Investigating the Foundations of Physical Law
5 Nilpotent quantum mechanics I

The Dirac equation

Many of the laws of physics can be seen as purely abstract statements about relations between fundamental quantities and the properties that emerge from the parameter group. While it is possible to derive a number of them by individual arguments, a more significant development is to try to explain the most fundamental virtually at one go from the packaging process that generates the algebraic group of order 64 connecting two vector spaces. This group turns out to be very significant indeed as it lies behind what seems to be the most fundamental equation in physics: Dirac's relativistic quantum mechanical equation for the fermion. At its deepest level, physics is only about fermions and their interactions, bosons, the only other particles, being generated by the interactions, and an equation for the fermion in its most general form will necessarily contain, explicitly or implicitly, many of the other principles that are significant in physics.

We have already made the claim that both the Dirac equation and the structure of the fermion can be derived, in principle, from the most efficient packaging of the parameters space, time, mass and charge. We now need to justify this claim by testing it against the Dirac equation as normally presented. Dirac started with the Schrödinger equation in which a second-order differential with respect to space is linked with a first-order differential with respect to time. Clearly, this is not relativistic, which requires both terms to be of the same order. Dirac needed a quantum version of Einstein's relativistic energy momentum equation, $E^2 - p^2 - m^2 = 0$, but quantizing this in the usual way gave the Klein-Gordon equation, with second-order differentials for space and time, which didn't give the required behaviour for the fermion. What he realised was needed was an equation that was *first-order* in the two variables. Effectively, this meant 'square-rooting' a second order equation to give *the same number* of linear terms. Now, this can be done if you use anticommuting operators, which will eliminate the cross-terms when the square root is squared again. The choice that may seem obvious now – quaternions – was not available to Dirac because of the general prohibition on their use, though he did have Pauli matrices. Dirac chose a set of 4×4 matrices and found an equation which worked spectacularly, especially in its automatic inclusion of the fermionic property of half-integral spin. In one of its main forms it looks something like:

$$\left(\gamma^{\mu}\partial_{\mu} + im\right)\psi = \left(\gamma^{0}\frac{\partial}{\partial t} + \gamma^{1}\frac{\partial}{\partial x} + \gamma^{2}\frac{\partial}{\partial y} + \gamma^{3}\frac{\partial}{\partial y} + im\right)\psi = 0$$

The γ terms are the matrices, and they are mutually anticommutative; γ^0 is a square root of 1 (actually the identity matrix I), and γ^1, γ^2, γ^3 are square roots of –1 (actually of $-I$), which are mutually orthogonal components of an object with vector properties ($\boldsymbol{\gamma}$) (the origin of the automatic inclusion of fermionic spin). The algebra is closed, but only with the addition of another matrix, γ^5, which is another square root of I, and is anticommutative to all the other matrices, but which does not appear in the equation. From the five generators, γ^0, γ^1, γ^2, γ^3, γ^5, we create a group of 64 possible combinations. The composition of the matrices is not unique, but it is usual to construct them from 2×2 Pauli matrices in the form

$$\boldsymbol{\gamma} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ -\boldsymbol{\sigma} & 0 \end{pmatrix} \qquad \gamma_0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \qquad \gamma_5 = \begin{pmatrix} 0 & -I \\ -I & 0 \end{pmatrix}$$

$$\gamma_1 = \begin{pmatrix} 0 & \sigma_x \\ -\sigma_x & 0 \end{pmatrix} \quad \gamma_2 = \begin{pmatrix} 0 & \sigma_y \\ -\sigma_y & 0 \end{pmatrix} \quad \gamma_3 = \begin{pmatrix} 0 & \sigma_z \\ -\sigma_z & 0 \end{pmatrix}$$

which can be expanded to:

$$\gamma_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \gamma_2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{pmatrix} \gamma_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\gamma_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \qquad I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The use of 4×4 matrices meant that the wavefunction ψ had to be expanded to a column vector with four components, which then meant that the equation had 4 'solutions', which could be identified with (fermion and antifermion) × (spin up and spin down). In fact it has the form of a spinor or a single component wavefunction multiplied by a spinor. If we make an explicit use of all the relevant matrices in the differential operator, the equation now becomes:

$$\begin{pmatrix} \frac{\partial}{\partial t} + im & 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial x} - i\frac{\partial}{\partial y} \\ 0 & \frac{\partial}{\partial t} + im & \frac{\partial}{\partial x} + i\frac{\partial}{\partial y} & -\frac{\partial}{\partial z} \\ -\frac{\partial}{\partial z} & -\frac{\partial}{\partial x} + i\frac{\partial}{\partial y} & -\frac{\partial}{\partial t} + im & 0 \\ -\frac{\partial}{\partial x} - i\frac{\partial}{\partial y} & \frac{\partial}{\partial z} & 0 & -\frac{\partial}{\partial t} + im \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} = 0 \qquad (1)$$

Now, Dirac seems nowhere to have investigated the true origin of his matrices, which, significantly form a group of order 64, including the identity matrix I and both real and complex versions of each matrix, and it is not a thing that you find in textbooks either, but, in fact, the way to construct the complete set of 4×4 matrices, *in all their variant forms*, is by using *two* sets of Pauli matrices, $\boldsymbol{\sigma}^1$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\sigma}^3$ and $\boldsymbol{\Sigma}^1$, $\boldsymbol{\Sigma}^2$, $\boldsymbol{\Sigma}^3$, which are commutative with each other. Again, this forms a group of order 64, with a minimum of 5 generating elements, which may be something like

$$\boldsymbol{\Sigma}^1 I,\ i\boldsymbol{\sigma}^1\boldsymbol{\Sigma}^3,\ i\boldsymbol{\sigma}^2\boldsymbol{\Sigma}^3,\ i\boldsymbol{\sigma}^3\boldsymbol{\Sigma}^3,\ i\boldsymbol{\Sigma}^2 I$$

or

$$\boldsymbol{\Sigma}^3 I,\ i\boldsymbol{\sigma}^1\boldsymbol{\Sigma}^1,\ i\boldsymbol{\sigma}^2\boldsymbol{\Sigma}^1,\ i\boldsymbol{\sigma}^3\boldsymbol{\Sigma}^1,\ i\boldsymbol{\Sigma}^2 I$$

and which, when we multiply them out, will be found to be identical in structure to γ^0, γ^1, γ^2, γ^3, γ^5. Now we know that our two Pauli algebras, $\boldsymbol{\sigma}^1$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\sigma}^3$ and $\boldsymbol{\Sigma}^1$, $\boldsymbol{\Sigma}^2$, $\boldsymbol{\Sigma}^3$, are exactly isomorphic to the two sets of vector units, **i**, **j**, **k**, and **i**, **j**, **k**, or two sets of complexified quaternions, that we have used before. Even using the two Pauli algebras, we can see that, in defining the group generators, we preserve the symmetry of one algebra ($\boldsymbol{\sigma}$) and break that of the other ($\boldsymbol{\Sigma}$).

The nilpotent Dirac equation

The standard form of the Dirac equation is not completely symmetric, as γ^5 is excluded although it is needed as one of the generators of the algebra. We can remedy this simply by premultiplying by $-i\gamma^5$:

$$-i\gamma^5\left(\gamma^\mu\partial_\mu + im\right)\psi = -i\gamma^5\left(\gamma^0\frac{\partial}{\partial t} + \gamma^1\frac{\partial}{\partial x} + \gamma^2\frac{\partial}{\partial y} + \gamma^3\frac{\partial}{\partial y} + im\right)\psi = 0\,.$$

Using the double Pauli version of the γ matrices we can convert this to

$$\boldsymbol{\Sigma}^2 I\left(\boldsymbol{\Sigma}^1 I\frac{\partial}{\partial t} + i\boldsymbol{\Sigma}^3\boldsymbol{\sigma}^1\frac{\partial}{\partial x} + i\boldsymbol{\Sigma}^3\boldsymbol{\sigma}^2\frac{\partial}{\partial y} + i\boldsymbol{\Sigma}^3\boldsymbol{\sigma}^3\frac{\partial}{\partial y} + im\right)\psi = 0$$

which then becomes

$$\left(-i\boldsymbol{\Sigma}^3 I\frac{\partial}{\partial t} - \boldsymbol{\Sigma}^1\boldsymbol{\sigma}^1\frac{\partial}{\partial x} - \boldsymbol{\Sigma}^1\boldsymbol{\sigma}^2\frac{\partial}{\partial y} - \boldsymbol{\Sigma}^1\boldsymbol{\sigma}^3\frac{\partial}{\partial y} + i\boldsymbol{\Sigma}^2 Im\right)\psi = 0\,,$$

and we can now begin to appreciate the new degree of symmetry we have created. Not only are all 5 generators included in the equation and one assigned to each term of the operator, but also the units of the two vector spaces, $\boldsymbol{\sigma}^1$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\sigma}^3$ and $\boldsymbol{\Sigma}^1$, $\boldsymbol{\Sigma}^2$, $\boldsymbol{\Sigma}^3$, are also all present, making the equation significantly different in its *physical* meaning to the one that we had before the transformation. This is now truly the story of two spaces, a fact which was completely lost in the original equation. We could continue

developing this form of the equation, but the physical meaning will become even more apparent if we replace our 5 Pauli generators by their equivalent from the double vector algebra, or, because of its origin in the combination of space, time, mass and charge, the vector-quaternion algebra. This will become even more convenient when we realise that the γmatrices, or their algebraic equivalents, are to be found within the wavefunction ψ, as well as in the operator. Of course, the signs of the quaternion units will be arbitrary, so, purely for convenience, we can replace the γ or double Pauli representations by, say,

$$-\boldsymbol{ik},\ -\boldsymbol{i}\mathbf{i},\ -\boldsymbol{i}\mathbf{j},\ -\boldsymbol{i}\mathbf{k},\ \boldsymbol{ij}$$

Our equation now becomes

$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-\boldsymbol{i}\mathbf{i}\frac{\partial}{\partial x}-\boldsymbol{i}\mathbf{j}\frac{\partial}{\partial y}-\boldsymbol{i}\mathbf{k}\frac{\partial}{\partial y}+\boldsymbol{j}m\right)\psi=\left(-\boldsymbol{k}\frac{\partial}{\partial t}-\boldsymbol{i}\boldsymbol{i}\nabla+\boldsymbol{j}m\right)\psi=0$$

and we can see that it represents a *purely mathematical* development from the original version, without any additional physical assumption. However, we obtain a major physical result as soon as we insert a plane wave solution, for a free particle, say,

$$\psi=A\,e^{-i(Et-\mathbf{p}.\mathbf{r})},$$

into this equation, for we immediately obtain

$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-\boldsymbol{i}\boldsymbol{i}\nabla+\boldsymbol{j}m\right)Ae^{-i(Et-\mathbf{p}.\mathbf{r})}=\left(\boldsymbol{ikE}+\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)Ae^{-i(Et-\mathbf{p}.\mathbf{r})}=0\,.$$

which has exactly the same form as the equation we obtained directly from the parameter group in the previous lecture, and which we can recognise as expressing a minimal statement of the variation of space and time within a structure created from the minimal algebraic generators of the parameter group.

We can see also that, if energy is conserved, the term in the bracket on the right-hand side is a nilpotent or square root of 0, and that the equation can only hold if A becomes identical to ($\boldsymbol{ikE}+\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m$) or a scalar multiple of it, which means that ψ is also a nilpotent. (Here, of course, we need to emphasize that the multivariate properties of $\mathbf{p}$ allow us to use the 'spin' terms $\mathbf{p}$ and $\boldsymbol{\nabla}$ instead of the 'helicity' terms $\boldsymbol{\sigma}.\mathbf{p}$ and $\boldsymbol{\sigma}.\boldsymbol{\nabla}$, where $\boldsymbol{\sigma}$ is a unit pseudovector of magnitude $-\mathbf{1}$, in a nilpotent structure, since $(\boldsymbol{\sigma}.\mathbf{p})^2=\mathbf{pp}=p^2$.) The very remarkable thing about this representation, and it carries over into cases where the fermion is not a free particle, is that the wavefunction is an explicit expression involving energy and momentum terms, ($\boldsymbol{ikE}+\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m$) $e^{-i(Et-\mathbf{p}.\mathbf{r})}$, a physical *object*, not a black box hidden behind the symbol ψ, which can only be seen as a calculating device. Quantum mechanics remains abstract and probabilistic, but it ceases to look arbitrary. We need to think outside the black box!

Of course, the spinor properties of the algebra still hold, even when we don't use a matrix representation, and ψ will not actually be a single term but rather a 4-component spinor, incorporating fermion / antifermion and spin up / down states. We can easily identify these possibilities with the arbitrary sign options for the iE and $\mathbf{p}$ (or $\boldsymbol{\sigma}.\mathbf{p}$) terms. This is easily accommodated in the nilpotent formalism by transforming $(i\mathbf{k}E + \mathbf{i}\mathbf{p} + \mathbf{j}m)$ into a column vector with four sign combinations of iE and $\mathbf{p}$, which may be written in abbreviated form as $(\pm i\mathbf{k}E \pm \mathbf{i}\mathbf{p} + \mathbf{j}m)$. Written out in full the four components are:

$$(i\mathbf{k}E + \mathbf{i}\mathbf{p} + \mathbf{j}m)$$
$$(i\mathbf{k}E - \mathbf{i}\mathbf{p} + \mathbf{j}m)$$
$$(-i\mathbf{k}E + \mathbf{i}\mathbf{p} + \mathbf{j}m)$$
$$(-i\mathbf{k}E - \mathbf{i}\mathbf{p} + \mathbf{j}m)$$

The signs are, of course, intrinsically arbitrary, but it is convenient to identify the four states by adopting a convention, say,

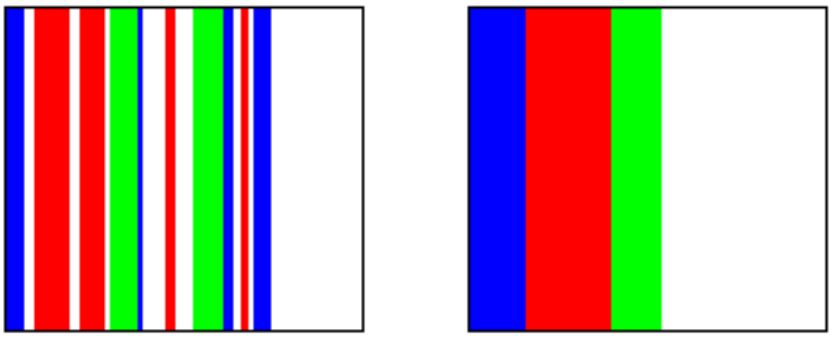
$(i\mathbf{k}E + \mathbf{i}\mathbf{p} + \mathbf{j}m)$	fermion spin up
$(i\mathbf{k}E - \mathbf{i}\mathbf{p} + \mathbf{j}m)$	fermion spin down
$(-i\mathbf{k}E + \mathbf{i}\mathbf{p} + \mathbf{j}m)$	antifermion spin down
$(-i\mathbf{k}E - \mathbf{i}\mathbf{p} + \mathbf{j}m)$	antifermion spin up

Once the convention is decided, however, the spin state of the particle (or, more conventionally, the helicity or handedness $\boldsymbol{\sigma}.\mathbf{p}$, the direction of spin relative to motion, with right-handed being positive) is fixed by the ratio of the signs of E and $\mathbf{p}$. So $\mathbf{i}\mathbf{p}$ / $i\mathbf{k}E$ has the same helicity as $(-\mathbf{i}\mathbf{p})$ / $(-i\mathbf{k}E)$, but the opposite helicity to $\mathbf{i}\mathbf{p}$ / $(-i\mathbf{k}E)$. In the nilpotent theory, the first or lead term represents the particle as observed, though all real fermions incorporate both spin states. If we wanted to represent an antifermion spin down, that term would become the lead term, and all the remaining terms would follow the same cycle of sign changes.

Conventionally, relativistic quantum mechanics uses the large matrix equation (1), but this has massive disadvantages, apart from the difficulty in writing down amplitudes and the hidden nature of the terms in the wavefunction. First of all, the γ matrices are not all real or all complex, so they don't have the same mathematical status and don't look symmetric – this is because we are trying to use a 2-dimensional algebra (complex numbers) to represent a 3-dimensional system. For the same reason, they are not symmetric as a layout, γ^3 being skewed with respect to the other two. These have damaging consequences, for example, creating the expression $\partial / \partial x + i\partial / \partial y$, which denies the rotation symmetry of the momentum operator and simply has no counterpart in nature. A similar problem happens with the energy operator $\partial / \partial t$, which is variously + and – with respect to im. This carries through at higher levels and means that we have to duplicate equations with alternative solutions, sometimes with

a broken connection between them. In addition, we have an enormous number of terms, 16 in the operator and 4 in the wavefunction, making 64 in total, which we have to consider at once. All this comes about because we haven't respected nature's choice of mathematics. If we do, then the result looks very different.

Using the quaternion operators, ***i***, ***j***, ***k***, almost as an extension of the way we use complex numbers to separate out the different parts into their appropriate compartments, we can unscramble or defragment the equation, so that, instead of looking like the left-hand side of our diagram, it looks more like the right, with energy, momentum and mass terms in both operator and amplitude, collected under the respective labels ***k***, ***i*** and ***j***:



$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-i\boldsymbol{i}\nabla+\boldsymbol{j}m\right)\left(\pm i\boldsymbol{k}E\pm\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{\mp i(Et\pm\mathbf{p}.\mathbf{r})}=0$$

Here, we have given the full specification for the spinor wavefunction, with four amplitudes and four corresponding phase factors. But, we can also make another massive simplification here. This is best seen by writing out the equation for each of the terms separately.

$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-i\boldsymbol{i}\nabla+\boldsymbol{j}m\right)\left(i\boldsymbol{k}E+\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{-i(Et-\mathbf{p}.\mathbf{r})}=0$$

$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-i\boldsymbol{i}\nabla+i\boldsymbol{j}m\right)\left(i\boldsymbol{k}E-\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{-i(Et+\mathbf{p}.\mathbf{r})}=0$$

$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-i\boldsymbol{i}\nabla+\boldsymbol{j}m\right)\left(-i\boldsymbol{k}E+\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{i(Et+\mathbf{p}.\mathbf{r})}=0$$

$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-i\boldsymbol{i}\nabla+i\boldsymbol{j}m\right)\left(-i\boldsymbol{k}E-\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{i(Et-\mathbf{p}.\mathbf{r})}=0$$

Here, we have one operator, four amplitude terms and four phase factors. But the reduction of the operator down from sixteen terms to one has given us the 'logical' space to transfer the variation in the *phase factor* to the operator, so extending the operator to four terms (written out in a column vector), operating on four amplitudes

(written out in a row vector) with a single, shared, phase. Written out as separate equations, this would become:

$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}-i\boldsymbol{i}\nabla+\boldsymbol{j}m\right)\left(i\boldsymbol{k}E+\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{-i(Et-\mathbf{p}.\mathbf{r})}=0$$
$$\left(-\boldsymbol{k}\frac{\partial}{\partial t}+i\boldsymbol{i}\nabla+\boldsymbol{j}m\right)\left(i\boldsymbol{k}E-\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{-i(Et-\mathbf{p}.\mathbf{r})}=0$$
$$\left(\boldsymbol{k}\frac{\partial}{\partial t}-i\boldsymbol{i}\nabla+\boldsymbol{j}m\right)\left(-i\boldsymbol{k}E+\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{-i(Et-\mathbf{p}.\mathbf{r})}=0$$
$$\left(\boldsymbol{k}\frac{\partial}{\partial t}+i\boldsymbol{i}\nabla+i\boldsymbol{j}m\right)\left(-i\boldsymbol{k}E-\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{-i(Et-\mathbf{p}.\mathbf{r})}=0$$

Since finding the phase factor is the biggest single problem in quantum mechanics, then we can expect a very significant increase in calculating power from this modification alone. Using again a compactified notation for row and column, this set of equations can be written:

$$\underset{row}{\left(\mp\boldsymbol{k}\frac{\partial}{\partial t}\mp i\boldsymbol{i}\nabla+\boldsymbol{j}m\right)}\underset{column}{\left(\pm i\boldsymbol{k}E\pm\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)}e^{-i(Et-\mathbf{p}.\mathbf{r})}=0$$

In this form of the equation, operator and amplitude look the same, especially if we write the first bracket in 'operator' form, with $i\partial/\partial t\rightarrow E$ and $-i\nabla\rightarrow\mathbf{p}$:

$$\left(\pm i\boldsymbol{k}E\pm\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)\left(\pm i\boldsymbol{k}E\pm\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m\right)e^{-i(Et-\mathbf{p}.\mathbf{r})}=0$$

Both are 4-component spinors. We note also that the Feynman prescription (obtainable from the Klein-4 group symmetry) of antifermions having reversed time is exactly obeyed. It seems to me that this is the form of the equation we would have arrived at if we could have derived it from first principles from the parameter group, using all available sign variations, rather than by 'reverse engineering' from the combination of the separately derived theories of special relativity and nonrelativistic quantum mechanics. It is apparent also that the nilpotent expression

$$(\pm i\boldsymbol{k}E\pm\boldsymbol{i}p_x\pm\boldsymbol{j}p_y\pm\boldsymbol{k}p_z+\boldsymbol{j}m)\,(i\boldsymbol{k}E\pm\boldsymbol{i}p_x\pm\boldsymbol{j}p_y\pm\boldsymbol{k}p_z+\boldsymbol{j}m)=0$$

or

$$(\pm i\boldsymbol{k}E\pm\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m)\,(\pm i\boldsymbol{k}E\pm\boldsymbol{i}\mathbf{p}+\boldsymbol{j}m)=0$$

can be used to create the Klein-Gordon equation that applies to bosons as well as fermions, by treating the E and $\mathbf{p}$ terms in *both* brackets as operators, acting on a particle with wavefunction $A\,e^{-i(Et-\mathbf{p}.\mathbf{r})}$, leading to

$$\left(\mp \boldsymbol{k}\frac{\partial}{\partial t} \mp i\boldsymbol{i}\nabla + \boldsymbol{j}m\right)\left(\mp \boldsymbol{k}\frac{\partial}{\partial t} \mp i\boldsymbol{i}\nabla + \boldsymbol{j}m\right)Ae^{-i(Et-\mathbf{p}.\mathbf{r})} = 0$$

or

$$\left(\nabla^2 - \frac{\partial^2}{\partial t^2} - m^2\right)Ae^{-i(Et-\mathbf{p}.\mathbf{r})} = 0\,.$$

It is clear that the Klein-Gordon equation will apply to any particle with phase factor $e^{-i(Et\,-\,\mathbf{p}.\mathbf{r})}$, as it leads to the product $(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$ $(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$. However, the Dirac equation will not apply to a particle with a scalar amplitude, as application of the single differential operator will only produce one factor containing $(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$, and so be unable to zero the result.

Using discrete differentiation

With a further refinement we can reduce the amount of information needed to specify relativistic quantum mechanics to just a two-term operator, while reducing the 'wavefunction' to an amplitude without a phase factor. Here, we use a discrete or anticommutative differentiation process, with a correspondingly discrete wavefunction. We start by defining a discrete differentiation (mainly following Kauffman) of the function F, which preserves the Leibniz chain rule, by taking:

$$\frac{\partial F}{\partial t} = [F, \mathcal{H}] = [F, E] \qquad \text{and} \qquad \frac{\partial F}{\partial X_i} = [F, P_i],$$

with $\mathcal{H} = E$ and P_i representing energy and momentum operators. Taking the case where velocity operators are not in evidence, we use $\partial F \,/\, \partial t$ rather than Kauffman's $dF \,/\, dt$. The mass term (which, of course, has only a passive role in quantum mechanics) disappears in the operator, though it is retained in the amplitude. Suppose we define a nilpotent amplitude

$$\psi = i\boldsymbol{k}E + \boldsymbol{i}\mathbf{i}P_1 + \boldsymbol{i}\mathbf{j}P_2 + \boldsymbol{i}\mathbf{k}P_3 + \boldsymbol{j}m$$

and an operator

$$\mathcal{D} = i\boldsymbol{k}\frac{\partial}{\partial t} + \boldsymbol{i}\mathbf{i}\frac{\partial}{\partial X_1} + \boldsymbol{i}\mathbf{j}\frac{\partial}{\partial X_2} + \boldsymbol{i}\mathbf{k}\frac{\partial}{\partial X_3},$$

with $$\frac{\partial \psi}{\partial t} = [\psi, \mathcal{H}] = [\psi, E] \qquad \text{and} \qquad \frac{\partial \psi}{\partial X_i} = [\psi, P_i],$$

After some basic algebraic manipulation, we obtain

$$\begin{aligned} -\mathcal{D}\psi = {} & i\psi(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{i}P_1 + \boldsymbol{i}\mathbf{j}P_2 + \boldsymbol{i}\mathbf{k}P_3 + \boldsymbol{j}m) \\ & + i(i\boldsymbol{k}E + \boldsymbol{i}\mathbf{i}P_1 + \boldsymbol{i}\mathbf{j}P_2 + \boldsymbol{i}\mathbf{k}P_3 + \boldsymbol{j}m)\psi - 2\,i(E^2 - {P_1}^2 - {P_2}^2 - {P_3}^2 - m^2). \end{aligned}$$

When ψ is nilpotent, then

$$\mathcal{D}\psi = \left(\boldsymbol{ik}\frac{\partial}{\partial t} + \boldsymbol{i}\nabla \right)\psi = 0 .$$

Generalising to four states, with $\mathcal{D}$ and ψ represented as 4-spinors, then

$$\mathcal{D}\psi = \left(\pm i\boldsymbol{k}\frac{\partial}{\partial t} \pm \boldsymbol{i}\nabla \right)\left(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{i}P_1 \pm \boldsymbol{i}\mathbf{j}P_2 \pm \boldsymbol{i}\mathbf{k}P_3 + \boldsymbol{j}m \right) = 0$$

becomes an equivalent of the nilpotent Dirac equation in this discrete calculus. The reason why this derivation is important is that it does not require the $\pm i$ (or $i\hbar$) term usually applied to the differential operators in canonical quantization, though, because our expression incorporates all four sign variations, we can include this if we want to without changing anything significant in the equation. This means that we can make a smooth transition between classical and quantum conditions, which points to an important connection between mass and the appearance of $\pm i$ terms in quantum mechanical operators and amplitudes.

Idempotents and nilpotents

Relativistic quantum mechanics was always assumed to require idempotent, rather than nilpotent wavefunctions, essentially because spinors are built up from primitive idempotents. Idempotents square to themselves rather than to zero, but, as we saw in lecture 2, they are closely related to nilpotents, and each can be converted into the other using quaternion multipliers. We can, in fact, make *exactly the same equation* look either idempotent or nilpotent simply by redistributing a single algebraic unit between the sections of the equation defined as operator and wavefunction. So, we can write the basic equation as either

$$\underset{operator}{[(\mp\ \boldsymbol{k}\partial/\partial t \mp\ i\boldsymbol{i}\nabla + \boldsymbol{j}m)\,\boldsymbol{j}]}\ \underset{idempotent\ \ wavefunction}{[\boldsymbol{j}(\pm\, i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\, e^{-i(Et - \mathbf{p}.\mathbf{r})}]} = 0$$

or

$$\underset{operator}{[(\mp\, \boldsymbol{k}\partial/\partial t \mp\ i\boldsymbol{i}\nabla + \boldsymbol{j}m)\,\boldsymbol{jj}]}\ \underset{nilpotent\ \ wavefunction}{[(\pm\, i\ \boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\, e^{-i(Et - \mathbf{p}.\mathbf{r})}]} = 0$$

without changing either physics or mathematics. The wavefunction, in fact, has both idempotent and nilpotent properties. Even the wavefunction in the original, 'conventional' version of the equation (1) was a nilpotent, because we brought out the nilpotent characteristics merely by multiplying the operator from the left, and not by changing the wavefunction. The idempotent aspect ensures that the 4-component wavefunction acts as a spinor, as in the conventional version of relativistic quantum mechanics.[1]

Besides helping to define spinor characteristics, the idempotent property, as we will show, has an additional special physical significance, related to vacuum. However, the nilpotent property has more powerful physical characteristics. Quantum mechanics is not just a set of mathematical formalisms which act as calculating devices. Each formalism tells us something about the nature of the physical system, and, as we will show, nilpotency is a statement of a *physical* principle, or one that is perhaps even more universal, rather than just a mathematical operation. As we will soon show, the nilpotent formalism is optimised for calculating efficiency, even by comparison with the nonrelativistic versions of quantum mechanics. One of the main reasons for this is the reduction to a single phase factor means that all aspects of a problem can be dealt with in a single calculation, but a more important one is that the extra constraint of nilpotency immediately forces on us an interpretation which connects every fermion holistically and nonlocally with the rest of the universe, in a way that makes much of the formal apparatus of relativistic quantum mechanics redundant. It seems to be impossible to make quantum physics any more minimal.

Pauli exclusion

If nilpotency is physically significant, what does it actually mean? We can answer this immediately: Pauli exclusion. A fermion, with a nilpotent wavefunction, let's say ψ_1, is automatically Pauli exclusive, because the combination state it forms with another fermion in an identical state, $\psi_1 \psi_1$, will necessarily be zero. Naturally occurring fermions, however, are not free particles, and the breakthrough occurs when we say that Pauli exclusion will occur for non-free fermions *for exactly the same reasons*. In these cases, we can redefine the operators E and $\mathbf{p}$ to incorporate any number of field terms or covariant derivatives, so that E now becomes, say, $i\partial / \partial t + e\phi + \ldots$, and $\mathbf{p}$ becomes, say, $-i\nabla + e\mathbf{A} + \ldots$, while preserving the overall structure of the nilpotent operator as $(\pm ikE \pm i\mathbf{p} + jm)$.

Clearly, if we do this, the E and $\mathbf{p}$ interpreted as eigenvalues in the amplitude will require more complicated expressions due to the presence of the additional terms in the operator. The phase factor will also no longer be the $e^{-i(Et - \mathbf{p}.\mathbf{r})}$ that we generate for the free particle; but the system will still be constrained by the need to maintain Pauli exclusion for all fermions, whether free or interacting, and its properties will be uniquely determined by the information that the operator has encoded. This will be true even if the external field terms are defined by expectation values, as with the Lamb shift, or in terms of quantum fields.

Writing an operator in the generic form $(\pm ikE \pm i\mathbf{p} + jm)$, with E and $\mathbf{p}$ now defined as generic terms involving differentials and associated potentials, means that we immediately specify the entire quantum mechanics of the system. The wavefunction and even *the equation itself* become redundant as independent sources of information.

The operator alone uniquely and completely determines the phase factor which will create a nilpotent amplitude. This does not require a special quantum mechanical equation, only a secondary functional one:

$$(\text{operator acting on phase factor})^2 = 0.$$

Even the spinor representation loses its independent status, as the first of the four terms, say ($i\mathbf{k}E + i\mathbf{p} + j\mathbf{m}$), will uniquely and completely specify the remaining three by automatic sign variation. (For this reason, we will often specify the operator in an abbreviated form using only the first term, and, when we write down an amplitude, we will often leave out the phase factor, which again contains no independent information.) Instead of needing an operator with 16 terms and an amplitude with 4 terms, we can specify the entire system with just a single term from an operator. Astonishingly, we don't need an equation at all to do quantum mechanics. The seemingly most fundamental structure in physics as we know it today is not an equation but an expression linking the symmetric algebras of four basic quantities, and the abstract properties that these algebras create, which we interpret as 'physics'.

Vacuum

Vacuum is an important concept in quantum physics but it is not well understood. In the nilpotent theory, however, it has a very clear meaning, which allows an immediate transition to quantum field theory, without any formal process of second quantization. It also invokes the fundamental idea of totality zero. We imagine creating a fermion in some particular state (determined by added potentials, interaction terms, etc) *ab initio*, that is, from absolutely nothing, a complete void or totality zero. We can then imagine vacuum as what is left in *nothing* – that is, everything other than the fermion. If we define the wavefunction of the fermion as, say, ψ_f, the wavefunction of vacuum will become $\psi_v = -\psi_f$. The superposition of fermion and vacuum will be initial zero state, $\psi_f + \psi_v = \psi_f - \psi_f = 0$, and, because the fermion is a nilpotent, the combination state

$$\psi_f \psi_v = -\psi_f \psi_{ff} = -(\pm i\mathbf{k}E \pm i\mathbf{p} + j\mathbf{m})\,(\pm i\mathbf{k}E \pm i\mathbf{p} + j\mathbf{m})$$

will also be zero. So, vacuum can be described as the 'hole' in the totality zero state produced by the creation of the fermion, and this becomes the 'rest of the universe' that the fermion sees and interacts with. So, if we want to 'create' a fermion with interacting field terms, then the 'rest of the universe' has to be 'constructed' simultaneously to make the existence of a fermion in that state possible.

Vacuum in this definition leads to a view of the universe as a zero totality, which, at the creation of every new fermion state, divides into two parts – the local fermion state and the nonlocal vacuum. These are connected with the simultaneous existence

of two vector spaces to create the mathematical structure: the real space with units **i**, **j**, **k** and the vacuum space with units **i**, **j**, **k** (derived from the mass-time-charge units of 1, i, $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$). The nilpotent formalism indicates that a fermion 'constructs' its own vacuum, or the entire 'universe' in which it operates, and we can consider the vacuum to be 'delocalised' to the extent that the fermion is 'localised'. We can consider the nilpotency as defining the interaction between the localised fermionic state and the delocalised vacuum, with which it is uniquely self-dual, the phase being the mechanism through which this is accomplished. We can also consider Pauli exclusion as saying that no two fermions can share the same vacuum.

Nilpotency now makes it possible for us to understand nonlocality, giving us a simple separation of local and nonlocal. The 'local' (which implies Lorentz or relativistic invariance) is defined as whatever happens inside the nilpotent structure ($\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m$), and the 'nonlocal' as whatever happens outside it. The bracketed term representing the fermion creation operator or wavefunction determines how conservation of energy applies to that fermion, as squaring the wavefunction and equating to zero gives us back the energy-momentum equation, and, of course, it is local, as the required Lorentzian structure is intrinsic. However, the addition and multiplication of nilpotent wavefunctions construct the nonlocal processes of superposition and combination, and these processes do not require a Lorentzian structure.

No single fermion can ever be isolated but must be interacting, and construct a 'space', so that its vacuum is not localised on itself. A point-like fermion necessarily requires a dispersed vacuum. Vacuum is intrinsically nonlocal. Because the fermion is localized, then the rest of the universe is necessarily nonlocalized. If the fermion is a point, as experiments suggests that it may be, then the rest of the universe is defined as everything outside that point. We can't define a single (noninteracting) fermion to exist – it can only be defined if we also define its vacuum. So the nonlocal connection which makes Pauli exclusion possible can be thought to occur through the vacua for each fermion.

Nilpotency isn't the usual way of expressing Pauli exclusion mathematically. In the standard interpretation, wavefunctions or amplitudes are also Pauli exclusive because they are antisymmetric, with nonzero

$$\psi_1\psi_2 - \psi_2\psi_1 = -(\psi_2\psi_1 - \psi_1\psi_2)$$

This, however, is automatic in the nilpotent formalism, where the expression becomes

$$(\pm i\mathbf{k}E_1 \pm i\mathbf{p}_1 + j m_1)\,(\pm i\mathbf{k}E_2 \pm i\mathbf{p}_2 + j m_2)$$
$$-(\pm i\mathbf{k}E_2 \pm i\mathbf{p}_2 + j m_2)\,(\pm i\mathbf{k}E_1 \pm i\mathbf{p}_1 + j m_1)$$
$$= 4\,\mathbf{p}_1\,\mathbf{p}_2 - 4\,\mathbf{p}_2\,\mathbf{p}_1 = 8\,i\,\mathbf{p}_1 \times \mathbf{p}_2.$$

which is clearly antisymmetric. But the result also tells us something new, for it requires a nilpotent wavefunction to have a **p** vector in spin space at a different orientation to any other. The instantaneous nonlocal correlation of the wavefunctions of all nilpotent could then require the intersection of the planes corresponding to all the different **p** vector directions. We can even consider these intersections as actually creating the *meaning* of Euclidean space, with an intrinsic spherical symmetry generated by the fermions themselves.

It seems likely that the universal collection of spin axis directions at any given instant is not repeatable, giving a unique direction for time, and – including energy as the 'time' component – making fermionic world-lines unique as well. A further connection with irreversibility or a unique time direction is suggested by the fact that an interaction between fermions with differently oriented vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ will produced a reduced magnitude of momentum within the system unless the even more 'organized' amount of rest mass is reduced. For nilpotents, $\mathbf{p}_1$ and $\mathbf{p}_2$ can never be parallel, and the vector sum of components in different directions will always be less than the scalar magnitudes added. In either case, the system will become less coherent or more entropic, as required by the second law of thermodynamics, the only law of physics which requires an irreversibility of time.

Pauli exclusion means that defining a fermion implies simultaneous definition of vacuum as 'the rest of the universe' with which it interacts. The nilpotent structure then provides energy-momentum conservation without requiring the system to be closed, since the E and **p** terms also contain all possible interactions. Again, we see that the nilpotent structure is naturally thermodynamic, and provides a route to a mathematization of nonequilibrium thermodynamics – all systems in this formulation are open systems. Also, the formation of any new state, which is determined by the nature of all other nilpotent states, is a creation event within a unique birth-ordering. Each 'creation' event (which includes any interaction and any change in parameters, as well as entirely new fermionic creations) also necessarily changes all existing states to some degree. In this sense, a nilpotent structure uniquely allows us to conceive of the infinite while only observing the finite.

We now have at least five different meanings for the expression

$$(\pm i\mathbf{k}E \pm i\mathbf{p} + j m)\,(\pm i\mathbf{k}E \pm i\mathbf{p} + j m)\,\phi \to 0$$

with ϕ an (optional) arbitrary scalar factor (phase, etc.):

classical	special relativity
operator × operator	Klein-Gordon equation
operator × wavefunction	Dirac equation
wavefunction × wavefunction	Pauli exclusion
fermion × vacuum	thermodynamics

It is characteristic of a theory in which duality is so deeply embedded to create such multiple meanings. The same is true of ***i***, ***j***, ***k***, which, as we will see, have many meanings, including charge operators, vacuum operators and generators of *P*, *C* and *T* symmetry transformations.

Yet another way of looking at Pauli exclusion is to say that the phase factors of all fermion wavefunctions must be unique. In effect, this means that the *E* and **p** values, and amplitudes are all unique, which is the same as saying that they are nilpotent. Here, we might return to our dual spaces, **i**, **j**, **k** and **i**, **j**, **k**, which can be said, in one interpretation to carry the entire information relating to physics, the first being rotationally symmetric and the second rotationally asymmetric. The first is real space or the space of measurement, and the second the vacuum space or space of interaction. The nilpotent condition requires these to be dual in terms of all physical information, although presenting it in quite different forms except in the construction of spinors (lecture 2). The nilpotency seen in vacuum space, which creates the unique phase factors for fermions, might present this phase as a unique direction on a set of axes defined by the values of *E*, **p** and *m*, by analogy with the unique direction of **p** on the axes of real space that we derived from the antisymmetric wavefunctions. The mapping of the unique phase on to both spin vector (using **i**, **j**, **k**) and to the axes whose directions were derived from ***i***, ***j***, ***k***, gives a good indication that the nilpotent structure is rightly considered an angular momentum operator, and the link of ***i***, ***j***, ***k*** with the three charge components shows the connection between angular momentum and charge which we discussed in lecture 4.

A number of significant results emerge automatically from the ***i***, ***j***, ***k*** or **i**, **j**, **k** representation. For example, half of the possibilities on one axis (those with $-m$) would be eliminated automatically (as being in the same direction as those with m), providing clear evidence that invariant mass cannot have two signs. Also eliminated would be fermions with zero m (since the directions would all be along the line $E = p$). In addition, such hypothetical massless particles would be impossible for fermions and antifermions with the same helicity, as E, p has the same direction as $-E$, $-p$.

Quantum mechanics and the quantum field

The nilpotent operator can be used to do ordinary relativistic quantum mechanics, once we have defined a probability density for a nilpotent wavefunction, ($\pm i\mathbf{k}E \pm \mathbf{i}\mathbf{p} +$

jm). In standard quantum mechanics, we create this by multiplying the wavefunction by its complex conjugate. Here, we use the *complex quaternion conjugate* ($\pm ikE \mp \mathbf{i}p - jm$) (the extra 'quaternion' resulting from the fact that the nilpotent wavefunction differs from a conventional one through premultiplication by a quaternion operator). The unit probability density is then be defined by

$$\frac{(\pm ikE \pm \mathbf{i}p + jm)}{\sqrt{2E}} \frac{(\pm ikE \mp \mathbf{i}p - jm)}{\sqrt{2E}} = 1,$$

the $1/\sqrt{2E}$ being a normalizing factor. If we assume that such factors automatically apply in calculations, we can also define ($\pm ikE \mp \mathbf{i}p - jm$) = – ($\mp ikE \pm \mathbf{i}p + jm$) as the 'reciprocal' of ($\pm ikE \pm \mathbf{i}p + jm$).

Rather more significantly, the nilpotent formalism not only creates quantum mechanics, but also implies a full quantum field theory in which the operators act on the entire quantum field, without requiring any formal process of second quantization. The transition to quantum field theory seemingly occurs at the point at which we choose to privilege the operator rather than the equation, and then apply Pauli exclusion to all fermionic states, whether free or bound, regardless of the number of interactions to which they are subject. A nilpotent operator, defined in this way from *absolutely nothing*, then becomes a creation operator acting on vacuum to create the fermion, together with all the interactions in which it is involved. As we have seen, no further mathematical formalism is necessary, and neither quantum mechanics nor quantum field theory requires specification by an equation. Once the operator is defined, the phase factor becomes an expression of all the possible variations in space and time which are encoded in the operator, and is uniquely defined with it. A fermion is thus specified as a set of space and time variations, with the mass term a purely passive quantity, and convenient, rather than necessary information. It would be difficult to imagine anything closer to the ideal which the parameter group embodies.

Spin and helicity

The derivation of fermion spin ½ illustrates the multivariate nature of the $\mathbf{p}$ operator. The nilpotent operator ($ikE + \mathbf{i}p + jm$) incorporates a Hamiltonian specified as $\mathcal{H} = -ik(\mathbf{i}p + jm) = -ij\mathbf{p} + iim$. If we *mathematically* define a quantity $\boldsymbol{\sigma} = -\mathbf{1}$ (the pseudovector of magnitude –1 already referred to), then

$$\begin{aligned}[\boldsymbol{\sigma}, \mathcal{H}] &= [-\mathbf{1}, -ij(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3) + iim] = [-\mathbf{1}, -ij(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3)] \\ &= 2ij(\mathbf{ij}p_2 + \mathbf{ik}p_3 + \mathbf{ji}p_1 + \mathbf{jk}p_3 + \mathbf{ki}p_1 + \mathbf{kj}p_2) \\ &= -2j(\mathbf{k}(p_2 - p_1) + \mathbf{j}(p_1 - p_3) + \mathbf{i}(p_3 - p_2)) \\ &= -2j\mathbf{1} \times \mathbf{p}.\end{aligned}$$

If $\mathbf{L}$ is an orbital angular momentum $\mathbf{r} \times \mathbf{p}$, then

$$\begin{aligned}[\mathbf{L}, \mathcal{H}] &= [\mathbf{r} \times \mathbf{p}, -ij\,(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3) + i\mathit{i}m] \\ &= [\mathbf{r} \times \mathbf{p}, -ij\,(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3)] \\ &= i\,[\mathbf{r}, -ij\,(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3)] \times \mathbf{p}\end{aligned}$$

But $$[\mathbf{r}, -ij\,(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3)] = i\mathbf{1}\ .$$

Hence $$[\mathbf{L}, \mathcal{H}] = j\mathbf{1} \times \mathbf{p},$$

and $\mathbf{L} + \boldsymbol{\sigma} / 2$ is a constant of the motion, because

$$[\mathbf{L} + \boldsymbol{\sigma} / 2, \mathcal{H}] = 0.$$

In this formalism, the spin ½ term characteristic of fermionic states emerges purely from the multivariate properties of the $\mathbf{p}$ operator, through the additional cross product term with its imaginary coefficient or pseudovector, exactly as Hestenes showed must result from this algebra (1966). Physically, we observe it as an intrinsic angular momentum term requiring a fermion to undergo a 4π, rather than 2π, rotation to return to its starting point. In our interpretation, this comes from the fact that a localised point-like fermion can only be created simultaneously with a mirror image nonlocalised vacuum state. The fermion on its own provides only half of the knowledge we require to specify the system, and this is equivalent to specifying only one of the two spaces. The system spends only half its time as fermion in real space, and the other half as an antifermion in vacuum space, as in the 4-component (spinor) wavefunction. The spin of fermion plus vacuum is, of course, single-valued (0).

We can define helicity ($\boldsymbol{\sigma}.\mathbf{p}$) as another constant of the motion because

$$[\boldsymbol{\sigma}.\mathbf{p}, \mathcal{H}] = [-p, -ij\,(\mathbf{i}p_1 + \mathbf{j}p_2 + \mathbf{k}p_3) + i\mathit{i}m] = 0$$

We have previously specified that, for a multivariate $\mathbf{p}$,

$$\mathbf{p}\mathbf{p} = (\boldsymbol{\sigma}.\mathbf{p})\,(\boldsymbol{\sigma}.\mathbf{p}) = pp = p^2.$$

So we can also use $\boldsymbol{\sigma}.\mathbf{p}$ ($\boldsymbol{\sigma}\mathbf{p}$) for $\mathbf{p}$ (or $\boldsymbol{\sigma}.\nabla$ ($\boldsymbol{\sigma}\nabla$) for ∇) in the nilpotent operator. As in the graphical representation we have just discussed, a hypothetical fermion / antifermion with zero mass would be reduced to two distinguishable states:

$$\begin{aligned}(ikE + i\,\boldsymbol{\sigma}.\mathbf{p} + jm) &\rightarrow (ikE - ip) \\ (-ikE + i\,\boldsymbol{\sigma}.\mathbf{p} + jm) &\rightarrow (-ikE - ip)\end{aligned}$$

each of which has only a single sign of helicity; $(ikE + ip)$ and $(-ikE + ip)$ are excluded by Pauli exclusion, if we choose the same sign conventions for $\mathbf{p}$. The

previous use of $\boldsymbol{\sigma} = -\mathbf{1}$ in deriving spin for states with positive energy suggests that the allowed spin direction for these states must be antiparallel, corresponding to left-handed helicity, with the negative energy, antistates having right-handed helicity. Numerically, $|\pm E| = p$, so the allowed states can be expressed as $\pm E(\boldsymbol{k} - i\boldsymbol{i})$. As in standard theory, if we multiply from the left by the projection operator $(1 - i\boldsymbol{j}) / 2 \equiv (1 - \gamma^5) / 2$ the allowed states will remain unchanged while the excluded ones are zeroed.

Because spin has emerged in this formalism from the specifically multivariate aspect of the operator $\mathbf{p}$, it is necessary to distinguish equations where the space variables are multivariate from those where they are not, as, for example, when polar coordinates are used. If fermions are point particles and their influence is spherically symmetric, then it will be convenient to express the influence of one point-source on another by changing the coordinates of the 'receiving' particle from Cartesian to polar, with the point-particle source at the centre of the coordinate system also defined as the centre of physical influence.

In such cases, an intrinsic spin is no longer structured into the formalism and an *explicit* spin (or total angular momentum) term has to be introduced. Dirac, however, has given a prescription for translating his equation into polar form, where the momentum operator acquires an additional (imaginary) spin (or total angular momentum) term, and we can easily adapt this to represent a polar transformation of the multivariate vector operator:

$$\nabla \rightarrow \left(\frac{\partial}{\partial r} + \frac{1}{r} \right) \pm i \frac{j + ½}{r} .$$

and use this to define a non-time varying nilpotent operator in polar coordinates:

$$\left(i\boldsymbol{k}E - i\boldsymbol{i}\nabla + \boldsymbol{j}m \right) \rightarrow \left(i\boldsymbol{k}E - i\boldsymbol{i}\left(\frac{\partial}{\partial r} + \frac{1}{r} \pm i \frac{j + ½}{r} \right) + \boldsymbol{j}m \right) .$$

If we take the use of polar coordinates as representing spherical symmetry with respect to a point source, then this operator has no nilpotent solutions unless the $i\boldsymbol{k}E$ term also contains an expression proportional to $1 / r$ to cancel those produced by the $i\boldsymbol{i}\nabla$ term. We recognize this as a Coulomb or inverse-square interaction component. In other words, simply defining a point source forces us to assume that a Coulomb interaction component is necessary for any nilpotent fermion defined with respect to it. All known forces have such components, together with an associated $U(1)$ symmetry. For the gravitational and electric forces, it is the main or complete description; for the strong force it is the one-gluon exchange; for the weak field it is the hypercharge and the B^0 gauge field. Its effect is connected purely with scale or magnitude and we can associate it with the coupling constant. In fundamental terms it arises because defining a point in any meaningful way in 3-dimensional space

requires a dual space which is structured on the basis of point charges. In addition, the only way of *fixing* a point in a nonconserved space with no identifiable units is to fix it in the space of a conserved quantity which is made to coincide, through nilpotency, with this one. By making this Coulomb component a consequence of nilpotency, we can also see it as a consequence of Pauli exclusion.

***Zitterbewegung* and Berry phase**

Originally, Dirac used $\boldsymbol{\alpha}$ and β operators in his equation, rather than $\boldsymbol{\gamma}$ and γ^0; $\boldsymbol{\alpha}$ was equivalent to $\boldsymbol{\gamma}\gamma^0$ and β to γ^0. So far, the constants $\hbar$ and c have been left out of our equations, but now it will be convenient to include them. Using these different conventions on symbolism, the nilpotent Hamiltonian becomes

$$\mathcal{H} = -\mathbf{ij}c\boldsymbol{\sigma}.\mathbf{p} - \mathbf{ii}mc^2 = -\mathbf{ij}c\mathbf{1}\mathbf{p} - - \mathbf{ii}mc^2 = \boldsymbol{\alpha}c\mathbf{p} - \mathbf{ii}mc^2.$$

With four separate spin states in the system, $\boldsymbol{\alpha} = -\mathbf{ij1}$ can be taken as a dynamical variable, and $\boldsymbol{\alpha}c = -\mathbf{ij1}c$ defined, in terms of the discrete commutator calculus, as a velocity operator, which, for a free particle, becomes:

$$\mathbf{v} = \dot{\mathbf{r}} = \frac{d\mathbf{r}}{dt} = \frac{1}{i\hbar}\,[\mathbf{r}, \mathcal{H}] = -\mathbf{ij1}c = c\boldsymbol{\alpha}\,.$$

Because we are now using a velocity operator, dF/dt must be distinguished from $\partial F/\partial t$. The equation of motion for the velocity operator then becomes:

$$\frac{d\boldsymbol{\alpha}}{dt} = \frac{1}{i\hbar}\,[\boldsymbol{\alpha}, \mathcal{H}] = \frac{2}{i\hbar}(c\mathbf{p} - \mathcal{H}\boldsymbol{\alpha}).$$

Schrödinger's solution of this well-known result first gave the equation of motion for the fermion from the Dirac equation:

$$\mathbf{r}(t) = \mathbf{r}(0) + \frac{c^2\mathbf{p}}{\mathcal{H}}\,t + \frac{\hbar\, c}{2i\,\mathcal{H}}\,[\boldsymbol{\alpha}(0) - c\mathcal{H}^{-1}\mathbf{p}](exp\,(2i\mathcal{H}t/h) - 1).$$

Though we have nothing to add to the derivation of this result, the nilpotent interpretation is important for our fundamental view of physics. The equation has classical analogues for all its terms, except the third, which Schrödinger interpreted as predicting a violent oscillatory motion or high-frequency vibration (which he called *zitterbewegung*) of the particle at frequency $\approx 2mc^2/\hbar$, and amplitude $\hbar/2mc$, which is related to the Compton wavelength for the particle and directly determined by the particle's rest mass.

The *zitterbewegung*, which can only be non-zero between states with equal momentum and opposite energy, is interpreted as a switching between the four fermionic states, which incorporate matter and antimatter and two directions of spin. It is certainly a vacuum effect, and we can interpret it as a continual redefinition of the localised fermionic state in relation to the nonlocal vacuum, without which it could not be defined at a point, and so an expression of the necessity of dual vector spaces in the description of a discrete particle. It is, in fact, a direct expression of the duality of these spaces. If two ways of presenting the same information can exist, then they will become equally probable and Nature will not privilege one over the other, though one may seem more convenient from our perspective.

Dirac saw *zitterbewegung* as implying that a fermion (or any massive particle) actually propagates along the light cone, oscillating between $+c$ and $-c$ at a frequency which determines its measured mass and momentum. According to this argument, a measured value of velocity can only be found by knowing positions at two different times. To find the instantaneous velocity, you have to reduce the time interval to zero, thus fixing the positions with exact precision, and so making the momentum value completely indeterminate according to Heisenberg uncertainty. The ultimate significance of *zitterbewegung* in this context may be that it locates rest mass as the result of defining a singularity.

Zitterbewegung can thus be seen as an intrinsic aspect of defining a fermion as a point-singularity through the nilpotent structure created by dual vector spaces. This brings us back to our topologies of simply- and multiply-connected spaces and the Berry phase (lecture 2). Some Berry phase phenomena involve a fermion with half-integral spin subjected to a cyclic adiabatic (i.e. effectively non-dissipative) process becoming single valued (i.e. with integer spin) in the presence of either another fermionic state, for example, an electron (Cooper pairing) or nucleus (Jahn-Teller effect), or an 'environment' whose origin is ultimately fermionic. This last could be, for example a vector potential (Aharonov-Bohm effect) or a magnetic flux line (quantum Hall effect). In each of these cases the Berry phase can be interpreted topologically, as in lecture 2, with the initial fermion travelling in a space that has changed from being simply- to multiply-connected by incorporating the other fermionic state or environment as a 'singularity'.

As previously mentioned, we can regard the unpaired fermion, defined as a pure physical singularity, as existing in its *own* multiply-connected space and thus naturally becoming a spin ½ particle. Experimental evidence to date suggests that fermions are point-like, and in this sense singularities; excluding anything that might be produced by gravity, and bosons as the products of fermion interactions, they are the only known physical singularities. Now, a physical singularity can only be defined with reference to a nonlocalised phase. The phase factor, as we have seen, contains all

the information about the singularity but extends everywhere, and overlaps with all other phase factors. It is this which enables two seemingly isolated singularities to interact, and which allows us to describe such interactions in terms of a quantum field. This will become more explicit when we do specific examples in the next lecture. Information from the dual spaces of one system (potentials or even distortions of its space-time structure) creates changes in the dual spaces of the other, via changes in the E and $\mathbf{p}$ terms of its operator, and, through the phase factor, of its amplitude. Even a pure vector potential (as in the Aharonov-Bohm effect, where it is created by a solenoid with no external field) will alter the $\mathbf{p}$ term and so produce these changes. Under cyclic adiabatic conditions, we can consider the E and p magnitudes of the combination to be equalised as in the formation of the bosonic-type states we will consider in the next lecture.

CPT Symmetry

If the lead term in the fermionic column vector, defines the fermion type, then we can show that the remaining terms are equivalent to the lead term, subjected to the respective symmetry transformations, P, T and C, by pre- and post-multiplication by the quaternion units $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ defining the *vacuum space*:

Parity	P	$\boldsymbol{i}\,(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{i} = (\pm i\boldsymbol{k}E \mp \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$
Time reversal	T	$\boldsymbol{k}\,(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{k} = (\mp i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$
Charge conjugation	C	$-\boldsymbol{j}\,(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{j} = (\mp i\boldsymbol{k}E \mp \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)$

The charge conjugation process could equally be represented by

Charge conjugation	C	$\boldsymbol{ij}\,(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{ij} = (\mp i\boldsymbol{k}E \mp \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m),$

showing its ultimate origin in a vector space where $\mathbf{ij} = i\mathbf{k}$. We can easily show by trying it out that the rules

$$CP \equiv T,\ PT \equiv C, \text{ and } CT \equiv P$$

necessarily apply, as also

$$TCP \equiv\ CPT \equiv \text{ identity}$$

as

$$\boldsymbol{k}\,(\boldsymbol{j}\,(\boldsymbol{i}\,(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{i})\,\boldsymbol{j})\boldsymbol{k} = \boldsymbol{kji}\,(\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m)\,\boldsymbol{ijk} = (\pm i\boldsymbol{k}E \pm \boldsymbol{i}\mathbf{p} + \boldsymbol{j}m).$$

It is clear that charge conjugation (or exchange of particle and antiparticle) is effectively defined in terms of parity and time reversal, rather than being an independent operation. This comes from the fact that only space and time are active elements, with their variation being the coded information that solely determines the phase factor and the entire nature of the fermion state, while the mass term (which connects with the charge conjugation transformation) is a passive element, which we

have shown can even be excluded from the operator without loss of information. The construction of a nilpotent amplitude effectively requires the loss of a sign degree of freedom in one component, E, $\mathbf{p}$ or m, and the passivity of mass makes it the term to which this will apply.

The *CPT* theorem is based on a combination of relativity and causality. Relativity says that the square of $(\pm i\mathbf{k}E \pm i\mathbf{p})$ or its conjugate, $(\pm i\mathbf{k}t \pm i\mathbf{r})$, is an invariant, and it is only when $\mathbf{j}m$ or the equivalent $\mathbf{j}\tau$ is added that we also get causality. Relativity and causality combined require a structure with $\mathbf{k}$, $\mathbf{i}$ and $\mathbf{j}$ on the same footing, and this is also a requirement for *CPT* symmetry. *C*, *P* and *T* are the symmetries concerning the algebraic signs of three of the four fundamental parameters, and so the *CPT* theorem is one of the most fundamental in physics. If the fourth parameter, mass, was not unipolar, there would be an *MCPT* theorem.

Note to p. 9

[1] The four idempotent terms in the column vector $\mathbf{j}(\pm i\mathbf{k}E \pm i\mathbf{p} + \mathbf{j}m)$ sum up to unity after normalization and, if combined with the appropriate quaternionic coefficients identifying their positions in the column, produce zero products between any two terms as required (which, in principle, demonstrates their intrinsic nilpotency). The nilpotent vector $(\pm i\mathbf{k}E \pm i\mathbf{p} + \mathbf{j}m)$ also has this intrinsic idempotent property, by absorbing the initial factor $\mathbf{j}$ into the quaternion coefficients.

Peter Rowlands
Physics Department, University of Liverpool
15 April 2013