± 1.

Now suppose we want the average of the y projections over a number of electrons. We measure the y projection over a number, N, of electrons and divide the sum by N. Averages are often indicated by a superscripted bar but it is a little easier to read the following if a new symbol is introduced which is the average of the projection in the y direction divided by μ.

Define

$$a_y = \bar{\varsigma_y} = \quad \text{average of y projections of } \mu \text{ divided by } \mu \tag{29}$$

The maximum value this can have would occur if all the ς_y were $+1$ and the minimum value would occur when they are all -1. Therefore a_y must lie in

the range

$$-1 \leq a_y \leq 1$$

with similar relations being true for the other two directions.

We can now define a vector made up of the averages of the projections divided by μ of the spin of many electrons in each of the three directions. Let

$$\vec{a} = \hat{x}a_x + \hat{y}a_y + \hat{z}a_z$$

Now calculate the square of the product $\underline{(\vec{\mu}\vec{a})}$ the same way that we calculated $\underline{(\vec{\mu}\vec{B})}$ in Eq. 22 recalling that we must use the Pauli matrix form of $\vec{\mu}$ so that the cross product terms will cancel.

$$\begin{aligned}
&\underline{(\vec{\mu}\vec{a})}^2 = \mu^2(a_x\sigma_{\mathbf{x}} + a_y\sigma_{\mathbf{y}} + a_z\sigma_{\mathbf{z}})(a_x\sigma_{\mathbf{x}} + a_y\sigma_{\mathbf{y}} + a_z\sigma_{\mathbf{z}}) = \\
&\mu^2(a_xa_x\sigma_{\mathbf{x}}\sigma_{\mathbf{x}} + a_xa_y\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} + a_xa_z\sigma_{\mathbf{x}}\sigma_{\mathbf{z}} + \\
&a_ya_x\sigma_{\mathbf{y}}\sigma_{\mathbf{x}} + a_ya_y\sigma_{\mathbf{y}}\sigma_{\mathbf{y}} + a_ya_z\sigma_{\mathbf{y}}\sigma_{\mathbf{z}} + \\
&a_za_x\sigma_{\mathbf{z}}\sigma_{\mathbf{x}} + a_za_y\sigma_{\mathbf{z}}\sigma_{\mathbf{y}} + a_za_z\sigma_{\mathbf{z}}\sigma_{\mathbf{z}}) = \\
&\mu^2(a_xa_x + a_ya_y + a_za_z)\mathbf{I} = \mu^2a^2\mathbf{I} = \mu^2a^2\mathbf{I}^2 \quad \text{since} \quad \mathbf{I} = \mathbf{I}^2
\end{aligned} \tag{30}$$

Comparing the top line and last line of Eq. 30 we have

$$(a_x\sigma_{\mathbf{x}} + a_y\sigma_{\mathbf{y}} + a_z\sigma_{\mathbf{z}}) = \pm a\mathbf{I} \tag{31}$$

In terms of the measurable quantities

$$(a_x\varsigma_{\mathbf{x}} + a_y\varsigma_{\mathbf{y}} + a_z\varsigma_{\mathbf{z}}) = \pm a \tag{32}$$

Each of the terms in the left bracket consists of an average "a" multiplied by a term of the cooresponding projection of $\vec{\mu}$ divided by μ. See Eq. 28.

Recall Eq. 29 reproduced below

$$a_y = \bar{\varsigma}_y = \quad \text{average of y projections of } \mu \text{ divided by } \mu$$

Individually each projection can have $\varsigma = \pm 1$ but if we average over many electrons so that each ς is replaced by its average, the corresponding a, we find that the left side of Eq. 32 is the sum of the squares of the a_s and that this sum is equal to $\pm$ the length of $\vec{a}$. Not the square of the length that one might expect.

Therefore we have the result that

$$(a_x^2 + a_y^2 + a_z^2) = \pm|a|$$

But it is also true that the sum of the squares of the components of any vector, the left term above, is the square of its length.

$$|a|^2 = |a|$$

so we are lead to the remarkable result that $|a| = 1$ Therefore the sum of the squares of the averages of the projections of $\vec{\mu}$ is μ^2.

$$(a_x^2 + a_y^2 + a_z^2)\mu^2 = \mu^2 \tag{33}$$

Now let us see how this explains the triple Stern-Gerlach experiment described earlier in this section. After passing through one Stern-Gerlach apparatus we concluded from Eq. 27 that

$$\text{For a z sorted beam} \quad \vec{\mu} = \mu(\varsigma_x\hat{x} + \varsigma_y\hat{y} + 1\hat{z})\mathbf{I}$$

Each ς can be ± 1 and we don't know for a single electron whether it is $+$ or $-$. *But we do know their averages* a *over a number of electrons!* In the sorted z-up beam, ς_z is always plus one so that its average a_z is one and its square is one. Therefore from Eq. 33 we have

$$a_x^2 + a_y^2 + 1 = 1$$

Both a_x and a_y must be zero! Therefore for z sorted electrons the average of x projections of their x spins is zero and the average of their y projections of spins is zero. The electrons are approximately evenly divided in these two directions.

Now we pass the z-up beam through a y directed gradient, a second Stern-Gerlach apparatus turned at right angles to the first. Then coming out of that we choose the right beam of a left-right directed split so that for that beam $a_y = \varsigma_y = +1$

By the same argument as that for z in the first sorting, in this beam, both a_x and a_z must be zero! Even though the beam had previously been selected for z-up, its average z spin projection is now zero! The second z sorting divides the electrons evenly between up and down. This is weird, particularly so when one considers that it has *resulted from an apparently unrelated property that the amplitude of the magnetic moment appear the same in all directions.*

The Source of Spin

So far we have not explained how the strange quantity spin arises. When it was first proposed as a solution to the spectral line split, (1925) the obvious suggestion was that it really was the result of spinning electric charge.

This explination at the time was rejected because it was argued that the surface velocity would have to be greater than the velocity of light.

We have rejected this argument on another basis. See page 25

A better explanation was given by Paul Dirac in [17] while pursuing a quite different objective.

The motive behind Dirac's work was that the Schrödiger equation, in spite of its spectacular success in predicting many of the properties of atoms, has what appears to be a serious flaw which is that it contains a second derivative

[17] Dirac, P. A. M. The quantum theory of the electron. Proc. R. Soc. (London) 1928 A 117 610-612

with respect to space coordinates but a first derivative with respect to time. Thus it can not be relativisticly correct, Lorentz covariant. It changes its form when viewed from a uniformly moving coordinate system.

A number of attempts were made to produce a similar wave equation replacing the first derivative with respect to time by a second derivative so that the result would be closer to Maxwell's equations for electromagnetic waves which are Lorentz covariant and were the origin of the Lorentz transformation.

Of particular interest was the Klein-Gordon equation [18] which has other applications but was not an adequate description of an electron.

The problem is that the second derivative with respect to time allows, in fact demands, negative electron energies. This appeared to be a fatal flaw. [19] Dirac's solution was, instead, to eliminate the second derivatives with respect to spatial coordinates to produce a Lorentz covariant result.

At first it may appear strange that one can arbitrarily get rid of second derivatives. However, it becomes less strange if one considers a familiar example such as the differential equation that defines the cosine function:

$$\frac{d^2}{dx^2}F(x) = -k^2F(x), \qquad F(0) = 1$$

By allowing a second dependent variable, this can be replaced by two first order differential equations

$$\frac{d}{dx}F(x) = kG(x)$$
$$\frac{d}{dx}G(x) = -kF(x)$$
$$F(0) = 1$$

Dirac's method is best illustrated by considering a specific, highly sim-

[18] See Leighton p662.

[19] The current view is that it is not, but the fatal flaw viewpoint led to Dirac's attack on the problem which is the topic of interest here.

plified, problem that can be solved easily, both with Schrödiger's equation, and by Dirac's method, and comparing the two solutions.

In his original work, [20] Dirac treats an electron in the usual three dimensions of space and one of time. This leads to 4 by 4 square matrices and a mass of algebra that makes his logic difficult to follow. In order to focus on the logic, we are going to use a cylindrical potential well and assume that the electrons orbit is a circle in the x,y plane instead of the usual product of radial dependence and angular dependence in three dimensions.

This will not necessarily lead to the description of any physically realizable situation and in fact will not always yield numerically correct results, but it does demonstrate all of the mathematical tools of Dirac's derivation and makes plausible his result that an electron has both an intrinsic magnetic moment and mechanical angular momentum.

Consider a circular infinitely deep potential well of radius r_0, whose inner potential is zero, and whose outside potential is infinite.

The Schrödiger equation to be solved is

$$\mathcal{E}\psi = \mathrm{H}(\mathrm{x}, \mathrm{y}, \mathrm{B})\psi \tag{34}$$

Where $\mathcal{E}$ is the total energy of a electron in the well including rest energy m_0c^2 in the Dirac solution but not, specifically, in the Schrödiger form. We will clarify this point later. H is the electron energy in terms of momentum and in this case including a magnetic field B whose effect on the electron we are soon to derive.

Although the problem is in circular geometry, it proves easier to solve in rectangular Cartesian coordinates and then convert to circular coordinates at the end.

[20] Proc. Roy. Soc. London A117, 610-624, 1928.

The Well According to Schrödiger

In the Schrödiger picture ψ is the usual quantum mechanical wave function which must satisfy the above equation inside the well with the condition that it be zero on the edges of the well: $r = r_0$.

In the Dirac version, ψ will be replaced by a two component function ψ_α and ψ_β both of which are zero outside the well.

Following the Schrödiger line we insert the following Hamiltonian into Eq. 34.

$$H = \frac{1}{2m_0}(p_x^2 + p_y^2) + V(x,y) + \text{Magnetic field-electron interaction.} \quad (35)$$

Explanation of the last term follows. The usual quantum mechanical procedure is to replace momentum and total energy operators as follows.

$$p_x --> \frac{\hbar}{i}\frac{\partial}{\partial x}, \qquad p_y --> \frac{\hbar}{i}\frac{\partial}{\partial y}, \qquad \mathcal{E} --> -\frac{\hbar}{i}\frac{\partial}{\partial t} \quad (36)$$

We will indicate replacement by tildes on the symbol so that

$$p_x --> \frac{\hbar}{i}\frac{\partial}{\partial x} = \tilde{p_x}$$

In order to include the effect of a magnetic field on the motion of a charge q we replace each component of the momentum operator $\tilde{p_x}$ for example, by $\tilde{p_x} + qA_x$ where A_x is the x component of the magnetic vector potential.

This step is by no means intuitive but derives from the familiar Lorentz force $\vec{F} = q(-\tilde{\nabla}\phi + \vec{v} \times \vec{B})$. The derivation is straightforward but tedious. It is worked out in detail in *Motion in a Magnetic Field* in the *Mathematics of Spin.*

Applying the replacements above to Eqs. 34 and 35 yields

$$i\hbar\frac{\partial\psi}{\partial t} = \mathcal{E}\psi = \frac{1}{2m_0}\left[(\tilde{p_x} + qA_x)^2 + (\tilde{p_y} + qA_y)^2\right]\psi + V\psi \quad (37)$$

[21] The $\tilde{p_x}$ are operators so that order of the terms must be preserved. $\tilde{p_x}A_x \neq A_x\tilde{p_x}$. Thus the following product has four distinct terms rather than three.

$$(\tilde{p_x} + qA_x)^2 = (\tilde{p_x} + qA_x)(\tilde{p_x} + qA_x) = -\hbar^2\frac{\partial^2}{\partial x^2} + qA_x\frac{\hbar}{i}\frac{\partial}{\partial x} + \frac{q\hbar}{i}\frac{\partial A_x}{\partial x} + q^2A_x^2$$

We get the right side of Eq. 37 by adding a similar y term to the above.

$$2\ m_0\ (\mathcal{E} - V)\psi = -\hbar^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\psi + \frac{q\hbar}{i}\left(A_x\frac{\partial}{\partial x} + A_y\frac{\partial}{\partial y}\right)\psi + \frac{q\hbar}{i}\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y}\right)\psi + q^2(A_x^2 + A_y^2)\psi \quad (38)$$

Now let us add a uniform magnetic field in the z direction, $B_0 = B_z$ produced by the curl of the vector potential $\vec{A}$

$$A_x = -\frac{yB_0}{2}, \qquad A_y = \frac{xB_0}{2}, \quad \text{Thus} \quad \nabla \times A = B_0. \quad (39)$$

With this replacement the third parenthesis in Eq. 38 above is easily seen to be zero while the second one can be replaced by

$$\left(A_x\frac{\partial}{\partial x} + A_y\frac{\partial}{\partial y}\right) = \frac{B_0}{2}\left(-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}\right) = \frac{B_0}{2}\frac{\partial}{\partial\theta} = \frac{i}{\hbar}\frac{B_0}{2}\tilde{p_\theta}$$

If in doubt about this step, see *Coordinate Conversion* in the *Mathematics of Spin* .

Now Eq. 38 divided by $2m_0$ becomes

$$(\mathcal{E} - V)\psi + \frac{\hbar^2}{2\ m_0}\nabla^2\psi - \frac{q^2}{2\ m_0}(A_x^2 + A_y^2)\psi - q\frac{B_0}{4m_0}\ \tilde{p_\theta}\psi = 0 \quad (40)$$

[21] q is the charge on the electron. Negative in this case.

where ∇^2 is the well known Laplacian operator except that in our case we have set the z dependent term to zero. For future reference we show it here both in Cartesian coordinates and in cylindrical polar coordinates. [22]

$$\nabla^2 = \left(\frac{\partial^2}{\partial \mathrm{x}^2} + \frac{\partial^2}{\partial \mathrm{y}^2}\right) = \frac{\partial^2}{\partial \mathrm{r}^2} + \frac{1}{\mathrm{r}}\frac{\partial}{\partial \mathrm{r}} + \frac{1}{\mathrm{r}^2}\frac{\partial^2}{\partial \theta^2} \tag{41}$$

In the last term in Eq. 40 above, the momentum operator $\tilde{\mathrm{p}}_\theta$ acts on ψ to yield an angular momentum p_θ. Thus the whole term.

$$\frac{\mathrm{qB}_0}{4\mathrm{m}_0}\tilde{\mathrm{p}}_\theta\psi$$

is an energy that for a fixed orbit, shifts $\mathcal{E}$ by

$$\frac{\mathrm{qB}_0\mathrm{p}_\theta}{4\mathrm{m}_0}$$

and is proportional to the angular momentum around the nucleus. It is the source of the ordinary Zeeman effect[23] except for a factor of two which is lost here due to neglecting the change in potential energy by assuming motion on at a fixed radius.[24]

[22] See *Mathematics of Spin*

[23] The shift of wavelength of a spectral line when a magnetic field is applied at the source of the light.

[24] Simple analogy: A harmonic oscillator in which half the energy exists as potential energy and half as kinetic.

Dirac's Solution

Now lets look at the same problem a la Dirac.

In order to insure compatibility with relativity, Dirac started from quantities that he knew behaved correctly under a Lorentz transformation and developed a wave equation from them. One such quantity is the rest mass m_0. The energy due to motion is then due to the difference between the apparent mass energy mc^2 and the conserved relativistic quantity m_0c^2.

Dirac's strategy was to produce a relativistically correct wave equation by replacing the Hamiltonian H (Total energy in terms of momentum.) in Schrödiger's equation by its relativistic equivalent:[25]

$$H = V(x, y, z) \pm \sqrt{m_0^2c^4 + c^2(p_x^2 + p_y^2 + p_z^2)} \tag{42}$$

He then applied the usual quantum mechanical procedure to convert this equation to an operator equation applied to the wave function:

$$p_x -- > \frac{\hbar}{i}\frac{\partial}{\partial x} = \tilde{p_x}, \qquad p_y -- > \frac{\hbar}{i}\frac{\partial}{\partial y} = \tilde{p_y}, \qquad p_z -- > \frac{\hbar}{i}\frac{\partial}{\partial z} = \tilde{p_z}. \tag{43}$$

[26] operating on the wave function ψ. With these substitutions, the Schrödiger equation becomes

(44)

$$\mathcal{E}\psi = H\psi = V\psi \pm \sqrt{m_0^2c^4 + \left(\frac{c\hbar}{i}\right)^2 \left(\left(\frac{\partial}{\partial x}\right)^2 + \left(\frac{\partial}{\partial y}\right)^2 + \left(\frac{\partial}{\partial z}\right)^2\right)}\ \psi$$

This definition of the Hamiltonian encounters difficulty. Partly because we don't know what to do with an operator that occurs under a square root

[25] See *Relativity* in the *Mathematics of Spin* .

[26] Note that p and $\tilde{p}$ have the same dimensions. $c\tilde{p}$ has dimensions of energy.

but more fundamentally because of the non linearity that it introduces. One cannot simply move V to the other side and square the remaining equation to get rid of the square root. This would involve squaring ψ which would destroy the linear superposition of wave functions on which quantum mechanics depends.

Dirac's solution was to make the quantity under the radical sign a perfect square. This cannot be done in terms of a single component wave function. His trick was to invoke a four component wave function which seems a rather startling break from previous physics but it was not quite as radical as it seems. Matrices and multi component variables had already been used in quantum mechanics and Dirac recognized that it was the trick that would work here.

As explained before, Dirac's solution lead to 4 by 4 matrices and very complicated looking equations.

We simplify the problem here by assuming a circular electron orbit at fixed radius in the z = 0 plane so that the electrons orbital velocity and momentum have only components in the plane. This allows us to use the 2x2 matrices that Pauli had carefully tailored for this kind of operation.

$$\sigma_{\mathbf{x}} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \qquad \sigma_{\mathbf{y}} = \begin{vmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{vmatrix}, \qquad \sigma_{\mathbf{z}} = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}. \tag{45}$$

These have, as before, the following properties:

$$\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} = -\sigma_{\mathbf{y}}\sigma_{\mathbf{x}}, \qquad \sigma_{\mathbf{y}}\sigma_{\mathbf{z}} = -\sigma_{\mathbf{z}}\sigma_{\mathbf{y}}, \qquad \sigma_{\mathbf{x}}\sigma_{\mathbf{z}} = -\sigma_{\mathbf{z}}\sigma_{\mathbf{x}} \tag{46}$$

and

$$\sigma_{\mathrm{x}}\sigma_{\mathrm{x}} = \mathbf{I}, \qquad \sigma_{\mathbf{y}}\sigma_{\mathbf{y}} = \mathbf{I}, \qquad \sigma_{\mathbf{z}}\sigma_{\mathbf{z}} = \mathbf{I}. \tag{47}$$

Now construct the following function for reasons which will appear below.

$$\left(\sigma_{\mathrm{z}}\mathrm{m}_0\mathrm{c}^2 + \mathrm{c}\sigma_{\mathrm{x}}\tilde{\mathrm{p}_{\mathrm{x}}} + \mathrm{c}\sigma_{\mathrm{y}}\tilde{\mathrm{p}_{\mathrm{y}}}\right) \tag{48}$$

[27] This quantity squared by the rules given as Eq. 46 and 47 and with $p_z = 0$ becomes the content of the radical in Eq. 42.

$$\left(m_0c^2\right)^2 + c^2(\tilde{p_x}^2 + \tilde{p_y}^2)$$

With this replacement Eq. 42 becomes

$$[(\mathcal{E} - V)\mathbf{I} \pm (c\sigma_x\tilde{p_x} + c\sigma_y\tilde{p_y} + \sigma_z m_0c^2)]\psi = 0 \tag{49}$$

[28]

$$\sigma_x = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}, \qquad \sigma_y = \begin{vmatrix} 0 & -i \\ i & 0 \end{vmatrix}, \qquad \sigma_z = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix}. \tag{50}$$

Since the σ are 2 x 2 matrices, we now need two equations each of which describes one of the two components of a new kind of wave function. Filling in the σ by their multipliers yields:

$$\left(\pm \begin{vmatrix} \mathcal{E} - V, & 0 \\ 0, & \mathcal{E} - V \end{vmatrix} + \begin{vmatrix} 0, & c\tilde{p_x} \\ c\tilde{p_x}, & 0 \end{vmatrix} + \begin{vmatrix} 0, -ic\tilde{p_y} \\ ic\tilde{p_y}, 0 \end{vmatrix} + \begin{vmatrix} m_0c^2, 0 \\ 0, -m_0c^2 \end{vmatrix}\right) \begin{vmatrix} \psi_\alpha \\ \psi_\beta \end{vmatrix} = 0$$

Or:

$$\left(\begin{vmatrix} \pm(\mathcal{E} - V) + m_0c^2, & 0 \\ 0, & \pm(\mathcal{E} - V) - m_0c^2 \end{vmatrix} + \begin{vmatrix} 0, & c\tilde{p_x} - ic\tilde{p_y} \\ c\tilde{p_x} + ic\tilde{p_y}, & 0 \end{vmatrix}\right) \begin{vmatrix} \psi_\alpha \\ \psi_\beta \end{vmatrix} = 0 \tag{51}$$

which expanded into two equations yields the following for a Dirac style treatment of an electron in a well.

[27] σ_z is not included in this line. We use it for another purpose in the next step.

[28] If the reader is disturbed by replacing the scalor $(\mathcal{E} - V)$ by its product with a unit matrix $\mathbf{I}$ be assured that this is resolved by Eq. 63 on page 53.

Corner Stone Equations

$$\begin{aligned}(\pm(\mathcal{E}-\mathrm{V})+\mathrm{m_0c^2})\psi_\alpha+(\mathrm{c\tilde{p}_x}-\mathrm{ic\tilde{p}_y})\psi_\beta=0\\(\mathrm{c\tilde{p}_x}+\mathrm{ic\tilde{p}_y})\psi_\alpha+(\pm(\mathcal{E}-\mathrm{V})-\mathrm{m_0c^2})\psi_\beta=0\end{aligned}\tag{52}$$

The two components of ψ can be separated by solving for one in terms of the other.

$$\begin{aligned}\left[((\mathcal{E}-\mathrm{V})^2-(\mathrm{m_0c^2})^2)-(\mathrm{c\tilde{p}_x}-\mathrm{ic\tilde{p}_y})(\mathrm{c\tilde{p}_x}+\mathrm{ic\tilde{p}_y})\right]\psi_\alpha=0\\\left[(\mathrm{c\tilde{p}_x}+\mathrm{ic\tilde{p}_y})(\mathrm{c\tilde{p}_x}-\mathrm{ic\tilde{p}_y})-((\mathcal{E}-\mathrm{V})^2-(\mathrm{m_0c^2})^2)\right]\psi_\beta=0\end{aligned}\tag{53}$$

These can be further simplified by noting that the total energy $\mathcal{E}$ of the electron consists mostly of its rest mass energy $\mathrm{m_0c^2}$ compared to which V is negligible. [29]

$$\begin{aligned}(\mathcal{E}-\mathrm{V})^2-(\mathrm{m_0c^2})^2=((\mathcal{E}-\mathrm{V})+(\mathrm{m_0c^2}))((\mathcal{E}-\mathrm{V})-(\mathrm{m_0c^2}))\approx\\2\mathrm{m_0c^2}((\mathcal{E}-\mathrm{V})-(\mathrm{m_0c^2}))=2\mathrm{m_0c^2}(\mathcal{E}'-\mathrm{V})\end{aligned}\tag{54}$$

where $\mathcal{E}'$ is the energy over and above the rest mass energy $\mathrm{m_0c^2}$. The reason for this last substitution is to make the energy terms resemble those of the Schrödiger form. This is a reasonable step since the rest mass energy does not appear except when the electron is annihilated. With these changes Eqs. 53 become

$$\left[2\mathrm{m_0c^2}(\mathcal{E}'-\mathrm{V})-(\mathrm{c\tilde{p}_x}-\mathrm{ic\tilde{p}_y})(\mathrm{c\tilde{p}_x}+\mathrm{ic\tilde{p}_y})\right]\psi_\alpha=0\tag{55}$$

[29] In fact we have assumed it zero in this example but carry it as a variable for possible future use.

$$\left[2m_0c^2(\mathcal{E}' - V) - (c\tilde{p_x} + ic\tilde{p_y})(c\tilde{p_x} - ic\tilde{p_y})\right]\psi_\beta = 0$$

Note we have reversed the positions of the terms in the second equation and reversed the signs to make it resemble, as far as possible, the equation above it.

Magnetic Moment

Now let us investigate the second term of the first equation above being careful to keep track of the order of operations and of the operand.

$$-(c\tilde{p_x} - ic\tilde{p_y})(c\tilde{p_x} + ic\tilde{p_y})\psi_\alpha$$

This product consists of four terms. As in the Schrödiger model we include the effect of a magnetic field by the substitution $\tilde{p} -> \tilde{p} + qA$ so that the first and last products become:

$$c\tilde{p_x}c\tilde{p_x}\psi_\alpha -> c^2(\tilde{p_x} + qA_x)^2\psi_\alpha$$

$$c\tilde{p_y}c\tilde{p_y}\psi_\alpha -> c^2(\tilde{p_y} + qA_y)^2\psi_\alpha$$

the sum of which divided by $2m_0c^2$ is the same as the rectangular bracket of Schrödiger's form Eq.37 shown below in the footnote. [30] However the cross product terms which do not occur in the Schrödiger form, $(c\tilde{p_x})(-ic\tilde{p_y}) + (ic\tilde{p_y})(c\tilde{p_x})$ gives us something new:

$$(ic^2\tilde{p_y}\tilde{p_x} - ic^2\tilde{p_x}\tilde{p_y})\psi_\alpha \tag{56}$$

[30]

$$i\hbar\frac{\partial\psi}{\partial t} = \mathcal{E}\psi = \frac{1}{2m_0}\left[(\tilde{p_x} + qA_x)^2 + (\tilde{p_y} + qA_y)^2\right]\psi + V\psi$$

The $\tilde{p_x}$ to $\tilde{p_x} + qA_x$ replacement applied to these terms yields:

$$ic^2\left[\left(\frac{\hbar}{i}\frac{\partial}{\partial y} + qA_y\right)\left(\frac{\hbar}{i}\frac{\partial}{\partial x} + qA_x)\right) - \left(\frac{\hbar}{i}\frac{\partial}{\partial x} + qA_x\right)\left(\frac{\hbar}{i}\frac{\partial}{\partial y} + qA_y\right)\right]\psi_\alpha \quad (57)$$

It is clear that the products of the first and last terms of each pair of parentheses cancel each other. The cross terms of these must be explored in detail. Each of these contains a $ic^2q\frac{\hbar}{i}$ term that can be factored out leaving

$$\left(A_y\frac{\partial}{\partial x} + \frac{\partial}{\partial y}A_x\right)\psi_\alpha - \left(A_x\frac{\partial}{\partial y} + \frac{\partial}{\partial x}A_y\right)\psi_\alpha$$

Recall that the constant magnetic field B_0 in the z direction is derived from the curl of $\vec{A}$. See Eq.39.

$$A_y = xB_0/2 \quad \text{and} \quad A_x = -yB_0/2$$

Thus the line above becomes:

$$\frac{B_0}{2}\left(x\frac{\partial\psi_\alpha}{\partial x} + \frac{\partial}{\partial y}(-y\psi_\alpha)\right) - \frac{B_0}{2}\left(-y\frac{\partial\psi_\alpha}{\partial y} + \frac{\partial}{\partial x}(x\psi_\alpha)\right) =$$

$$\frac{B_0}{2}\left(x\frac{\partial\psi_\alpha}{\partial x} - \psi_\alpha - y\frac{\partial\psi_\alpha}{\partial y}\right) - \frac{B_0}{2}\left(-y\frac{\partial\psi_\alpha}{\partial y} + \psi_\alpha + x\frac{\partial\psi_\alpha}{\partial x}\right) = -B_0\psi_\alpha$$

Now restoring the temporarily dropped factor $ic^2q\frac{\hbar}{i}$ yields a term

$$-c^2q\hbar B_0\psi_\alpha$$

which gets added to the last term within the bracket in the first of Eq. 55.

Assembling the parts of Eqs. 55 and dividing by m_0c^2 yields a pair of equations each of which is almost the same as the Schrödiger equation Eq.

40 shown below in the footnote.[31] where ∇ is the Laplacian operator defined in Eq. 41.

Therefore we now have

$$\left[(\mathcal{E}' - \mathrm{V}) + \frac{\hbar^2}{2\ \mathrm{m_0}}\nabla^2 - \frac{\mathrm{q}^2}{2\ \mathrm{m_0}}(\mathrm{A_x^2} + \mathrm{A_y^2}) - \mathrm{q}\frac{\mathrm{B_0}}{4\mathrm{m_0}}\ \tilde{\mathrm{p}}_\theta - \frac{\mathrm{q}\hbar}{2\mathrm{m_0}}\mathrm{B_0}\right]\psi_\alpha = 0 \tag{58}$$

$$\left[(\mathcal{E}' - \mathrm{V}) + \frac{\hbar^2}{2\ \mathrm{m_0}}\nabla^2 - \frac{\mathrm{q}^2}{2\ \mathrm{m_0}}(\mathrm{A_x^2} + \mathrm{A_y^2}) - \mathrm{q}\frac{\mathrm{B_0}}{4\mathrm{m_0}}\ \tilde{\mathrm{p}}_\theta + \frac{\mathrm{q}\hbar}{2\mathrm{m_0}}\mathrm{B_0}\right]\psi_\beta = 0$$

One observes that the second equation 55 for ψ_β is the same as the first for ψ_α except that the last two terms are reversed which reverses the order of the terms in Eq. 56 which in turn changes the sign of the last term in Eq. 58 above.

Since these are the same as the Schrödiger equation except for the energy represented by the last term, the solutions are the same except that they correspond to energies $\mathcal{E}'$ offset from the Schrödiger solutions by the negative of the last term.

Thus without solving for ψ we can immediately see that the energy depends upon three terms that depend on the magnetic field. One of these depends on the angular momentum of the electron in its orbit p_θ. This term and the A^2 term are the same for ψ_α and ψ_β and to the corresponding term in the Schrödiger equation. However the last term

$$\frac{\mathrm{q}\hbar}{2\mathrm{m_0}}\mathrm{B_0}$$

31

$$\left[(\mathcal{E} - \mathrm{V}) + \frac{\hbar^2}{2\ \mathrm{m_0}}\nabla^2 - \frac{\mathrm{q}^2}{2\ \mathrm{m_0}}(\mathrm{A_x^2} + \mathrm{A_y^2}) - \mathrm{q}\frac{\mathrm{B_0}}{4\mathrm{m_0}}\ \tilde{\mathrm{p}}_\theta\right]\psi = 0$$

is new and switches sign between ψ_α and ψ_β. This is the electron's magnetic moment. It is either parallel or antiparallel to B_0 and adds or subtracts from the energy level that would satisfy the boundary conditions for the well under the Schrödiger equation.

It is important to recognize that this property of adding or subtracting the spin from the energy, which Dirac referred to as "duplexity" does not arise from the possible double value of the square root in Eq.42. It came from the need of a two component ψ function to produce a Lorentz invariant wave equation and is a completely relativistic effect. The double value of the square root in Eq. 42 causes the two components to exchange places.

Angular Momentum of the Electron

In the familer version of quantum mechanics, one finds the average value of an observable from

$$\mathrm{Av(O)} = \int \psi^* \mathbf{O_{op}} \psi \mathrm{dx}$$

[32]

where we are assuming that ψ has been normalized so that

$$\int \psi^* \psi \mathrm{dx} = 1$$

and $\mathbf{O_{op}}$ is the operator associated with the observable O. The analogous operation for a two component ψ and a 2×2 operator would be

$$\mathrm{Av(O)} = \int < \psi | \mathbf{O_{op}} | \psi > \mathrm{dx}$$

where $|\psi >$ is a two component ψ_α, ψ_β column vector, and $< \psi|$ is a two component row vector with each component the complex conjugate of the corresponding component of $|\psi >$. $\mathbf{O_{op}}$ is a two by two matrix.

Once again the condition applys that

$$\int < \psi || \psi > \mathrm{dx} = 1$$

The result of Av(O) is a single number.

We are going to use this to investigate the angular momentum of the electron in the well. It first it appears as if one needs to know ψ before one

[32] dx is used here as an abriviation for whatever space one is considering. Thus it could be dxdydz for example.

can get any useful result, but there is at least one important case where the result does not depend on knowledge of ψ, when the $\mathbf{O_{op}} = 0$.

The relevance of the above is to demonstrate the peculiar fact that the time rate of change of angular momentum calculated from the Hamiltonian in 49 is not zero. Note, we are not referring to the angular momentum, but to its time rate of change. One would not necessarily expect the angular momentum of an electron in its orbit to be zero. However, in the absence of external influences (torques) we should expect the time rate of change to be zero.

Angular momentum as a function of time does not appear easily from a direct calculation unless ψ is known. However, we show in the *Mathematics of Spin* that the rate of change of the average of an observable is

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{Av}(\mathbf{O}) = \int < \psi|\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{O_{op}}|\psi > \mathrm{dx} \tag{59}$$

which, when the Hamiltonian does not contain time, $\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{O_{op}}$ is given by [33]

$$\left(\frac{\hbar}{\mathrm{i}}\right)\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{O_{op}} = \mathbf{O_{op}H} - \mathbf{HO_{op}}.$$

where $\mathbf{H}$ is the appropriate Hamiltonian, in this case the square bracket of Eq. 49 solved for $\mathcal{E}$.

$$\mathbf{H} = [\mathrm{V}\mathbf{I} \pm (\mathrm{c}\sigma_{\mathrm{x}}\tilde{\mathrm{p}_{\mathrm{x}}} + \mathrm{c}\sigma_{\mathrm{y}}\tilde{\mathrm{p}_{\mathrm{y}}} + \sigma_{\mathrm{z}}\mathrm{m}_0\mathrm{c}^2)] \tag{60}$$

Classically

$$\mathrm{p}_\theta = \mathrm{x}\mathrm{p}_{\mathrm{y}} - \mathrm{y}\mathrm{p}_{\mathrm{x}}$$

but in this formulation, the momentum operator is not a scalar quantity but a 2x2 matrix which in Cartesian coordinates is

$$\mathbf{O_{op}} = \frac{\hbar}{\mathrm{i}}\left(\mathrm{x}\frac{\partial}{\partial \mathrm{y}} - \mathrm{y}\frac{\partial}{\partial \mathrm{x}}\right)\mathbf{I} \quad \text{which is also} \quad \mathbf{I}\frac{\hbar}{\mathrm{i}}\frac{\partial}{\partial\theta} \tag{61}$$

[33] Sometimes called the equation of motion.

Therefore the operator for the average angular momentum about z becomes

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathrm{Av}\ (\mathrm{p}_\theta) = \frac{\mathrm{d}}{\mathrm{dt}}\mathrm{Av}(\mathrm{xp_y} - \mathrm{yp_x}) =$$

$$\int < \psi| \left(\mathrm{x}\frac{\partial}{\partial \mathrm{y}} - \mathrm{y}\frac{\partial}{\partial \mathrm{x}}\right) \mathbf{H} - \mathbf{H} \left(\mathrm{x}\frac{\partial}{\partial \mathrm{y}} - \mathrm{y}\frac{\partial}{\partial \mathrm{x}}\right) |\psi > \mathrm{dx}$$

and is in general not zero. However we are going to find that we can add a term to p_θ operator that will make the sum zero for any ψ.

The Hamiltonian from Eq. 49

$$\mathbf{H} = \mathrm{V}\mathbf{I} \pm (\mathrm{c}\sigma_\mathrm{x}\tilde{\mathrm{p}_\mathrm{x}} + \mathrm{c}\sigma_\mathrm{y}\tilde{\mathrm{p}_\mathrm{y}} + \sigma_\mathrm{z}\mathrm{m_0c^2})$$

consists of four parts two of which are constant diagonal matrices $|\mathbf{C.D.M.}|$. It is easily seen that the quantity

$$|\mathbf{C.D.M.}|\mathbf{H} - \mathbf{H}|\mathbf{C.D.M.}| = 0$$

Thus the only non zero parts of $\mathbf{O_{op}H} - \mathbf{HO_{op}}$would come from

$$\mathbf{H'} = \mathrm{c}\sigma_\mathrm{x}\tilde{\mathrm{p}_\mathrm{x}} + \mathrm{c}\sigma_\mathrm{y}\tilde{\mathrm{p}_\mathrm{y}} = \frac{\mathrm{c}\hbar}{\mathrm{i}} \left(\sigma_\mathrm{x}\frac{\partial}{\partial \mathrm{x}} + \sigma_\mathrm{y}\frac{\partial}{\partial \mathrm{y}}\right)$$

and from Eq. 61.

$$\mathbf{Op} = \frac{\hbar}{\mathrm{i}} \left(\mathrm{x}\frac{\partial}{\partial \mathrm{y}} - \mathrm{y}\frac{\partial}{\partial \mathrm{x}}\right) \mathbf{I}$$

Evaluating $\mathbf{OpH'} - \mathbf{H'Op}$ looks a bit long but it is straight forward manipulation with the slight restriction that one must be careful to preserve the order of differentiation of products. For example:

$$\left(\mathrm{x}\frac{\partial}{\partial \mathrm{y}}\right) \frac{\partial}{\partial \mathrm{x}} = \mathrm{x}\frac{\partial^2}{\partial \mathrm{x}\partial \mathrm{y}} \quad \text{is not the same as} \quad \frac{\partial}{\partial \mathrm{x}} \left(\mathrm{x}\frac{\partial}{\partial \mathrm{y}}\right) = \mathrm{x}\frac{\partial^2}{\partial \mathrm{x}\partial \mathrm{y}} + \frac{\partial}{\partial \mathrm{y}}$$

We are going to use two theorems which are proven in the *Mathematics of Spin* . These are:

$$p_a\sigma_b = \sigma_b p_a \quad \text{where a and b are any combination of x,y or z.}$$

and

$$\sigma_x\sigma_y = i\sigma_z \quad \sigma_y\sigma_z = i\sigma_x \quad \sigma_z\sigma_x = i\sigma_y$$

Recall that $\sigma_x\sigma_y = -\sigma_y\sigma_x$ etc.

With these in hand we proceed to calculate:

$$\mathbf{OpH' - H'Op} =$$

$$c\left(\frac{\hbar}{i}\right)^2\left[\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)\left(\sigma_x\frac{\partial}{\partial x} + \sigma_y\frac{\partial}{\partial y}\right) - \left(\sigma_x\frac{\partial}{\partial x} + \sigma_y\frac{\partial}{\partial y}\right)\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)\right]$$

The terms multipling $c\left(\frac{\hbar}{i}\right)^2\sigma_x$ are

$$\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)\frac{\partial}{\partial x} - \frac{\partial}{\partial x}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) =$$

$$x\frac{\partial^2}{\partial y\partial x} - y\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x}\left(x\frac{\partial}{\partial y}\right) + \frac{\partial}{\partial x}\left(y\frac{\partial}{\partial x}\right) =$$

$$x\frac{\partial^2}{\partial y\partial x} - y\frac{\partial^2}{\partial x^2} - x\frac{\partial^2}{\partial y\partial x} - \frac{\partial}{\partial y} + y\frac{\partial^2}{\partial x^2} = -\frac{\partial}{\partial y}$$

The terms multipling $c\left(\frac{\hbar}{i}\right)^2\sigma_y$ are

$$\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right)\frac{\partial}{\partial y} - \frac{\partial}{\partial y}\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) =$$

$$x\frac{\partial^2}{\partial y^2} - y\frac{\partial^2}{\partial x \partial y} - \frac{\partial}{\partial y}\left(x\frac{\partial}{\partial y}\right) + \frac{\partial}{\partial y}\left(y\frac{\partial}{\partial x}\right) =$$

$$x\frac{\partial^2}{\partial y^2} - y\frac{\partial^2}{\partial x \partial y} - x\frac{\partial^2}{\partial y^2} + y\frac{\partial^2}{\partial y \partial x} + \frac{\partial}{\partial x} = \frac{\partial}{\partial x}$$

So that

$$\frac{\hbar}{i}\frac{d}{dt}\mathbf{O_{op}} = \mathbf{OpH'} - \mathbf{H'Op} = c\left(\frac{\hbar}{i}\right)^2\left(-\sigma_x\frac{\partial}{\partial y} + \sigma_y\frac{\partial}{\partial x}\right)$$

or

$$\frac{d}{dt}\mathbf{O_{op}} = c\left(\frac{\hbar}{i}\right)\left(-\sigma_x\frac{\partial}{\partial y} + \sigma_y\frac{\partial}{\partial x}\right) = c(\sigma_y\tilde{p_x} - \sigma_x\tilde{p_y}) \tag{62}$$

This is of course an operator. It has no value until it operates on a ψ function. However, it is clear that when it is used as the operator inside Eq. 59 the value of the integral is not necessarily zero.

We remind the reader that the angular momentum of a body moving in the x,y plane is pointed in the z direction. Therefore in order to repair the difficulty above, we are going to have to add something that points in the z direction.

Let us return to the original trick that was used to linearize the radical and define what is meant by a momentum vector. See page 42.

We define $\vec{\tilde{p}}$ as

$$\vec{\tilde{p}} = \tilde{p_x}\sigma_x + \tilde{p_y}\sigma_y + \tilde{p_z}\sigma_z$$

We can assure ourselves that this still has the property required on page 42 by calculating the dot product of this vector with itself.

$$\vec{\tilde{p}} \cdot \vec{\tilde{p}} = \mathbf{I}(\tilde{p_x}^2 + \tilde{p_y}^2 + \tilde{p_z}^2) \tag{63}$$

We are assuming that $\tilde{p_z}$ is zero because the motion is in the x, y plane. From the above it seems reasonable to assume that a directional vector is given by

$$\vec{x} = x\sigma_x + y\sigma_y + z\sigma_z$$

The sigma have taken the role of unit vectors in conventional vector analysis.

Therefore if one is going to define an angular momentum in the z direction, it has to be multiplied by σ_z.

Now if one calculates the average time rate of change of $\sigma_z\hbar$ one finds that

$$\sigma_z\hbar\mathbf{H}' - \mathbf{H}'\sigma_z\hbar = c\hbar(\sigma_z\sigma_x\tilde{p_x} + \sigma_z\sigma_y\tilde{p_y}) - c\hbar(\sigma_x\sigma_z\tilde{p_x} + \sigma_y\sigma_z\tilde{p_y}) =$$

$$2c\hbar(\sigma_y\tilde{p_x} - \sigma_x\tilde{p_y}).$$

which is $2\hbar$ multiplied by the right side of Eq. 62.

Thus if $-\sigma_z\frac{\hbar}{2}$ is added to $\mathbf{O_{op}}$ a new operator is formed which yields zero for the average time rate of momentum. *The result does not depend upon the value of*ψ .

To quote Dirac, "This result one can interpret by saying that the electron has a spin angular momentum $\hbar/2$."[34]

Or in his original paper[35] "We can interpret this result by saying that the electron has a spin angular momentum $\frac{\hbar}{2}\sigma$..... Note that he does not say, nor imply, that this is the same as classical angular momentum but rather that it can be so interpreted.

The fact that it is not the same as classical angular momentum is less than startling when one recalls that the magnetic moment does not correspond to a ball of rotating charge as noted in the second section *Stern-Gerlach* of this

[34]p 266 Dirac

[35]Darwin, 'Roy.Soc.Proc., A,Vol 116, p227 (1927)

chapter. This is a new kind of entity and cannot be expected to be an obvious extension of familiar things.

Ref Pauli W 1927 Zur Quantenmechanik des magnetischen Elektrons Z. Phys. 43 601-23 Useful values and their dimensions.

$$\begin{array}{l} \text{For hydrogen r} = 5.3 \times 10^{-11}\text{meters.} \\ \hbar = 1.05 \times 10^{-34}\text{Joule sec.} = 6.5 \times 10^{-16} \text{ electron volt seconds} \\ \text{c} = 3 \times 10^{8} \text{ meters/sec.} \\ \text{electron volt} = 1.6 \times 10^{-19} \text{ Joules.} \\ \xi = 6 \times 10^{-16}\text{Joule} = 3700 \text{ electron volts.} \\ \mathcal{E}\text{lectron mass} = 9 \times 10^{-31}\text{kgm.} \\ \mathcal{E}\text{lectron charge} = 6.5 \times 10^{-19} \text{ Coulombs} \\ \text{Ionization energy about 13eV} \end{array} \tag{64}$$