Июль 2020 г. **УСПЕХИ ФИЗИЧЕСКИХ НАУК** *Том 190, № 7*

МЕТОДИЧЕСКИЕ ЗАМЕТКИ

Ошибался ли Зоммерфельд?

(К истории появления спина в релятивистских волновых уравнениях)

С.В. Петров

В заметке излагается краткая история возникновения спина электрона в релятивистских волновых уравнениях. Волновое уравнение Дирака было получено в 1928 году, причём он предполагал получить уравнение для "простейшей" частицы, обладающей нулевым спином. Но, как сообщил впоследствии Дирак на Европейской конференции по физике частиц (Будапешт, 4–9 июля 1977 г.), для него оказалось большим сюрпризом, что полученное им уравнение описывает состояния частицы со спином $1/2$.

Ключевые слова: спин, релятивистские волновые уравнения

PACS numbers: **01.65.+g**, **03.65.–w** DOI: https://doi.org/10.3367/UFNr.2019.11.038695

15 лет тому назад в *УФН* было опубликовано письмо в редакцию [1] под названием "Формула Зоммерфельда и теория Дирака", в котором критическому анализу подвергалось совпадение формул энергетического спектра атома водорода, полученных в рамках "старой квантовой теории" (квантование по Бору–Зоммерфельду) и в результате решения уравнения Дирака. Результат анализа был сформулирован уже в аннотации: "Удивительное совпадение формулы тонкой структуры А. Зоммерфельда и П. Дирака оказывается результатом ошибки первого автора".

На наш взгляд, ситуация, во-первых, не так однозначна, как это представлено в [1], а во-вторых, касается столь важной проблемы, как связь между квантовой механикой и механикой классической, а более конкретно: как возникает спин при переходе от классической динамики релятивистской частицы к динамике квантовой [1]. Именно эта причина и побудила нас к написанию заметки, несмотря на значительный отрезок времени, прошедшего с момента публикации [1].

По твёрдому убеждению автора [1], совпадение, упомянутое выше, невозможно, ибо в противном случае следует признать, что "эклектическая теория каким-то образом оказалась эквивалентной последовательной теории Дирака". Отметим прежде всего, что совпадение формул Зоммерфельда и Дирака не означает, конечно, эквивалентности теорий, речь идёт только о численном совпадении значений энергии [2]. Ошибка Зоммерфельда, по мнению автора письма [1], заключалась в том, что он неправильно проквантовал величину орбитального момента L: вместо полуцелых значений $L = (l + 1/2)\hbar$, как это следует из квазиклассики, Зоммерфельд положил $L = n_{\varphi}\hbar$, где n_{φ} — целое число. Поступи он "правильно", он получил бы значения энергии, совпадающие с собственными значениями уравнения Клейна–Фока–Гордона, что, как подчёркнуто в [1], "и должно быть в бесспиновом случае". Вообще-то в 1916 году, когда Зоммерфельд опубликовал свою работу [2], трудно было следовать рецептам квазиклассики, появившимся 10 лет спустя [3–5]. Более того, в 1916 г. никто и не подозревал о существовании спина электрона. Дирак говорил, что "в то время (т.е. вплоть до 1924 года — *С.П.*) о спине и понятия не было", хотя "некоторые физики думали об этом" [6]. Так, в 1918 г. Комптон высказал идею, что электрону присущи как трансляционное, так и вращательное движения [7], а тремя годами позднее пришёл к заключению об электроне как "первичной магнитной частице", вращающейся "подобно миниатюрному гироскопу" [8]. Идею внутреннего магнитного момента электрона для описания спектров щелочных металлов пытался использовать Крониг, но его гипотеза не получила одобрения со стороны Паули, Гейзенберга и Крамерса [9]. Наконец, в 1925–1926 гг. Уленбек и Гаудсмит опубликовали два коротких сообщения, в которых аргументировали введение спина [10, 11]. Второе из этих сообщений сопровождалось письмом Бора, в котором он поддерживал идею вращающегося электрона.

Зоммерфельд следовал по пути, проложенному Бором [12], который учёл, что электрон удерживается на избранной круговой орбите радиуса a благодаря равенству кулоновской и центробежной сил

$$\frac{mv^2}{a} = \frac{Ze^2}{a^2}\,. \tag{1}$$

Это, чисто классическое, условие Бор дополнил условием квантовым, согласно которому величина орбитального

[1] Разумеется, здесь нас интересует исторический аспект вопроса, а не современное состояние теории.

[2] В частности, только в теории Дирака возникает адекватная классификация состояний.

С.В. Петров. Московский государственный университет им. М.В. Ломоносова, химический факультет, Ленинские горы 1, стр. 3, 119992 Москва, Российская Федерация
E-mail: spswix@rambler.ru

Статья поступила 9 сентября 2019 г.

момента $L = mva$ электрона не может быть любой, а должна удовлетворять условию

$$2\pi L = nh\,, \tag{2}$$

где n — целое число начиная с 1. Из этих двух соотношений легко получить энергию электрона на избранной орбите

$$E_n = -\,\frac{2\pi^2 m Z^2 e^4}{h^2}\,\frac{1}{n^2}\,. \tag{3}$$

Этот результат оказался в полном согласии с формулой Бальмера, а спустя 13 лет выяснилось, что значения энергии (3) являются собственными значениями уравнения Шрёдингера. Очевидно, что такое полное согласие было бы невозможно, допусти Бор полуцелые значения n в условии (2). Развивая идеи Бора, Зоммерфельд проквантовал все плоские эллиптические орбиты с помощью условий, наложенных на переменные действия [13]

$$J_\varphi = \oint p_\varphi\,\mathrm{d}\varphi = n_\varphi h\,, \tag{4}$$

$$J_r = \oint p_r\,\mathrm{d}r = n_r h\,, \tag{5}$$

где r и φ — полярные координаты в плоскости орбиты, p_r и p_φ — сопряжённые импульсы, а n_r и n_φ — целые числа. Первое из этих условий сводится к боровскому условию (2).Чтобы реализовать условие (5), необходимо учесть, что функция Гамильтона является интегралом движения, равным энергии электрона E. Тогда легко получить p_r как явную функцию радиальной переменной r и взять интеграл (5). В результате получается та же самая формула [3] энергетических уровней (3), в которой $n = n_r + n_\varphi$. Уже здесь, в нерелятивистском случае, возникает ситуация, о которой упоминает автор [1], а именно, "эклектическая теория" приводит к тем же результатам, что и "последовательная теория" Шрёдингера.

Следующий шаг по пути, начатому Бором, Зоммерфельд сделал, пытаясь объяснить тонкую структуру водородных спектральных линий, экспериментально обнаруженную Майкельсоном [14]. Он сохранил те же условия квантования (4) и (5), но при вычислении интеграла (5) вместо нерелятивистской функции Гамильтона использовал релятивистскую:

$$H = c\sqrt{p^2 + m^2 c^2} - \frac{Ze^2}{r}\,. \tag{6}$$

В полярных координатах на плоскости она записывается в виде

$$H = c\sqrt{p_r^2 + \frac{p_\varphi^2}{r^2} + m^2 c^2} - \frac{Ze^2}{r}\,. \tag{7}$$

Используя условия (4) и (5) и учитывая, что (как и в нерелятивистском случае) энергия E и величина орбитального момента L являются интегралами движения, Зоммерфельд получил

$$E_{n_r n_\varphi} = \frac{mc^2}{\sqrt{1 + \alpha^2 Z^2 \Big/ \left(n_r + \sqrt{n_\varphi^2 - \alpha^2 Z^2}\right)^2}}\,. \tag{8}$$

Согласно этой формуле боровские уровни энергии расщепляются, поскольку теперь энергия (в отличие от нерелятивистского случая) не может быть определена одним квантовым числом $n = n_r + n_\varphi$, что и объясняет тонкую структуру бальмеровских линий. Безразмерная величина $\alpha = 2\pi e/hc \approx 1/137$, автоматически появившаяся при выводе формулы (8) и получившая название "постоянная тонкой структуры", полностью определяет величину расщепления энергетических уровней.

Энергетические уровни дираковского атома водорода [15] определяются главным квантовым числом $n = 1, 2, \ldots$ и квантовым числом полного углового момента (т.е. суммы орбитального и спинового моментов электрона) j, которое (при заданном n) принимает полуцелые значения $1/2, 3/2, \ldots, n - 1/2$

$$E_{nj} = \frac{mc^2}{\sqrt{1 + \left[\alpha^2 Z^2 \Big/ \left(n - (j+1/2) + \sqrt{(j+1/2)^2 - \alpha^2 Z^2}\right)^2\right]}}\,. \tag{9}$$

Численное совпадение обеих формул очевидно. Достаточно в (8) квантовое число n_φ заменить на $(j + 1/2)$, а n_r на $n - (j + 1/2)$.

Вернёмся к ключевому вопросу, вызвавшему как письмо [1], так и настоящую заметку, а именно, к утверждению, что полученные Зоммерфельдом значения энергии должны совпадать с собственными значениями уравнения Клейна – Фока – Гордона, а не уравнения Дирака.

Первое релятивистское волновое уравнение было опубликовано в работах Клейна [16], Фока [17] и Гордона [18]. Легко усмотреть аналогию между происхождением уравнения Клейна – Фока – Гордона [4] и нерелятивистского уравнения Шрёдингера.

Волновое уравнение Шрёдингера для свободной частицы (поскольку мы пытаемся понять, как появляется спин в релятивистских волновых уравнениях, достаточно рассмотреть случай свободной частицы)

$$\mathrm{i}\hbar\,\frac{\partial \Psi}{\partial t} = -\,\frac{\hbar^2}{2m}\,\Delta\Psi \tag{10}$$

формально может быть получено из соотношения между энергией и импульсом свободной частицы

$$E = \frac{1}{2m}p^2\,, \tag{11}$$

если энергии E поставить в соответствие оператор $\mathrm{i}\hbar\,(\partial/\partial t)$, а импульсу $\mathbf{p}$ — оператор $(\hbar/\mathrm{i})\,\nabla$.

Релятивистским аналогом соотношения (11) является выражение

$$\frac{E^2}{c^2} - p^2 = m^2 c^2\,, \tag{12}$$

[3] То, что решение более общей задачи приводит к тому же результату (3), что и частный случай круговых орбит, не удивляло Зоммерфельда, поскольку (по его выражениию [13]) "в этом семействе эллипсов каждый эллипс энергетически эквивалентен вполне определённому боровскому кругу". Сегодня об этом феномене мы говорим как о вырождении состояний по энергии.

[4] По печальной традиции имя Фока обычно не упоминается в названии уравнения (см. по этому поводу публикации в *УФН* [19 – 21]). Следует отметить также, что к созданию уравнения Клейна – Фока – Гордона были причастны де Бройль и Шрёдингер [6].

где $\mathbf{p}$ теперь — релятивистский импульс частицы. В левой части соотношения (12) — скалярное произведение[5] 4-импульса $\mathbf{P} = (E/c, \mathbf{p})$ самого на себя, являющееся, как и скалярное произведение двух любых релятивистских 4-векторов, инвариантом. Релятивистское волновое уравнение может быть получено, если 4-импульсу $\mathbf{P}$ поставить в соответствие оператор $\hat{\mathbf{P}}$, четыре компоненты которого суть умноженные на $\hbar/\mathrm{i}$ операторы дифференцирования по компонентам 4-радиуса-вектора $\mathbf{R} = (ct, \mathbf{r})$[6]:

$$\left(-\frac{\hbar^2}{c^2}\frac{\partial^2}{\partial t^2} + \hbar^2\Delta - m^2c^2\right)\Psi = 0\,. \tag{13}$$

Первые два слагаемых в скобках образуют умноженное на $\hbar^2$ скалярное произведение 4-градиента самого на себя, так что оператор, действующий на волновую функцию, релятивистски инвариантен. Следовательно, чтобы волновое уравнение (13) (как и любое релятивистское уравнение) согласно принципам специальной теории относительности не меняло своей формы при преобразовании Лоренца, т.е. было ковариантным, волновая функция Ψ должна быть скалярной функцией только координат и времени, поскольку в уравнении (13) нет членов, ответственных за спин. Следуя канонам квантовой механики (замена динамических переменных соответствующими операторами) и специальной теории относительности (требование релятивистской инвариантности), мы автоматически приходим к волновому уравнению (13), описывающему движение бесспиновой частицы. Изначально каких-либо дополнительных требований при получении волнового уравнения для частицы именно с нулевым спином не было.

Релятивистское волновое уравнение для электрона было опубликовано Дираком в 1928 году [22]. Предшествующие два года он посвятил формулировке квантовой механики на символическом языке, положив в основу абстрактные векторы состояний[7]. По мнению Дирака, "символический метод глубже проникает в природу вещей" и "позволяет выразить физические законы в ясной и сжатой форме", а метод представлений (волновая механика Шрёдингера и матричная механика Гейзенберга – Борна – Иордана) удобнее использовать для решения конкретных задач [23].

Отправным пунктом при выводе релятивистского волнового уравнения было абсолютное неприятие Дираком уравнения Клейна – Фока – Гордона. Исходя из справедливости принципа суперпозиции в течение всего промежутка времени (по крайней мере пока система не подвержена какому-либо возмущению), уравнение, описывающее изменение вектора состояния, должно быть линейным дифференциальным уравнением 1-го порядка по времени [23], в то время как уравнение Клейна – Фока – Гордона содержит вторую временну́ю производную. Спустя полвека Дирак вспоминал [6], что в 1927 г. "Бор, казалось, был полностью удовлетворён теорией Клейна – Гордона" и его точка зрения "совпадала с мнением большинства физиков в то время, — возможно, даже с мнением всех их". И это при том, что в рамках теории Клейна – Фока – Гордона возникали неразрешимые проблемы с вычислением вероятностей динамических переменных[8]: "Если бы вам захотелось найти вероятность того, что импульс принимает заданные значения, то вы вообще не смогли бы ответить на этот вопрос. Точно так же для других динамических переменных совсем не удалось бы получить никакой информации о их вероятностях" [6].

Требованию линейности по временно́й производной отвечает волновое уравнение шрёдингеровского типа

$$\mathrm{i}\hbar\,\frac{\partial\Psi}{\partial t} = H\Psi \tag{14}$$

с релятивистским гамильтонианом в правой части

$$H = c\sqrt{\hat{p}^2 + m^2c^2}\,. \tag{15}$$

Однако такая форма записи волнового уравнения не устраивала Дирака, поскольку в силу равноправия всех четырёх координат любой точки пространства-времени правильная релятивистская теория должна быть полностью симметричной по отношению к производным по времени и декартовым координатам и, следовательно, волновое уравнение должно быть линейным и по пространственным компонентам 4-градиента. Таким образом, перед Дираком возникла проблема: как извлечь квадратный корень в выражении (15)? Как вспоминал впоследствии Дирак [6], к решению он пришёл "в какой-то мере случайно". В 1927 году существование спина электрона сомнений не вызывало. И в частности, были известны двухрядные матрицы Паули σ_1, σ_2 и σ_3. И вот однажды, как пишет Дирак [6], "играясь с математикой", он обнаружил "очень интересный результат, а именно":

$$(\sigma_1\hat{p}_x + \sigma_2\hat{p}_y + \sigma_3\hat{p}_z)^2 = \hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2\,, \tag{16}$$

"и таким образом получилось нечто вроде квадратного корня для $\hat{p}_x^2 + \hat{p}_y^2 + \hat{p}_z^2$". Однако в выражении (15) под корнем — сумма четырёх квадратов, поэтому трёх матриц Паули недостаточно и следовало бы добавить четвёртую матрицу, обладающую теми же свойствами, что и матрицы Паули, а именно: её квадрат должен быть единичной матрицей и она должна антикоммутировать с каждой из матриц Паули. Но такой матрицы не оказалось и "в течение нескольких недель это было серьёзным препятствием" для Дирака, пока он не обнаружил, что, для того чтобы получить гамильтониан, линейный по декартовым компонентам оператора импульса

$$H_{\mathrm{D}} = c(\alpha_1\hat{p}_x + \alpha_2\hat{p}_y + \alpha_3\hat{p}_z) + mc^2\beta\,, \tag{17}$$

в качестве величин $\alpha_i (i = 1, 2, 3)$ и β, удовлетворяющих тем же требованиям, что и матрицы Паули

$$\begin{cases} \alpha_i\alpha_k + \alpha_k\alpha_i = 0\,, \\ \alpha_i\beta + \beta\alpha_i = 0\,, \\ \alpha_i^2 = \beta^2 = 1\,, \end{cases} \tag{18}$$

можно выбрать эрмитовы матрицы 4×4, которые не должны зависеть от компонент $\hat{\mathbf{p}}$ (в противном случае

[5] Мы имеем в виду скалярное произведение в псевдоэвклидовом пространстве Минковского.

[6] Оператор $\hat{\mathbf{P}}$ с точностью до множителя $\hbar/\mathrm{i}$ является 4-градиентом.

[7] В законченном виде символический метод был изложен в 1930 г. в знаменитых "Принципах" [23].

[8] Поскольку уравнение Клейна – Фока – Гордона является дифференциальным уравнением 2-го порядка по времени, плотность вероятности приходится определять как билинейную форму по Ψ и $\partial\Psi/\partial t$ [24], в результате чего она может принимать как положительные, так и отрицательные значения и, следовательно, теряет свой смысл как плотность вероятности.

гамильтониан H_{D} не будет линейным) и не должны быть функциями координат и времени, поскольку для свободной частицы все точки четырёхмерного пространства-времени должны быть равноправными. В качестве таких матриц Дирак взял матрицы:

$$\alpha_i = \begin{pmatrix} O & \sigma_i \\ \sigma_i & O \end{pmatrix}, \tag{19}$$

$$\beta = \begin{pmatrix} I & O \\ O & -I \end{pmatrix}, \tag{20}$$

где O и I — двумерные нулевая и единичная матрицы. Независимость матриц α_i и β от "традиционных" динамических переменных неизбежно приводит к выводу, что эти матрицы "описывают некоторые новые степени свободы, относящиеся к какому-то внутреннему движению электрона" [23]. Другими словами, если бы к 1928 г. не существовало экспериментальных данных, свидетельствующих о внутреннем угловом моменте электрона, релятивистское волновое уравнение Дирака явилось бы теоретическим предсказанием его существования.

Из матричной структуры гамильтониана Дирака (17) следует, что волновая функция должна рассматриваться как 4-мерный столбец [9], каждый элемент которого является функцией координат и времени, а волновое уравнение представляет собой систему из четырёх связанных дифференциальных уравнений первого порядка. Наконец, непосредственная проверка показала [22] (см. также [23]), что полученное волновое уравнение ковариантно относительно преобразований Лоренца. Итак, Дирак получил "истинное релятивистское уравнение" [6], безупречное как в рамках квантовой механики, так и специальной теории относительности, при этом оказалось, что оно законно только для частиц со спином 1/2. Однако ход рассуждений Дирака не даёт ответа на вопрос: почему полученное уравнение описывает движение частицы с половинным спином, а не каким-либо другим? Ведь, казалось бы, что использование матриц Паули как раз говорит о стремлении Дирака получить волновое уравнение для частиц со спином 1/2. Сам Дирак в связи с этим говорил [6], что его целью было получить уравнение для "простейшей частицы и оказалось, что таковой является частица со спином половина. Для меня это было большим сюрпризом, потому что я считал, что простейшая частица, естественно, должна иметь нулевой спин, а спин половина придётся вводить позднее как некоторое усложнение, после того как удастся решить задачу о бесспиновой частице. Но всё оказалось совсем по-другому".

Таким образом, как и отсутствие спина в уравнении Клейна – Фока – Гордона, так и появление спина 1/2 в уравнении Дирака происходит автоматически, без каких-либо предварительных требований, обусловливающих наличие (или отсутствие) спина в волновых уравнениях.

Совпадение же результатов Зоммерфельда и Дирака (или Клейна – Фока – Гордона при соответствующей замене целого числа n_φ на полуцелое $(l+1/2)$ в условии (4)), как и совпадение результатов Бора и Шрёдингера в нерелятивистском случае, является, по-видимому, случайным. "Кажется, случайности такого рода бывают при поисках понимания Природы" [6].

[9] Биспинор в современной терминологии.

Список литературы

1. Грановский Я И *УФН* **174** 577 (2004); Granovskii Ya I *Phys. Usp.* **47** 523 (2004)
2. Sommerfeld A *Ann. Physik* **51** 125 (1916) По новой нумерации томов, принятой на сайте Wiley Online Library, Vol. 356
3. Wentzel G *Z. Phys.* **38** 518 (1926)
4. Kramers H A *Z. Phys.* **39** 828 (1926)
5. Brillouin L *C.R. Acad. Sci.* **183** 24 (1926)
6. Дирак П А М *УФН* **128** 681 (1979); Dirac P A M *Sov. Phys. Usp.* **22** 648 (1979)
7. Compton A H *J. Washington Acad. Sci.* **8** 1 (1918)
8. Compton A H *J. Franklin Inst.* **192** 145 (1921)
9. Jammer M *The Conceptual Development of Quantum Mechanics* (New York: Mc Graw-Hill Book Co., 1967); Пер. на русск. яз.: Джеммер М *Эволюция понятий квантовой механики* (М.: Наука, 1985)
10. Uhlenbeck G E, Goudsmit S A *Naturwissenschaften* **13** 953 (1925)
11. Uhlenbeck G E, Goudsmit S *Nature* **117** 264 (1926)
12. Bohr N *Philos. Mag. 6* **26** 1 (1913); *Philos. Mag. 6* **26** 476 (1913); *Philos. Mag. 6* **26** 857 (1913)
13. Sommerfeld A *Atombau und Spektrallinien* Vol. 1 (Braunschweig: F. Vieweg and Sohn, 1951); Пер. на русск. яз.: Зоммерфельд А *Строение атома и спектры* Т. 1 (М.: ГИТТЛ, 1956)
14. Michelson A A *Philos. Mag. 5* **31** 338 (1891)
15. Messiah A *Quantum Mechanics* Vol. 2 (Amsterdam: North-Holland Publ. Co., 1961); Пер. на русск. яз.: Мессиа А *Квантовая механика* Т. 2 (М.: Наука, 1979)
16. Klein O *Z. Phys.* **37** 855 (1926)
17. Fock V *Z. Phys.* **38** 242 (1926)
18. Gordon W *Z. Phys.* **40** 117 (1926)
19. Окунь Л Б *УФН* **180** 871 (2010); Okun L B *Phys. Usp.* **535** 835 (2010)
20. Дайсон Ф *УФН* **180** 859 (2010); Dyson F *Phys. Usp.* **53** 825 (2010)
21. Фок В А *УФН* **180** 874 (2010); Fock V A *Phys. Usp.* **53** 839 (2010); Fock V *Z. Phys.* **39** 226 (1926)
22. Dirac P A M *Proc. R. Soc. Lond. A* **117** 610 (1928)
23. Dirac P A M *The Principles of Quantum Mechanics* 4th ed. (Oxford: Clarendon Press, 1958); Пер. на русск. яз.: Дирак П А М *Принципы квантовой механики* (М.: Физматлит, 1960)
24. Давыдов А С *Квантовая механика* (М.: Физматлит, 1963); Пер. на англ. яз.: Davydov A S *Quantum Mechanics* (Oxford: Pergamon Press, 1976)

Was Sommerfeld wrong? (To the history of the appearance of spin in relativistic wave equations)

S.V. Petrov
Lomonosov Moscow State University, Department of Chemistry, Leninskie gory 1, str. 3, 119992 Moscow, Russian Federation
E-mail: *spswix@rambler.ru*

This article presents a brief history of the appearance of electron spin in relativistic wave equations. Dirac derived his wave equation in 1928 with the intention to obtain an equation for the "simplest" particle with spin zero. But, as Dirac later announced at the European Conference on Particle Physics (Budapest, 4-9 July 1977), it was a big surprise for him that the equation described the states of a particle with spin 1/2.

Keywords: spin, relativistic wave equations

PACS numbers: **01.65.+g**, **03.65.−w**

Bibliography — 24 references
Uspekhi Fizicheskikh Nauk **190** (7) 777 – 780 (2020)
DOI: https://doi.org/10.3367/UFNr.2019.11.038695

Received 9 September 2019
Physics – Uspekhi **63** (7) (2020)
DOI: https://doi.org/10.3367/UFNe.2019.11.038695