Dirac equation

From Wikipedia, the free encyclopedia

In particle physics, the **Dirac equation** is a relativistic wave equation derived by British physicist Paul Dirac in 1928. In its free form, or including electromagnetic interactions, it describes all spin-$\frac{1}{2}$ massive particles such as electrons and quarks for which parity is a symmetry. It is consistent with both the principles of quantum mechanics and the theory of special relativity,[1] and was the first theory to account fully for special relativity in the context of quantum mechanics. It was validated by accounting for the fine details of the hydrogen spectrum in a completely rigorous way.

The equation also implied the existence of a new form of matter, *antimatter*, previously unsuspected and unobserved and which was experimentally confirmed several years later. It also provided a *theoretical* justification for the introduction of several component wave functions in Pauli's phenomenological theory of spin; the wave functions in the Dirac theory are vectors of four complex numbers (known as bispinors), two of which resemble the Pauli wavefunction in the non-relativistic limit, in contrast to the Schrödinger equation which described wave functions of only one complex value. Moreover, in the limit of zero mass, the Dirac equation reduces to the Weyl equation.

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—represents one of the great triumphs of theoretical physics. This accomplishment has been described as fully on a par with the works of Newton, Maxwell, and Einstein before him.[2] In the context of quantum field theory, the Dirac equation is reinterpreted to describe quantum fields corresponding to spin-$\frac{1}{2}$ particles.

Contents

- 1 Mathematical formulation
  - 1.1 Making the Schrödinger equation relativistic
  - 1.2 Dirac's coup
  - 1.3 Covariant form and relativistic invariance
  - 1.4 Conservation of probability current
  - 1.5 Solutions
  - 1.6 Comparison with the Pauli theory
  - 1.7 Comparison with the Weyl theory
  - 1.8 Dirac Lagrangian
- 2 Physical interpretation
  - 2.1 Identification of observables
  - 2.2 Hole theory
  - 2.3 In quantum field theory
- 3 Other formulations
  - 3.1 As a differential equation in one real component
  - 3.2 Curved spacetime
  - 3.3 The algebra of physical space
- 4 See also

- 4.1 Articles on the Dirac equation
- 4.2 Other equations
- 4.3 Other topics
- 5 References
  - 5.1 Selected papers
  - 5.2 Textbooks
- 6 External links

Mathematical formulation

The Dirac equation in the form originally proposed by Dirac is:[3]

Dirac equation *(original)*

$$\left(\beta mc^2 + c\left(\sum_{n=1}^{3}\alpha_n p_n\right)\right)\psi(x,t) = i\hbar\frac{\partial\psi(x,t)}{\partial t}$$

where $\psi = \psi(x, t)$ is the wave function for the electron of rest mass m with spacetime coordinates x, t. The p_1, p_2, p_3 are the components of the momentum, understood to be the momentum operator in the Schrödinger equation. Also, c is the speed of light, and $\hbar$ is the Planck constant divided by 2π. These fundamental physical constants reflect special relativity and quantum mechanics, respectively.

Dirac's purpose in casting this equation was to explain the behavior of the relativistically moving electron, and so to allow the atom to be treated in a manner consistent with relativity. His rather modest hope was that the corrections introduced this way might have a bearing on the problem of atomic spectra. Up until that time, attempts to make the old quantum theory of the atom compatible with the theory of relativity, attempts based on discretizing the angular momentum stored in the electron's possibly non-circular orbit of the atomic nucleus, had failed – and the new quantum mechanics of Heisenberg, Pauli, Jordan, Schrödinger, and Dirac himself had not developed sufficiently to treat this problem. Although Dirac's original intentions were satisfied, his equation had far deeper implications for the structure of matter and introduced new mathematical classes of objects that are now essential elements of fundamental physics.

The new elements in this equation are the 4×4 matrices α_k and β, and the four-component wave function ψ. There are four components in ψ because the evaluation of it at any given point in configuration space is a bispinor. It is interpreted as a superposition of a spin-up electron, a spin-down electron, a spin-up positron, and a spin-down positron (see below for further discussion).

The 4×4 matrices α_k and β are all Hermitian and have squares equal to the identity matrix:

$$\alpha_i^2 = \beta^2 = I_4$$

and they all mutually anticommute (if i and j are distinct):

$$\alpha_i\alpha_j + \alpha_j\alpha_i = 0$$
$$\alpha_i\beta + \beta\alpha_i = 0$$

The single symbolic equation thus unravels into four coupled linear first-order partial differential equations for the four quantities that make up the wave function. These matrices and the form of the wave function have a deep mathematical significance. The algebraic structure represented by the gamma matrices had been created some 50 years earlier by the English mathematician W. K. Clifford. In turn, Clifford's ideas had emerged from the mid-19th-century work of the German mathematician Hermann Grassmann in his *Lineale Ausdehnungslehre* (*Theory of Linear Extensions*). The latter had been regarded as well-nigh incomprehensible by most of his contemporaries. The appearance of something so seemingly abstract, at such a late date, and in such a direct physical manner, is one of the most remarkable chapters in the history of physics.

Making the Schrödinger equation relativistic

The Dirac equation is superficially similar to the Schrödinger equation for a massive free particle:

$$-\frac{\hbar^2}{2m}\nabla^2\phi = i\hbar\frac{\partial}{\partial t}\phi.$$

The left side represents the square of the momentum operator divided by twice the mass, which is the non-relativistic kinetic energy. Because relativity treats space and time as a whole, a relativistic generalization of this equation requires that space and time derivatives must enter symmetrically as they do in the Maxwell equations that govern the behavior of light — the equations must be differentially of the *same order* in space and time. In relativity, the momentum and the energies are the space and time parts of a spacetime vector, the four-momentum, and they are related by the relativistically invariant relation

$$E^2 = m^2c^4 + p^2c^2$$

which says that the length of this four-vector is proportional to the rest mass m. Substituting the operator equivalents of the energy and momentum from the Schrödinger theory, we get the Klein-Gordon equation describing the propagation of waves, constructed from relativistically invariant objects,

$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right)\phi = \frac{m^2c^2}{\hbar^2}\phi$$

with the wave function ϕ being a relativistic scalar: a complex number which has the same numerical value in all frames of reference. Space and time derivatives both enter to second order. This has a telling consequence for the interpretation of the equation. Because the equation is second order in the time derivative, one must specify initial values both of the wave function itself and of its first-time derivative in order to solve definite problems. Since both may be specified more or less arbitrarily, the wave function cannot maintain its former role of determining the probability density of finding the electron in a given state of motion. In the Schrödinger theory, the probability density is given by the positive definite expression

$$\rho = \phi^*\phi$$

and this density is convected according to the probability current vector

$$J = -\frac{i\hbar}{2m}(\phi^* \nabla \phi - \phi \nabla \phi^*)$$

with the conservation of probability current and density following from the continuity equation:

$$\nabla \cdot J + \frac{\partial \rho}{\partial t} = 0.$$

The fact that the density is positive definite and convected according to this continuity equation implies that we may integrate the density over a certain domain and set the total to 1, and this condition will be maintained by the conservation law. A proper relativistic theory with a probability density current must also share this feature. Now, if we wish to maintain the notion of a convected density, then we must generalize the Schrödinger expression of the density and current so that space and time derivatives again enter symmetrically in relation to the scalar wave function. We are allowed to keep the Schrödinger expression for the current, but must replace the probability density by the symmetrically formed expression

$$\rho = \frac{i\hbar}{2mc^2}(\psi^* \partial_t \psi - \psi \partial_t \psi^*).$$

which now becomes the 4th component of a spacetime vector, and the entire probability 4-current density has the relativistically covariant expression

$$J^\mu = \frac{i\hbar}{2m}(\psi^* \partial^\mu \psi - \psi \partial^\mu \psi^*)$$

The continuity equation is as before. Everything is compatible with relativity now, but we see immediately that the expression for the density is no longer positive definite – the initial values of both ψ and $\partial_t \psi$ may be freely chosen, and the density may thus become negative, something that is impossible for a legitimate probability density. Thus, we cannot get a simple generalization of the Schrödinger equation under the naive assumption that the wave function is a relativistic scalar, and the equation it satisfies, second order in time.

Although it is not a successful relativistic generalization of the Schrödinger equation, this equation is resurrected in the context of quantum field theory, where it is known as the Klein–Gordon equation, and describes a spinless particle field (e.g. pi meson). Historically, Schrödinger himself arrived at this equation before the one that bears his name but soon discarded it. In the context of quantum field theory, the indefinite density is understood to correspond to the *charge* density, which can be positive or negative, and not the probability density.

Dirac's coup

Dirac thus thought to try an equation that was *first order* in both space and time. One could, for example, formally (i.e. by abuse of notation) take the relativistic expression for the energy

$$E = c\sqrt{p^2 + m^2 c^2}\,,$$

replace p by its operator equivalent, expand the square root in an infinite series of derivative operators, set up an eigenvalue problem, then solve the equation formally by iterations. Most physicists had little faith in such a process, even if it were technically possible.

As the story goes, Dirac was staring into the fireplace at Cambridge, pondering this problem, when he hit upon the idea of taking the square root of the wave operator thus:

$$\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} = \left(A\partial_x + B\partial_y + C\partial_z + \frac{i}{c}D\partial_t\right)\left(A\partial_x + B\partial_y + C\partial_z + \frac{i}{c}D\partial_t\right).$$

On multiplying out the right side we see that, in order to get all the cross-terms such as $\partial_x\partial_y$ to vanish, we must assume

$$AB + BA = 0,\ \ldots$$

with

$$A^2 = B^2 = \ldots = 1.$$

Dirac, who had just then been intensely involved with working out the foundations of Heisenberg's matrix mechanics, immediately understood that these conditions could be met if A, B, C and D are *matrices*, with the implication that the wave function has *multiple components*. This immediately explained the appearance of two-component wave functions in Pauli's phenomenological theory of spin, something that up until then had been regarded as mysterious, even to Pauli himself. However, one needs at least 4 × 4 matrices to set up a system with the properties required — so the wave function had *four* components, not two, as in the Pauli theory, or one, as in the bare Schrödinger theory. The four-component wave function represents a new class of mathematical object in physical theories that makes its first appearance here.

Given the factorization in terms of these matrices, one can now write down immediately an equation

$$\left(A\partial_x + B\partial_y + C\partial_z + \frac{i}{c}D\partial_t\right)\psi = \kappa\psi$$

with κ to be determined. Applying again the matrix operator on both sides yields

$$\left(\nabla^2 - \frac{1}{c^2}\partial_t^2\right)\psi = \kappa^2\psi.$$

On taking $\kappa = \frac{mc}{\hbar}$ we find that all the components of the wave function *individually* satisfy the relativistic energy–momentum relation. Thus the sought-for equation that is first-order in both space and time is

$$\left(A\partial_x + B\partial_y + C\partial_z + \frac{i}{c}D\partial_t - \frac{mc}{\hbar}\right)\psi = 0.$$

Setting

$$A = i\beta\alpha_1\,,\ B = i\beta\alpha_2\,,\ C = i\beta\alpha_3\,,\ D = \beta\,,$$

and because $D^2 = \beta^2 = I_4\,,$

we get the Dirac equation as written above.

Covariant form and relativistic invariance

To demonstrate the relativistic invariance of the equation, it is advantageous to cast it into a form in which the space and time derivatives appear on an equal footing. New matrices are introduced as follows:

$$D = \gamma^0\,,$$
$$A = i\gamma^1\,,\ B = i\gamma^2\,,\ C = i\gamma^3\,,$$

and the equation takes the form (remembering the definition of the covariant components of the 4-gradient and especially that $\partial_0 = \frac{1}{c}\partial_t$)

Dirac equation

$$i\hbar\gamma^\mu\partial_\mu\psi - mc\psi = 0$$

where there is an implied summation over the values of the twice-repeated index $\mu = 0, 1, 2, 3,$ and ∂_μ is the 4-gradient. In practice one often writes the gamma matrices in terms of 2 × 2 sub-matrices taken from the Pauli matrices and the 2 × 2 identity matrix. Explicitly the standard representation is

$$\gamma^0 = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \gamma^1 = \begin{pmatrix} 0 & \sigma_x \\ -\sigma_x & 0 \end{pmatrix}, \gamma^2 = \begin{pmatrix} 0 & \sigma_y \\ -\sigma_y & 0 \end{pmatrix}, \gamma^3 = \begin{pmatrix} 0 & \sigma_z \\ -\sigma_z & 0 \end{pmatrix}.$$

The complete system is summarized using the Minkowski metric on spacetime in the form

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu} I_4$$

where the bracket expression

$$\{a, b\} = ab + ba$$

denotes the anticommutator. These are the defining relations of a Clifford algebra over a pseudo-orthogonal 4-dimensional space with metric signature $(+ - - -)$. The specific Clifford algebra employed in the Dirac equation is known today as the Dirac algebra. Although not recognized as such by Dirac at the time the equation was formulated, in hindsight the introduction of this *geometric algebra* represents an enormous stride forward in the development of quantum theory.

The Dirac equation may now be interpreted as an eigenvalue equation, where the rest mass is proportional to an eigenvalue of the 4-momentum operator, the proportionality constant being the speed of light:

$$P_{\mathrm{op}}\psi = mc\psi.$$

Using $\not{\partial}$ (pronounced: "d-slash"[4]) in Feynman slash notation, which includes the gamma matrices as well as a summation over the spinor components in the derivative itself, the Dirac equation becomes:

$$i\hbar\not{\partial}\psi - mc\psi = 0$$

In practice, physicists often use units of measure such that $\hbar = c = 1$, known as natural units. The equation then takes the simple form

Dirac equation *(natural units)*

$$(i\not{\partial} - m)\psi = 0$$

A fundamental theorem states that if two distinct sets of matrices are given that both satisfy the Clifford relations, then they are connected to each other by a similarity transformation:

$$\gamma^{\mu\prime} = S^{-1}\gamma^{\mu}S.$$

If in addition the matrices are all unitary, as are the Dirac set, then S itself is unitary;

$$\gamma^{\mu\prime} = U^{\dagger}\gamma^{\mu}U.$$

The transformation U is unique up to a multiplicative factor of absolute value 1. Let us now imagine a Lorentz transformation to have been performed on the space and time coordinates, and on the derivative operators, which form a covariant vector. For the operator $\gamma^{\mu}\partial_{\mu}$ to remain invariant, the gammas must transform among themselves as a contravariant vector with respect to their spacetime index. These new gammas will themselves satisfy the Clifford relations, because of the orthogonality of the Lorentz transformation. By the fundamental theorem, we may replace the new set by the old set subject to a unitary transformation. In the new frame, remembering that the rest mass is a relativistic scalar, the Dirac equation will then take the form

$$(iU^{\dagger}\gamma^{\mu}U\partial'_{\mu} - m)\psi(x', t') = 0$$
$$U^{\dagger}(i\gamma^{\mu}\partial'_{\mu} - m)U\psi(x', t') = 0.$$

If we now define the transformed spinor

$$\psi' = U\psi$$

then we have the transformed Dirac equation in a way that demonstrates manifest relativistic invariance:

$$(i\gamma^{\mu}\partial'_{\mu} - m)\psi'(x', t') = 0.$$

Thus, once we settle on any unitary representation of the gammas, it is final provided we transform the spinor according to the unitary transformation that corresponds to the

given Lorentz transformation. The various representations of the Dirac matrices employed will bring into focus particular aspects of the physical content in the Dirac wave function (see below). The representation shown here is known as the *standard* representation – in it, the wave function's upper two components go over into Pauli's 2-spinor wave function in the limit of low energies and small velocities in comparison to light.

The considerations above reveal the origin of the gammas in *geometry*, hearkening back to Grassmann's original motivation – they represent a fixed basis of unit vectors in spacetime. Similarly, products of the gammas such as $\gamma_\mu\gamma_\nu$ represent *oriented surface elements*, and so on. With this in mind, we can find the form of the unit volume element on spacetime in terms of the gammas as follows. By definition, it is

$$V = \frac{1}{4!}\epsilon_{\mu\nu\alpha\beta}\gamma^\mu\gamma^\nu\gamma^\alpha\gamma^\beta.$$

For this to be an invariant, the epsilon symbol must be a tensor, and so must contain a factor of $\sqrt{g}$, where g is the determinant of the metric tensor. Since this is negative, that factor is *imaginary*. Thus

$$V = i\gamma^0\gamma^1\gamma^2\gamma^3.$$

This matrix is given the special symbol γ^5, owing to its importance when one is considering improper transformations of spacetime, that is, those that change the orientation of the basis vectors. In the standard representation, it is

$$\gamma_5 = \begin{pmatrix} 0 & I_2 \\ I_2 & 0 \end{pmatrix}.$$

This matrix will also be found to anticommute with the other four Dirac matrices:

$$\gamma^5\gamma^\mu + \gamma^\mu\gamma^5 = 0$$

It takes a leading role when questions of *parity* arise because the volume element as a directed magnitude changes sign under a spacetime reflection. Taking the positive square root above thus amounts to choosing a handedness convention on spacetime .

Conservation of probability current

By defining the adjoint spinor

$$\bar{\psi} = \psi^\dagger\gamma^0$$

where $\psi^\dagger$ is the conjugate transpose of ψ, and noticing that

$$(\gamma^\mu)^\dagger\gamma^0 = \gamma^0\gamma^\mu\,,$$

we obtain, by taking the Hermitian conjugate of the Dirac equation and multiplying from the right by γ^0, the adjoint equation:

$$\bar{\psi}(-i\gamma^\mu\partial_\mu - m) = 0$$

where ∂_μ is understood to act to the left. Multiplying the Dirac equation by ψ from the left, and the adjoint equation by ψ from the right, and subtracting, produces the law of conservation of the Dirac current:

$$\partial_\mu \left(\bar{\psi}\gamma^\mu\psi\right) = 0.$$

Now we see the great advantage of the first-order equation over the one Schrödinger had tried – this is the conserved current density required by relativistic invariance, only now its 4th component is *positive definite* and thus suitable for the role of a probability density:

$$J^0 = \bar{\psi}\gamma^0\psi = \psi^\dagger\psi.$$

Because the probability density now appears as the fourth component of a relativistic vector and not a simple scalar as in the Schrödinger equation, it will be subject to the usual effects of the Lorentz transformations such as time dilation. Thus, for example, atomic processes that are observed as rates, will necessarily be adjusted in a way consistent with relativity, while those involving the measurement of energy and momentum, which themselves form a relativistic vector, will undergo parallel adjustment which preserves the relativistic covariance of the observed values.

Solutions

See Dirac spinor for details of solutions to the Dirac equation. Note that since the Dirac operator acts on 4-tuples of square-integrable functions, its solutions should be members of the same Hilbert space. The fact that the energies of the solutions do not have a lower bound is unexpected – see the hole theory section below for more details.

Comparison with the Pauli theory

The necessity of introducing half-integer spin goes back experimentally to the results of the Stern–Gerlach experiment. A beam of atoms is run through a strong inhomogeneous magnetic field, which then splits into N parts depending on the intrinsic angular momentum of the atoms. It was found that for silver atoms, the beam was split in two—the ground state therefore could not be integer, because even if the intrinsic angular momentum of the atoms were as small as possible, 1, the beam would be split into three parts, corresponding to atoms with $L_z = -1, 0, +1$. The conclusion is that silver atoms have net intrinsic angular momentum of $\frac{1}{2}$. Pauli set up a theory which explained this splitting by introducing a two-component wave function and a corresponding correction term in the Hamiltonian, representing a semi-classical coupling of this wave function to an applied magnetic field, as so in SI units: (Note that bold faced characters imply Euclidean vectors in 3 dimensions, where as the Minkowski four-vector A_μ can be defined as $A_\mu = (\Phi/c, -\mathbf{A})$.)

$$H = \frac{1}{2m}(\boldsymbol{\sigma}\cdot(\mathbf{p} - e\mathbf{A}))^2 + e\phi.$$

Here $\mathbf{A}$ and ϕ represent the components of the electromagnetic four-potential in their standard SI units, and the three sigmas are the Pauli matrices. On squaring out the first term, a residual interaction with the magnetic field is found, along with the usual classical Hamiltonian of a charged particle interacting with an applied field in SI units:

$$H = \frac{1}{2m}(\mathbf{p} - e\mathbf{A})^2 + e\phi - \frac{e\hbar}{2m}\boldsymbol{\sigma}\cdot\mathbf{B}.$$

This Hamiltonian is now a 2 × 2 matrix, so the Schrödinger equation based on it must use a two-component wave function. On introducing the external electromagnetic 4-vector potential into the Dirac equation in a similar way, known as minimal coupling, it takes the form :

$$(\gamma^\mu(i\hbar\partial_\mu - \frac{e}{c}A_\mu) - mc)\psi = 0\,.$$

A second application of the Dirac operator will now reproduce the Pauli term exactly as before, because the spatial Dirac matrices multiplied by i, have the same squaring and commutation properties as the Pauli matrices. What is more, the value of the gyromagnetic ratio of the electron, standing in front of Pauli's new term, is explained from first principles. This was a major achievement of the Dirac equation and gave physicists great faith in its overall correctness. There is more however. The Pauli theory may be seen as the low energy limit of the Dirac theory in the following manner. First the equation is written in the form of coupled equations for 2-spinors with the SI units restored:

$$\begin{pmatrix} (mc^2 - E + e\phi) & c\boldsymbol{\sigma}\cdot(\mathbf{p} - e\mathbf{A}) \\ -c\boldsymbol{\sigma}\cdot(\mathbf{p} - e\mathbf{A}) & \left(mc^2 + E - e\phi\right) \end{pmatrix} \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

so

$$\begin{aligned} (E - e\phi)\psi_+ - c\boldsymbol{\sigma}\cdot(\mathbf{p} - e\mathbf{A})\,\psi_- &= mc^2\psi_+ \\ -(E - e\phi)\psi_- + c\boldsymbol{\sigma}\cdot(\mathbf{p} - e\mathbf{A})\,\psi_+ &= mc^2\psi_- \end{aligned}$$

Assuming the field is weak and the motion of the electron non-relativistic, we have the total energy of the electron approximately equal to its rest energy, and the momentum going over to the classical value,

$$\begin{aligned} E - e\phi &\approx mc^2 \\ \mathbf{p} &\approx m\mathbf{v} \end{aligned}$$

and so the second equation may be written

$$\psi_- \approx \frac{1}{2mc}\boldsymbol{\sigma}\cdot(\mathbf{p} - e\mathbf{A})\,\psi_+$$

which is of order $\frac{v}{c}$ – thus at typical energies and velocities, the bottom components of the Dirac spinor in the standard representation are much suppressed in comparison to the top components. Substituting this expression into the first equation gives after some rearrangement

$$(E - mc^2)\psi_+ = \frac{1}{2m}[\boldsymbol{\sigma}\cdot(\mathbf{p} - e\mathbf{A})]^2\psi_+ + e\phi\psi_+$$

The operator on the left represents the particle energy reduced by its rest energy, which is just the classical energy, so we recover Pauli's theory if we identify his 2-spinor with the

top components of the Dirac spinor in the non-relativistic approximation. A further approximation gives the Schrödinger equation as the limit of the Pauli theory. Thus, the Schrödinger equation may be seen as the far non-relativistic approximation of the Dirac equation when one may neglect spin and work only at low energies and velocities. This also was a great triumph for the new equation, as it traced the mysterious i that appears in it, and the necessity of a complex wave function, back to the geometry of spacetime through the Dirac algebra. It also highlights why the Schrödinger equation, although superficially in the form of a diffusion equation, actually represents the propagation of waves.

It should be strongly emphasized that this separation of the Dirac spinor into large and small components depends explicitly on a low-energy approximation. The entire Dirac spinor represents an *irreducible* whole, and the components we have just neglected to arrive at the Pauli theory will bring in new phenomena in the relativistic regime – antimatter and the idea of creation and annihilation of particles.

Comparison with the Weyl theory

In the limit $m \to 0$, the Dirac equation reduces to the Weyl equation, which describes relativistic massless spin-$1/2$ particles.[5]

Dirac Lagrangian

Both the Dirac equation and the Adjoint Dirac equation can be obtained from (varying) the action with a specific Lagrangian density that is given by:

$$\mathcal{L} = i\hbar c\overline{\psi}\gamma^\mu \partial_\mu \psi - mc^2\overline{\psi}\psi$$

If one varies this with respect to ψ one gets the Adjoint Dirac equation. Meanwhile, if one varies this with respect to $\overline{\psi}$ one gets the Dirac equation.

Physical interpretation

Identification of observables

The critical physical question in a quantum theory is—what are the physically observable quantities defined by the theory? According to the postulates of quantum mechanics, such quantities are defined by Hermitian operators that act on the Hilbert space of possible states of a system. The eigenvalues of these operators are then the possible results of measuring the corresponding physical quantity. In the Schrödinger theory, the simplest such object is the overall Hamiltonian, which represents the total energy of the system. If we wish to maintain this interpretation on passing to the Dirac theory, we must take the Hamiltonian to be

$$H = \gamma^0 \left[mc^2 + c\gamma^k \left(p_k - \frac{q}{c}A_k\right)\right] + qA^0.$$

where, as always, there is an implied summation over the twice-repeated index $k = 1, 2, 3$. This looks promising, because we see by inspection the rest energy of the particle and, in case $A = 0$, the energy of a charge placed in an electric potential qA^0. What about the term involving the vector potential? In classical electrodynamics, the energy of a charge moving in an applied potential is

$$H = c\sqrt{\left(p - \frac{q}{c}A\right)^2 + m^2c^2} + qA^0.$$

Thus, the Dirac Hamiltonian is fundamentally distinguished from its classical counterpart, and we must take great care to correctly identify what is observable in this theory. Much of the apparently paradoxical behavior implied by the Dirac equation amounts to a misidentification of these observables.

Hole theory

The negative E solutions to the equation are problematic, for it was assumed that the particle has a positive energy. Mathematically speaking, however, there seems to be no reason for us to reject the negative-energy solutions. Since they exist, we cannot simply ignore them, for once we include the interaction between the electron and the electromagnetic field, any electron placed in a positive-energy eigenstate would decay into negative-energy eigenstates of successively lower energy. Real electrons obviously do not behave in this way, or they would disappear by emitting energy in the form of photons.

To cope with this problem, Dirac introduced the hypothesis, known as **hole theory**, that the vacuum is the many-body quantum state in which all the negative-energy electron eigenstates are occupied. This description of the vacuum as a "sea" of electrons is called the Dirac sea. Since the Pauli exclusion principle forbids electrons from occupying the same state, any additional electron would be forced to occupy a positive-energy eigenstate, and positive-energy electrons would be forbidden from decaying into negative-energy eigenstates.

If an electron is forbidden from simultaneously occupying positive-energy and negative-energy eigenstates, then the feature is known as Zitterbewegung, which arises from the interference of positive-energy and negative-energy states, would have to be considered to be an unphysical prediction of time-dependent Dirac theory. This conclusion may be inferred from the explanation of hole theory given in the preceding paragraph. Recent results have been published in Nature [R. Gerritsma, G. Kirchmair, F. Zaehringer, E. Solano, R. Blatt, and C. Roos, Nature 463, 68-71 (2010)] in which the Zitterbewegung feature was simulated in a trapped-ion experiment. This experiment impacts the hole interpretation if one infers that the physics-laboratory experiment is not merely a check on the mathematical correctness of a Dirac-equation solution but the measurement of a real effect whose detectability in electron physics is still beyond reach.

Dirac further reasoned that if the negative-energy eigenstates are incompletely filled, each unoccupied eigenstate – called a **hole** – would behave like a positively charged particle. The hole possesses a *positive* energy since energy is required to create a particle–hole pair from the vacuum. As noted above, Dirac initially thought that the hole might be the proton, but Hermann Weyl pointed out that the hole should behave as if it had the same mass as an electron, whereas the proton is over 1800 times heavier. The hole was eventually identified as the positron, experimentally discovered by Carl Anderson in 1932.

It is not entirely satisfactory to describe the "vacuum" using an infinite sea of negative-energy electrons. The infinitely negative contributions from the sea of negative-energy electrons have to be canceled by an infinite positive "bare" energy and the contribution to the charge density and current coming from the sea of negative-energy electrons is exactly canceled by an infinite positive "jellium" background so that the net electric

charge density of the vacuum is zero. In quantum field theory, a Bogoliubov transformation on the creation and annihilation operators (turning an occupied negative-energy electron state into an unoccupied positive energy positron state and an unoccupied negative-energy electron state into an occupied positive energy positron state) allows us to bypass the Dirac sea formalism even though, formally, it is equivalent to it.

In certain applications of condensed matter physics, however, the underlying concepts of "hole theory" are valid. The sea of conduction electrons in an electrical conductor, called a Fermi sea, contains electrons with energies up to the chemical potential of the system. An unfilled state in the Fermi sea behaves like a positively charged electron, though it is referred to as a "hole" rather than a "positron". The negative charge of the Fermi sea is balanced by the positively charged ionic lattice of the material.

In quantum field theory

In quantum field theories such as quantum electrodynamics, the Dirac field is subject to a process of second quantization, which resolves some of the paradoxical features of the equation.

Other formulations

The Dirac equation can be formulated in a number of other ways.

As a differential equation in one real component

Generically (if a certain linear function of electromagnetic field does not vanish identically), three out of four components of the spinor function in the Dirac equation can be algebraically eliminated, yielding an equivalent fourth-order partial differential equation for just one component. Furthermore, this remaining component can be made real by a gauge transform.[6]

Curved spacetime

This article has developed the Dirac equation in flat spacetime according to special relativity. It is possible to formulate the Dirac equation in curved spacetime.

The algebra of physical space

This article developed the Dirac equation using four vectors and Schrödinger operators. The Dirac equation in the algebra of physical space uses a Clifford algebra over the real numbers, a type of geometric algebra.

See also

The Dirac equation appears on the floor of Westminster Abbey on the plaque commemorating Paul Dirac's life, which was inaugurated on November 13, 1995.[7]

Articles on the Dirac equation	Other topics

- Dirac field
- Dirac spinor
- Klein paradox
- Nonlinear Dirac equation

Other equations

- Breit equation
- Einstein–Maxwell–Dirac equations
- Klein–Gordon equation
- Rarita–Schwinger equation
- Two-body Dirac equations
- Fermionic field
- Feynman checkerboard
- Foldy–Wouthuysen transformation
- Bohr–Sommerfeld theory
- Quantum electrodynamics
- Theoretical and experimental justification for the Schrödinger equation

References

1. P.W. Atkins (1974). *Quanta: A handbook of concepts*. Oxford University Press. p. 52. ISBN 0-19-855493-1.
2. T.Hey, P.Walters (2009). *The New Quantum Universe*. Cambridge University Press. p. 228. ISBN 978-0-521-56457-1.
3. Dirac, P.A.M. (1982) [1958]. *Principles of Quantum Mechanics*. International Series of Monographs on Physics (4th ed.). Oxford University Press. p. 255. ISBN 978-0-19-852011-5.
4. see for example Brian Pendleton: Quantum Theory 2012/2013, section 4.3 The Dirac Equation (http://www2.ph.ed.ac.uk/~bjp/qt/rqt.pdf)
5. Tommy Ohlsson (22 September 2011). *Relativistic Quantum Physics: From Advanced Quantum Mechanics to Introductory Quantum Field Theory*. Cambridge University Press. p. 86. ISBN 978-1-139-50432-4.
6. Akhmeteli, Andrey (2011). "One real function instead of the Dirac spinor function" (PDF). *Journal of Mathematical Physics*. **52** (8): 082303. arXiv:1008.4828. Bibcode:2011JMP....52h2303A. doi:10.1063/1.3624336.
7. Gisela Dirac-Wahrenburg. "Paul Dirac". Dirac.ch. Retrieved 2013-07-12.

Selected papers

- Dirac, P. A. M. (1928). "The Quantum Theory of the Electron" (PDF). *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. **117** (778): 610. Bibcode:1928RSPSA.117..610D. doi:10.1098/rspa.1928.0023. JSTOR 94981.
- Dirac, P. A. M. (1930). "A Theory of Electrons and Protons". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. **126** (801): 360. Bibcode:1930RSPSA.126..360D. doi:10.1098/rspa.1930.0013. JSTOR 95359.
- Anderson, Carl (1933). "The Positive Electron". *Physical Review*. **43** (6): 491. Bibcode:1933PhRv...43..491A. doi:10.1103/PhysRev.43.491.
- Frisch, R.; Stern, O. (1933). "Über die magnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. I". *Zeitschrift für Physik*. **85**: 4. Bibcode:1933ZPhy...85....4F. doi:10.1007/BF01330773.
- M. Arminjon; F. Reifler (2013). "Equivalent forms of Dirac equations in curved spacetimes and generalized de Broglie relations". *Brazilian Journal of Physics*. **43** (1–2): 64–77. arXiv:1103.3201. Bibcode:2013BrJPh..43...64A. doi:10.1007/s13538-012-0111-0.

Textbooks

- Halzen, Francis; Martin, Alan (1984). *Quarks & Leptons: An Introductory Course in Modern Particle Physics*. John Wiley & Sons.
- Shankar, R. (1994). *Principles of Quantum Mechanics* (2nd ed.). Plenum.
- Bjorken, J D; Drell, S. *Relativistic Quantum mechanics*.
- Thaller, B. (1992). *The Dirac Equation*. Texts and Monographs in Physics. Springer.
- Schiff, L.I. (1968). *Quantum Mechanics* (3rd ed.). McGraw-Hill.
- Griffiths, D.J. (2008). *Introduction to Elementary Particles* (2nd ed.). Wiley-VCH. ISBN 978-3-527-40601-2.

External links

- The Dirac Equation (http://www.mathpages.com/home/kmath654/kmath654.htm) at MathPages
- The Nature of the Dirac Equation, its solutions, and Spin (http://www.mc.maricopa.edu/~kevinlg/i256/Nature_Dirac.pdf)
- Dirac equation for a spin ½ particle (http://electron6.phys.utk.edu/qm2/modules/m9/dirac.htm)
- Pedagogic Aids to Quantum Field Theory (http://www.quantumfieldtheory.info/) click on Chap. 4 for a step-by-small-step introduction to the Dirac equation, spinors, and relativistic spin/helicity operators.
- BBC Documentary *Atom 3 The Illusion of Reality* (https://www.youtube.com/watch?v=-yflI8_CJzs)

Categories: Quantum field theory | Spinors | Partial differential equations | Fermions | Paul Dirac | Dirac equation | 1928 introductions

---