

WOLFGANG PAULI

Exclusion principle and quantum mechanics

Nobel Lecture, December 13, 1946

The history of the discovery of the « exclusion principle », for which I have received the honor of the Nobel Prize award in the year 1945, goes back to my students days in Munich. While, in school in Vienna, I had already obtained some knowledge of classical physics and the then new Einstein relativity theory, it was at the University of Munich that I was introduced by Sommerfeld to the structure of the atom - somewhat strange from the point of view of classical physics. I was not spared the shock which every physicist, accustomed to the classical way of thinking, experienced when he came to know of Bohr's « basic postulate of quantum theory » for the first time. At that time there were two approaches to the difficult problems connected with the quantum of action. One was an effort to bring abstract order to the new ideas by looking for a key to translate classical mechanics and electrodynamics into quantum language which would form a logical generalization of these. This was the direction which was taken by Bohr's « correspondence principle ». Sommerfeld, however, preferred, in view of the difficulties which blocked the use of the concepts of kinematical models, a direct interpretation, as independent of models as possible, of the laws of spectra in terms of integral numbers, following, as Kepler once did in his investigation of the planetary system, an inner feeling for harmony. Both methods, which did not appear to me irreconcilable, influenced me. The series of whole numbers 2, 8, 18, 32... giving the lengths of the periods in the natural system of chemical elements, was zealously discussed in Munich, including the remark of the Swedish physicist, Rydberg, that these numbers are of the simple form $2n^2$, if n takes on all integer values. Sommerfeld tried especially to connect the number 8 and the number of corners of a cube.

A new phase of my scientific life began when I met Niels Bohr personally for the first time. This was in 1922, when he gave a series of guest lectures at Göttingen, in which he reported on his theoretical investigations on the Periodic System of Elements. I shall recall only briefly that the essential progress made by Bohr's considerations at that time was in explaining, by means of the spherically symmetric atomic model, the formation of the intermediate

shells of the atom and the general properties of the rare earths. The question, as to why all electrons for an atom in its ground state were not bound in the innermost shell, had already been emphasized by Bohr as a fundamental problem in his earlier works. In his Göttingen lectures he treated particularly the closing of this innermost K-shell in the helium atom and its essential connection with the two non-combining spectra of helium, the ortho- and para-helium spectra. However, no convincing explanation for this phenomenon could be given on the basis of classical mechanics. It made a strong impression on me that Bohr at that time and in later discussions was looking for a *general* explanation which should hold for the closing of *every* electron shell and in which the number 2 was considered to be as essential as 8 in contrast to Sommerfeld's approach.

Following Bohr's invitation, I went to Copenhagen in the autumn of 1922, where I made a serious effort to explain the so-called « anomalous Zeeman effect », as the spectroscopists called a type of splitting of the spectral lines in a magnetic field which is different from the normal triplet. On the one hand, the anomalous type of splitting exhibited beautiful and simple laws and Landé had already succeeded to find the simpler splitting of the spectroscopic terms from the observed splitting of the lines. The most fundamental of his results thereby was the use of half-integers as magnetic quantum numbers for the doublet-spectra of the alkali metals. On the other hand, the anomalous splitting was hardly understandable from the standpoint of the mechanical model of the atom, since very general assumptions concerning the electron, using classical theory as well as quantum theory, always led to the same triplet. A closer investigation of this problem left me with the feeling that it was even more unapproachable. We know now that at that time one was confronted with two logically different difficulties simultaneously. One was the absence of a general key to translate a given mechanical model into quantum theory which one tried in vain by using classical mechanics to describe the stationary quantum states themselves. The second difficulty was our ignorance concerning the proper classical model itself which could be suited to derive at all an anomalous splitting of spectral lines emitted by an atom in an external magnetic field. It is therefore not surprising that I could not find a satisfactory solution of the problem at that time. I succeeded, however, in generalizing Landé's term analysis for very strong magnetic fields², a case which, as a result of the magneto-optic transformation (Paschen-Back effect), is in many respects simpler. This early work

was of decisive importance for the finding of the exclusion principle.

Very soon after my return to the University of Hamburg, in 1923, I gave there my inaugural lecture as *Privatdozent* on the Periodic System of Elements. The contents of this lecture appeared very unsatisfactory to me, since the problem of the closing of the electronic shells had been clarified no further. The only thing that was clear was that a closer relation of this problem to the theory of multiplet structure must exist. I therefore tried to examine again critically the simplest case, the doublet structure of the alkali spectra. According to the point of view then orthodox, which was also taken over by Bohr in his already mentioned lectures in Göttingen, a non-vanishing angular momentum of the atomic core was supposed to be the cause of this doublet structure.

In the autumn of 1924 I published some arguments against this point of view, which I definitely rejected as incorrect and proposed instead of it the assumption of a new quantum theoretic property of the electron, which I called a « two-valuedness not describable classically »³. At this time a paper of the English physicist, Stoner, appeared⁴ which contained, besides improvements in the classification of electrons in subgroups, the following essential remark: For a given value of the principal quantum number is the number of energy levels of a single electron in the alkali metal spectra in an external magnetic field the same as the number of electrons in the closed shell of the rare gases which corresponds to this principal quantum number.

On the basis of my earlier results on the classification of spectral terms in a strong magnetic field the general formulation of the exclusion principle became clear to me. The fundamental idea can be stated in the following way: The complicated numbers of electrons in closed subgroups are reduced to the simple number *one* if the division of the groups by giving the values of the four quantum numbers of an electron is carried so far that every degeneracy is removed. An entirely non-degenerate energy level is already « closed », if it is occupied by a single electron; states in contradiction with this postulate have to be excluded. The exposition of this general formulation of the exclusion principle was made in Hamburg in the spring of 1925⁵, after I was able to verify some additional conclusions concerning the anomalous Zeeman effect of more complicated atoms during a visit to Tübingen with the help of the spectroscopic material assembled there.

With the exception of experts on the classification of spectral terms, the physicists found it difficult to understand the exclusion principle, since no meaning in terms of a model was given to the fourth degree of freedom of

general assumptions. I had always the feeling and I still have it today, that this is a deficiency. Of course in the beginning I hoped that the new quantum mechanics, with the help of which it was possible to deduce so many half-empirical formal rules in use at that time, will also rigorously deduce the exclusion principle. Instead of it there was for electrons still an exclusion: not of particular states any longer, but of whole classes of states, namely the exclusion of all classes different from the antisymmetrical one. The impression that the shadow of some incompleteness fell here on the bright light of success of the new quantum mechanics seems to me unavoidable. We shall resume this problem when we discuss relativistic quantum mechanics but wish to give first an account of further results of the application of wave mechanics to systems of several like particles.

In the paper of Heisenberg, which we are discussing, he was also able to give a simple explanation of the existence of the two non-combining spectra of helium which I mentioned in the beginning of this lecture. Indeed, besides the rigorous separation of the wave functions into symmetry classes with respect to space-coordinates and spin indices together, there exists an approximate separation into symmetry classes with respect to space coordinates alone. The latter holds only so long as an interaction between the spin and the orbital motion of the electron can be neglected. In this way the para- and ortho-helium spectra could be interpreted as belonging to the class of symmetrical and antisymmetrical wave functions respectively in the space coordinates alone. It became clear that the energy difference between corresponding levels of the two classes has nothing to do with magnetic interactions but is of a new type of much larger order of magnitude, which one called exchange energy.

Of more fundamental significance is the connection of the symmetry classes with general problems of the statistical theory of heat. As is well known, this theory leads to the result that the entropy of a system is (apart from a constant factor) given by the logarithm of the number of quantum states of the whole system on a so-called energy shell. One might first expect that this number should be equal to the corresponding volume of the multi-dimensional phase space divided by h^f , where h is Planck's constant and f the number of degrees of freedom of the whole system. However, it turned out that for a system of N like particles, one had still to divide this quotient by $N!$ in order to get a value for the entropy in accordance with the usual postulate of homogeneity that the entropy has to be proportional to the mass for a given inner state of the substance. In this way a qualitative distinction between

was discovered, I proposed to use the assumption of a nuclear spin to interpret the hyperfine-structure of spectral lines¹⁵. This proposal met on the one hand strong opposition from many sides but influenced on the other hand Goudsmit and Uhlenbeck in their claim of an electron spin. It was only some years later that my attempt to interpret the hyperfine-structure could be definitely confirmed experimentally by investigations in which also Zeeman himself participated and which showed the existence of a magneto-optic transformation of the hyperfine-structure as I had predicted it. Since that time the hyperfine-structure of spectral lines became a general method of determining the nuclear spin.

In order to determine experimentally also the symmetry class of the nuclei, other methods were necessary. The most convenient, although not the only one, consists in the investigation of band spectra due to a molecule with two like atoms¹⁶. It could easily be derived that in the ground state of the electron configuration of such a molecule the states with even and odd values of the rotational quantum number are symmetric and antisymmetric respectively for a permutation of the space coordinates of the two nuclei. Further there exist among the $(2I+1)^2$ spin states of the pair of nuclei, $(2I+1)(I+1)$ states symmetrical and $(2I+1)I$ states antisymmetrical in the spins, since the $(2I+1)$ states with two spins in the same direction are necessarily symmetrical. Therefore the conclusion was reached: If the total wave function of space coordinates and spin indices of the nuclei is symmetrical, the ratio of the weight of states with an even rotational quantum number to the weight of states with an odd rotational quantum number is given by $(I+1) : I$. In the reverse case of an antisymmetrical total wave function of the nuclei, the same ratio is $I : (I+1)$. Transitions between one state with an even and another state with an odd rotational quantum number will be extremely rare as they can only be caused by an interaction between the orbital motions and the spins of the nuclei. Therefore the ratio of the weights of the rotational states with different parity will give rise to two different systems of band spectra with different intensities, the lines of which are alternating.

The first application of this method was the result that the protons have the spin $\frac{1}{2}$ and fulfill the exclusion principle just as the electrons. The initial difficulties to understand quantitatively the specific heat of hydrogen molecules at low temperatures were removed by Dennison's hypothesis¹⁷, that at this low temperature the thermal equilibrium between the two modifications of the hydrogen molecule (ortho- H_2 : odd rotational quantum numbers,

of the vacuum derived from the quantized field becomes infinite, a result which is directly connected with the fact that the system considered has an infinite number of degrees of freedom. It is clear that this zero-point energy has no physical reality, for instance it is not the source of a gravitational field. Formally it is easy to subtract constant infinite terms which are independent of the state considered and never change; nevertheless it seems to me that already this result is an indication that a fundamental change in the concepts underlying the present theory of quantized fields will be necessary.

In order to clarify certain aspects of relativistic quantum theory I have discussed here, different from the historical order of events, the one-valued fields first. Already earlier Dirac²³ had formulated his relativistic wave equations corresponding to material particles with spin $\frac{1}{2}$ using a pair of so-called spinors with two components each. He applied these equations to the problem of one electron in an electromagnetic field. In spite of the great success of this theory in the quantitative explanation of the fine structure of the energy levels of the hydrogen atom and in the computation of the scattering cross section of one photon by a free electron, there was one consequence of this theory which was obviously in contradiction with experience. The energy of the electron can have, according to the theory, both positive and negative values, and, in external electromagnetic fields, transitions should occur from states with one sign of energy to states with the other sign. On the other hand there exists in this theory a four-vector satisfying the continuity equation with a fourth component corresponding to a density which is definitely positive.

It can be shown that there is a similar situation for all fields, which, like the spinors, transform for rotations in ordinary space according to two-valued representations, thus changing their sign for a full rotation. We shall call briefly such quantities « two-valued ». From the relativistic wave equations of such quantities one can always derive a four-vector bilinear in the field components which satisfies the continuity equation and for which the fourth component, at least after integration over the space, gives an essentially positive quantity. On the other hand, the expression for the total energy can have both the positive and the negative sign.

Is there any means to shift the minus sign from the energy back to the density of the four-vector? Then the latter could again be interpreted as charge density in contrast to particle density and the energy would become positive as it ought to be. You know that Dirac's answer was that this could actually be achieved by application of the exclusion principle. In his lecture

delivered here in Stockholm¹⁰ he himself explained his proposal of a new interpretation of his theory, according to which in the actual vacuum all the states of negative energy should be occupied and only deviations of this state of smallest energy, namely holes in the sea of these occupied states are assumed to be observable. It is the exclusion principle which guarantees the stability of the vacuum, in which all states of negative energy are occupied. Furthermore the holes have all properties of particles with positive energy and positive electric charge, which in external electromagnetic fields can be produced and annihilated in pairs. These predicted positrons, the exact mirror images of the electrons, have been actually discovered experimentally.

The new interpretation of the theory obviously abandons in principle the standpoint of the one-body problem and considers a many-body problem from the beginning. It cannot any longer be claimed that Dirac's relativistic wave equations are the only possible ones but if one wants to have relativistic field equations corresponding to particles, for which the value $\frac{1}{2}$ of their spin is known, one has certainly to assume the Dirac equations. Although it is logically possible to quantize these equations like classical fields, which would give symmetrical states of a system consisting of many such particles, this would be in contradiction with the postulate that the energy of the system has actually to be positive. This postulate is fulfilled on the other hand if we apply the exclusion principle and Dirac's interpretation of the vacuum and the holes, which at the same time substitutes the physical concept of charge density with values of both signs for the mathematical fiction of a positive particle density. A similar conclusion holds for all relativistic wave equations with two-valued quantities as field components. This is the other step (historically the earlier one) in the direction of an understanding of the connection between spin and symmetry class.

I can only shortly note that Dirac's new interpretation of empty and occupied states of negative energy can be formulated very elegantly with the help of the formalism of Jordan and Wigner mentioned before. The transition from the old to the new interpretation of the theory can indeed be carried through simply by interchanging the meaning of one of the operators with that of its hermitian conjugate if they are applied to states originally of negative energy. The infinite « zero charge » of the occupied states of negative energy is then formally analogous to the infinite zero-point energy of the quantized one-valued fields. The former has no physical reality either and is not the source of an electromagnetic field.

In spite of the formal analogy between the quantization of the one-valued fields leading to ensembles of like particles with symmetrical states and to particles fulfilling the exclusion principle described by two-valued operator quantities, depending on space and time coordinates, there is of course the fundamental difference that for the latter there is no limiting case, where the mathematical operators can be treated like classical fields. On the other hand we can expect that the possibilities and the limitations for the applications of the concepts of space and time, which find their expression in the different concepts of charge density and particle density, will be the same for charged particles with integer and with half-integer spins.

The difficulties of the present theory become much worse, if the interaction of the electromagnetic field with matter is taken into consideration, since the well-known infinities regarding the energy of an electron in its own field, the so-called self-energy, then occur as a result of the application of the usual perturbation formalism to this problem. The root of this difficulty seems to be the circumstance that the formalism of field quantization has only a direct meaning so long as the sources of the field can be treated as continuously distributed, obeying the laws of classical physics, and so long as only averages of field quantities over finite space-time regions are used. The electrons themselves, however, are essentially non-classical field sources.

At the end of this lecture I may express my critical opinion, that a correct theory should neither lead to infinite zero-point energies nor to infinite zero charges, that it should not use mathematical tricks to subtract infinities or singularities, nor should it invent a « hypothetical world » which is only a mathematical fiction before it is able to formulate the correct interpretation of the actual world of physics.

From the point of view of logic, my report on « Exclusion principle and quantum mechanics » has no conclusion. I believe that it will only be possible to write the conclusion if a theory will be established which will determine the value of the fine-structure constant and will thus explain the atomistic structure of electricity, which is such an essential quality of all atomic sources of electric fields actually occurring in Nature.

1. A. Landé, *Z. Physik*, 5 (1921) 231 and *Z. Physik*, 7(1921) 398, *Physik. Z.*, 22 (1921) 417.
2. W. Pauli, *Z. Physik*, 16 (1923) 155.
3. W. Pauli, *Z. Physik*, 31 (1925) 373.
4. E. C. Stoner, *Phil. Mag.*, 48 (1924) 719.
5. W. Pauli, *Z. Physik*, 31 (1925) 765.
6. S. Goudsmit and G. Uhlenbeck, *Naturwiss.*, 13 (1925) 953, *Nature*, 117 (1926) 264.
7. L. H. Thomas, *Nature*, 117 (1926) 514, and *Phil. Mag.*, 3 (1927) 1. Compare also J. Frenkel, *Z. Physik*, 37 (1926) 243.
8. Compare *Rapport du Sixième Conseil Solvay de Physique, Paris, 1932*, pp. 217-225.
9. For this earlier stage of the history of the exclusion principle compare also the author's note in *Science*, 103 (1946) 213, which partly coincides with the first part of the present lecture.
10. The Nobel Lectures of W. Heisenberg, E. Schrodinger, and P. A. M. Dirac are collected in *Die moderne Atomtheorie*, Leipzig, 1934.
11. The articles of N. Bohr are collected in *Atomic Theory and the Description of Nature*, Cambridge University Press, 1934. See also his article « Light and Life », *Nature*, 131 (1933) 421, 457.
12. W. Heisenberg, *Z. Physik*, 38 (1926) 411 and 39 (1926) 499.
13. E. Fermi, *Z. Physik*, 36 (1926) 902.
P. A. M. Dirac, *Proc. Roy. Soc. London*, A 112 (1926) 661.
14. S. N. Bose, *Z. Physik*, 26 (1924) 178 and 27 (1924) 384.
A. Einstein, *Berl. Ber.*, (1924) 261; (1925) 1, 1S.
15. W. Pauli, *Naturwiss.*, 12 (1924) 741.
16. W. Heisenberg, *Z. Physik*, 41 (1927) 239, F. Hund, *Z. Physik*, 42 (1927) 39.
17. D. M. Dennison, *Proc. Roy. Soc. London*, A 115 (1927) 483.
18. R. de L. Kronig, *Naturwiss.*, 16 (1928) 335.
W. Heitler und G. Herzberg, *Naturwiss.*, 17 (1929) 673.
19. G. N. Lewis and M. F. Ashley, *Phys. Rev.*, 43 (1933) 837.
G. M. Murphy and H. Johnston, *Phys. Rev.*, 45 (1934) 550 and 46 (1934) 95.
20. Compare for the following the author's report in *Rev. Mod. Phys.*, 13 (1941) 203, in which older literature is given. See also W. Pauli and V. Weisskopf, *Helv. Phys. Acta*, 7 (1934) 809.
21. N. Bohr and L. Rosenfeld, *Kgl. Danske Videnskab. Selskab. Mat. Fys. Medd.*, 12 [8] (1933).
22. P. Jordan and E. Wigner, *Z. Physik*, 47 (1928) 631.
Compare also V. Fock, *Z. Physik*, 75 (1932) 622.
23. W. Pauli, *Ann. Inst. Poincaré*, 6 (1936) 137 and *Phys. Rev.*, 58 (1940) 716.
24. L. Landau and R. Peierls, *Z. Physik*, 69 (1931) 56.
Compare also the author's article in *Handbuch der Physik*, 24, Part 1, 1933, Chap. A, §2.
25. P. A. M. Dirac, *Proc. Roy. Soc. London*, A 117 (1928) 610.