Discovery of Dirac equation by Dirac

Masatsugu Sei Suzuki
Department of Physics, SUNY at Binghamton
(Date: April 16, 2017)

I read a book written by H.S. Kragh, titled as Dirac Scientific Biography. I was very impressed with the discovery of Dirac equation by Dirac. His paper on the quantum theory of the electron by Dirac was published by Proceedings of the Royal Society (received on January 02, 1928). Dirac's theory was a product of his emerging general philosophy of physics. He wanted the theory to be founded on general principles rather than on any particular model of the electron. Dirac was guided by two invariance requirements: first, the space-time properties of the equation should transform according to the theory of relativity. Second, the quantum properties should transform according to the transformation theory of quantum mechanics.

For simplicity we do not use any notations which are conventionally used for the discussion of relativity such as covariant and contravariant vector components.). Here the discovery of Dirac equation is summarized, based on the book of Kragh.

1. Discovery of Dirac equation (by Dirac, Kragh 1990)

First Dirac considered the quantity $(\boldsymbol{\sigma}\cdot\boldsymbol{p})$, where $\boldsymbol{\sigma}$ is the Pauli spin operator;

$$\sigma_1 = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Dirac found a circle in an identity he had noticed when he played with Pauli spin matrices, namely

$$\sqrt{p_1^{\,2} + p_2^{\,2} + p_3^{\,2}} = \sigma_1 p_1 + \sigma_2 p_2 + \sigma_3 p_3 = \boldsymbol{\sigma}\cdot\boldsymbol{p} \qquad (1)$$

or

$$(\boldsymbol{\sigma}\cdot\boldsymbol{p})^2 = \boldsymbol{p}^2$$

Note that

$$(\hat{\boldsymbol{\sigma}}\cdot\boldsymbol{A})(\hat{\boldsymbol{\sigma}}\cdot\boldsymbol{B}) = (\boldsymbol{A}\cdot\boldsymbol{B})\hat{1} + i\hat{\boldsymbol{\sigma}}\cdot(\boldsymbol{A}\times\boldsymbol{B}) \qquad \text{(formula)}$$

with ***A*** and ***B*** are arbitrary vector, or directly, we can derive as follows.

$$
\begin{aligned}
(\boldsymbol{\sigma}\cdot\boldsymbol{p})^2 &= (\sigma_1 p_1 + \sigma_2 p_2 + \sigma_3 p_3)(\sigma_1 p_1 + \sigma_2 p_2 + \sigma_3 p_3) \\
&= \sigma_1 p_1 \sigma_1 p_1 + \sigma_1 p_1 \sigma_2 p_2 + \sigma_1 p_1 \sigma_3 p_3 \\
&\quad + \sigma_2 p_2 \sigma_1 p_1 + \sigma_2 p_2 \sigma_2 p_2 + \sigma_2 p_2 \sigma_3 p_3 \\
&\quad + \sigma_3 p_3 \sigma_1 p_1 + \sigma_3 p_3 \sigma_2 p_2 + \sigma_3 p_3 \sigma_3 p_3 \\
&= \sigma_1^{\ 2} p_1^{\ 2} + \sigma_2^{\ 2} p_2^{\ 2} + \sigma_3^{\ 2} p_3 + (\sigma_1\sigma_2 + \sigma_2\sigma_1) p_1 p_2 \\
&\quad + (\sigma_2\sigma_3 + \sigma_3\sigma_2) p_2 p_3 + (\sigma_3\sigma_1 + \sigma_1\sigma_3) p_2 p_3 \\
&= p_1^{\ 2} + p_2^{\ 2} + p_3^{\ 2} \\
&= \boldsymbol{p}^2
\end{aligned}
$$

since

$$\sigma_1^{\ 2} = \sigma_2^{\ 2} = \sigma_3^{\ 2} = 1$$

$$\sigma_1\sigma_2 + \sigma_2\sigma_1 = 0\,, \qquad \sigma_2\sigma_3 + \sigma_3\sigma_2 = 0\,, \qquad \sigma_3\sigma_1 + \sigma_1\sigma_3 = 0$$

At that stage Dirac may have tried to use the quantity $\boldsymbol{\sigma}\cdot\boldsymbol{p}$ as the Hamiltonian in a wave equation; that is, he may have considered

$$(\boldsymbol{\sigma}\cdot\boldsymbol{p})\psi = i\hbar\frac{\partial}{\partial t}\psi \tag{2}$$

where ψ is a two-component wave function. This equation is Lorentz invariant, contains a spin of one-half, and is of first order in the time derivative. But it does not contain a mass term, it does not apply to electron. Hence Dirac had to reconsider the possible significance of Eq.(1): If it could be generated to four squares instead of two, it would obviously indicate a solution; for then linearization of the type wanted

$$\sqrt{p_1^{\ 2} + p_2^{\ 2} + p_3^{\ 2} + m^2c^2} = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta mc\,, \tag{3}$$

But were there coefficients with this property, and if so what did they look like? Dirac argued the linear wave equation as provisionally given by Eqs.(1) and (3), has to contain the KG (Klein-Gordon) equation as its square. In this way he was able to

$$
\begin{aligned}
&(\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta mc)^2 \\
&= (\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta mc)(\alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 + \beta mc) \\
&= \alpha_1^{\ 2} p_1^{\ 2} + +\alpha_2^{\ 2} p_2^{\ 2} + \alpha_3^{\ 2} p_3^{\ 2} + \beta^2 m^2 c^2 + (\alpha_1\alpha_2 + \alpha_2\alpha_1) p_1 p_2 + (\alpha_2\alpha_3 + \alpha_3\alpha_2) p_2 p_3 + \\
&+ (\alpha_3\alpha_3 + \alpha_3\alpha_2) p_2 p_3 + (\alpha_1\beta + \beta\alpha_1) mcp_1 + (\alpha_2\beta + \beta\alpha_2) mcp_2 + (\alpha_3\beta + \beta\alpha_3) mcp_3 \\
&= p_1^{\ 2} + p_2^{\ 2} + p_3^{\ 2} + m^2c^2
\end{aligned}
$$

where $\mu, \nu = 0,1,2,3$

$$\{\alpha_\mu, \alpha_\nu\} = [\alpha_\mu, \alpha_\nu]_+ = 0 \qquad \text{for} \qquad \mu \neq \nu$$

$$(4)$$

$$\alpha_\mu{}^2 = 1$$

For convenience we use $\beta = \alpha_0$. Dirac knew that a set of similar conditions are fulfilled by the spin matrices, of which there are, however, only three. So he naturally tried to take

$$\alpha_j = \sigma_j$$

and sought for another 2x2 candidate for β. However, such a candidate does not exist, and Dirac realized that working with 2x2 matrices just would not do. Then he got again one of those invaluable ideas out of the blue.

"I (Dirac) suddenly realized that there was no need to stick to quantities, which can be represented by matrices with just two rows and columns. Why not go four rows and columns (4x4)? This idea solved the problem, and found the explicit form of the α matrices.

Dirac matrices (4x4 matrices)

$\alpha_x, \alpha_y, \alpha_z$: odd matrices

β: even matrx.

$$\alpha_x = \alpha_1 = \sigma_1 \otimes \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \sigma_1 \\ \sigma_1 & 0 \end{pmatrix}$$

$$\alpha_y = \alpha_2 = \sigma_1 \otimes \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i1 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \sigma_2 \\ \sigma_2 & 0 \end{pmatrix}$$

$$\alpha_z = \alpha_3 = \sigma_1 \otimes \sigma_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \sigma_3 \\ \sigma_3 & 0 \end{pmatrix}$$

$$\beta = \alpha_0 = \sigma_3 \otimes I_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}$$

For $i \neq j$ (where *i*, *j* = 1, 2, 3)

$$\begin{aligned} \alpha_i \alpha_j + \alpha_j \alpha_i &= \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \begin{pmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{pmatrix} + \begin{pmatrix} 0 & \sigma_j \\ \sigma_j & 0 \end{pmatrix} \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_i \sigma_i + \sigma_i \sigma_i & 0 \\ 0 & \sigma_i \sigma_i + \sigma_i \sigma_i \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \alpha_i^2 &= \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \\ &= \begin{pmatrix} \sigma_i^2 & 0 \\ 0 & \sigma_i^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= 1_4 \end{aligned}$$

For *i* (=1, 2, 3),

$$\begin{aligned} \alpha_i \alpha_0 + \alpha_4 \alpha_0 &= \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix} + \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix} \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -\sigma_i I_2 \\ \sigma_i I_2 & 0 \end{pmatrix} + \begin{pmatrix} 0 & \sigma_i I_2 \\ -\sigma_i I_2 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

With the linearization successfully carried out, the ice was broken. The next stage to formulate the wave equation for a free electron – was easy. Equations (1) and (3) immediately yielded

$$i\hbar \frac{\partial}{c\partial t}\psi = (\boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta mc)\psi \qquad \text{or} \qquad i\hbar \frac{\partial}{\partial t}\psi = [c(\boldsymbol{\alpha}\cdot\boldsymbol{p}) + \beta mc^2]\psi$$

which is known as the Dirac equation.

Dirac reduced a physical problem to a mathematical one, and mathematics forced him to accept the use of 4x4 matrices as coefficients. This again forced him to accept a four-component wave function. Though logical enough this was a bold proposal since there was no physical justification for the two extra components.

2. Spin of electron and antiparticle (antiparticle)

It was validated by accounting for the fine details of the hydrogen spectrum in a completely rigorous way. Although Dirac did not at first fully appreciate the importance of his results, the entailed explanation of spin as a consequence of the union of quantum mechanics and relativity—and the eventual discovery of the positron (antiparticle of electron)—represents one of the great triumphs of theoretical physics.

From his theory in the presence of an external magnetic field, Dirac was able to deduce the correct spin from the first principle upon which his equation was built. This was a great and unexpected triumph.

$$\boldsymbol{\mu} = -\mu_B \boldsymbol{\sigma}$$

where $\mu_B = \dfrac{e\hbar}{2mc}$.

REFERENCES

H.S. Kragh, Dirac Scientific Biography (Cambridge, 1990).