Quantised Singularities in the Electromagnetic Field.

By P. A. M. Dirac, F.R.S., St. John's College, Cambridge.

(Received May 29, 1931.)

§ 1. *Introduction.*

The steady progress of physics requires for its theoretical formulation a mathematics that gets continually more advanced. This is only natural and to be expected. What, however, was not expected by the scientific workers of the last century was the particular form that the line of advancement of the mathematics would take, namely, it was expected that the mathematics would get more and more complicated, but would rest on a permanent basis of axioms and definitions, while actually the modern physical developments have required a mathematics that continually shifts its foundations and gets more abstract. Non-euclidean geometry and non-commutative algebra, which were at one time considered to be purely fictions of the mind and pastimes for logical thinkers, have now been found to be very necessary for the description of general facts of the physical world. It seems likely that this process of increasing abstraction will continue in the future and that advance in physics is to be associated with a continual modification and generalisation of the axioms at the base of the mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation.

There are at present fundamental problems in theoretical physics awaiting solution, *e.g.*, the relativistic formulation of quantum mechanics and the nature of atomic nuclei (to be followed by more difficult ones such as the problem of life), the solution of which problems will presumably require a more drastic revision of our fundamental concepts than any that have gone before. Quite likely these changes will be so great that it will be beyond the power of human intelligence to get the necessary new ideas by direct attempts to formulate the experimental data in mathematical terms. The theoretical worker in the future will therefore have to proceed in a more indirect way. The most powerful method of advance that can be suggested at present is to employ all the resources of pure mathematics in attempts to perfect and generalise the mathematical formalism that forms the existing basis of theoretical physics, and *after* each success in this direction, to try to interpret the new mathematical features in terms of physical entities (by a process like Eddington's Principle of Identification).

A recent paper by the author* may possibly be regarded as a small step according to this general scheme of advance. The mathematical formalism at that time involved a serious difficulty through its prediction of negative kinetic energy values for an electron. It was proposed to get over this difficulty, making use of Pauli's Exclusion Principle which does not allow more than one electron in any state, by saying that in the physical world almost all the negative-energy states are already occupied, so that our ordinary electrons of positive energy cannot fall into them. The question then arises as to the physical interpretation of the negative-energy states, which on this view really exist. We should expect the uniformly filled distribution of negative-energy states to be completely unobservable to us, but an unoccupied one of these states, being something exceptional, should make its presence felt as a kind of hole. It was shown that one of these holes would appear to us as a particle with a positive energy and a positive charge and it was suggested that this particle should be identified with a proton. Subsequent investigations, however, have shown that this particle necessarily has the same mass as an electron† and also that, if it collides with an electron, the two will have a chance of annihilating one another much too great to be consistent with the known stability of matter.‡

It thus appears that we must abandon the identification of the holes with protons and must find some other interpretation for them. Following Oppenheimer,§ we can assume that in the world as we know it, *all*, and not merely nearly all, of the negative-energy states for electrons are occupied. A hole, if there were one, would be a new kind of particle, unknown to experimental physics, having the same mass and opposite charge to an electron. We may call such a particle an anti-electron. We should not expect to find any of them in nature, on account of their rapid rate of recombination with electrons, but if they could be produced experimentally in high vacuum they would be quite stable and amenable to observation. An encounter between two hard γ-rays (of energy at least half a million volts) could lead to the creation simultaneously of an electron and anti-electron, the probability of occurrence of this process being of the same order of magnitude as that of the collision of the two γ-rays on the assumption that they are spheres of the same size as classical

* 'Proc. Roy. Soc.,' A, vol. 126, p. 360 (1930).

† H. Weyl, 'Gruppentheorie und Quantenmechanik,' 2nd ed. p. 234 (1931).

‡ I. Tamm, 'Z. Physik,' vol. 62, p. 545 (1930); J. R. Oppenheimer, 'Phys. Rev.,' vol. 35, p. 939 (1930); P. Dirac, 'Proc. Camb. Philos. Soc.,' vol. 26, p. 361 (1930).

§ J. R. Oppenheimer, 'Phys. Rev.,' vol. 35, p. 562 (1930).

electrons. This probability is negligible, however, with the intensities of γ-rays at present available.

The protons on the above view are quite unconnected with electrons. Presumably the protons will have their own negative-energy states, all of which normally are occupied, an unoccupied one appearing as an anti-proton. Theory at present is quite unable to suggest a reason why there should be any differences between electrons and protons.

The object of the present paper is to put forward a new idea which is in many respects comparable with this one about negative energies. It will be concerned essentially, not with electrons and protons, but with the reason for the existence of a smallest electric charge. This smallest charge is known to exist experimentally and to have the value e given approximately by*

$$hc/e^2 = 137. \tag{1}$$

The theory of this paper, while it looks at first as though it will give a theoretical value for e, is found when worked out to give a connection between the smallest electric charge and the smallest magnetic pole. It shows, in fact, a symmetry between electricity and magnetism quite foreign to current views. It does not, however, force a complete symmetry, analogous to the fact that the symmetry between electrons and protons is not forced when we adopt Oppenheimer's interpretation. Without this symmetry, the ratio on the left-hand side of (1) remains, from the theoretical standpoint, completely undetermined and if we insert the experimental value 137 in our theory, it introduces quantitative differences between electricity and magnetism so large that one can understand why their qualitative similarities have not been discovered experimentally up to the present.

§ 2. *Non-integrable Phases for Wave Functions.*

We consider a particle whose motion is represented by a wave function ψ, which is a function of x, y, z and t. The precise form of the wave equation and whether it is relativistic or not, are not important for the present theory. We express ψ in the form

$$\psi = Ae^{i\gamma}, \tag{2}$$

where A and γ are real functions of x, y, z and t, denoting the amplitude and phase of the wave function. For a given state of motion of the particle, ψ will be determined except for an arbitrary constant numerical coefficient, which must be of modulus unity if we impose the condition that ψ shall be normalised.

* h means Planck's constant divided by 2π.

The indeterminacy in ψ then consists in the possible addition of an arbitrary constant to the phase γ. Thus the value of γ at a particular point has no physical meaning and only the difference between the values of γ at two different points is of any importance.

This immediately suggests a generalisation of the formalism. We may assume that γ has no definite value at a particular point, but only a definite difference in values for any two points. We may go further and assume that this difference is not definite unless the two points are neighbouring. For two distant points there will then be a definite phase difference only relative to some curve joining them and different curves will in general give different phase differences. The total change in phase when one goes round a closed curve need not vanish.

Let us examine the conditions necessary for this non-integrability of phase not to give rise to ambiguity in the applications of the theory. If we multiply ψ by its conjugate complex ϕ we get the density function, which has a direct physical meaning. This density is independent of the phase of the wave function, so that no trouble will be caused in this connection by any indeterminacy of phase. There are other more general kinds of applications, however, which must also be considered. If we take two different wave functions ψ_m and ψ_n, we may have to make use of the product $\phi_m\psi_n$. The integral

$$\int \phi_m \psi_n \, dx \, dy \, dz$$

is a number, the square of whose modulus has a physical meaning, namely, the probability of agreement of the two states. In order that the integral may have a definite modulus the integrand, although it need not have a definite phase at each point, must have a definite phase difference between any two points, whether neighbouring or not. Thus the change in phase in $\phi_m\psi_n$ round a closed curve must vanish. This requires that the change in phase in ψ_n round a closed curve shall be equal and opposite to that in ϕ_m and hence the same as that in ψ_m. We thus get the general result :— *The change in phase of a wave function round any closed curve must be the same for all the wave functions.*

It can easily be seen that this condition, when extended so as to give the same uncertainty of phase for transformation functions and matrices representing observables (referring to representations in which x, y and z are diagonal) as for wave functions, is sufficient to insure that the non-integrability of phase gives rise to no ambiguity in all applications of the theory. Whenever a ψ_n appears, if it is not multiplied into a ϕ_m, it will at

any rate be multiplied into something of a similar nature to a ϕ_m, which will result in the uncertainty of phase cancelling out, except for a constant which does not matter. For example, if ψ_n is to be transformed to another representation in which, say, the observables ξ are diagonal, it must be multiplied by the transformation function $(\xi \,|\, xyzt)$ and integrated with respect to x, y and z. This transformation function will have the same uncertainty of phase as a ϕ, so that the transformed wave function will have its phase determinate, except for a constant independent of ξ. Again, if we multiply ψ_n by a matrix $(x'y'z't' \,|\, \alpha \,|\, x''y''z''t)$, representing an observable α, the uncertainty in the phase as concerns the column [specified by x'', y'', z'', t] will cancel the uncertainty in ψ_n and the uncertainty as concerns the row will survive and give the necessary uncertainty in the new wave function $\alpha\psi_n$. The superposition principle for wave functions will be discussed a little later and when this point is settled it will complete the proof that all the general operations of quantum mechanics can be carried through exactly as though there were no uncertainty in the phase at all.

The above result that the change in phase round a closed curve must be the same for all wave functions means that this change in phase must be something determined by the dynamical system itself (and perhaps also partly by the representation) and must be independent of which state of the system is considered. As our dynamical system is merely a simple particle, it appears that the non-integrability of phase must be connected with the field of force in which the particle moves.

For the mathematical treatment of the question we express ψ, more generally than (2), as a product

$$\psi = \psi_1 e^{i\beta}, \tag{3}$$

where ψ_1 is any ordinary wave function (*i.e.*, one with a definite phase at each point) whose modulus is everywhere equal to the modulus of ψ. The uncertainty of phase is thus put in the factor $e^{i\beta}$. This requires that β shall not be a function of x, y, z, t having a definite value at each point, but β must have definite derivatives

$$\kappa_x = \frac{\partial\beta}{\partial x}, \qquad \kappa_y = \frac{\partial\beta}{\partial y}, \qquad \kappa_z = \frac{\partial\beta}{\partial z}, \qquad \kappa_0 = \frac{\partial\beta}{\partial t},$$

at each point, which do not in general satisfy the conditions of integrability $\partial\kappa_x/\partial y = \partial\kappa_y/\partial x$, etc. The change in phase round a closed curve will now be, by Stokes' theorem,

$$\int (\boldsymbol{\kappa},\, \mathbf{ds}) = \int (\mathrm{curl}\, \boldsymbol{\kappa},\, \mathbf{dS}), \tag{4}$$

where **ds** (a 4-vector) is an element of arc of the closed curve and **dS** (a 6-vector) is an element of a two-dimensional surface whose boundary is the closed curve. The factor ψ_1 does not enter at all into this change in phase.

It now becomes clear that the non-integrability of phase is quite consistent with the principle of superposition, or, stated more explicitly, that if we take two wave functions ψ_m and ψ_n both having the same change in phase round any closed curve, any linear combination of them $c_m\psi_m + c_n\psi_n$ must also have this same change in phase round every closed curve. This is because ψ_m and ψ_n will both be expressible in the form (3) with the same factor $e^{i\beta}$ (*i.e.*, the same κ's) but different ψ_1's, so that the linear combination will be expressible in this form with the same $e^{i\beta}$ again, and this $e^{i\beta}$ determines the change in phase round any closed curve. We may use the same factor $e^{i\beta}$ in (3) for dealing with all the wave functions of the system, but we are not obliged to do so, since only curl $\boldsymbol{\kappa}$ is fixed and we may use κ's differing from one another by the gradient of a scalar for treating the different wave functions.

From (3) we obtain

$$-ih\frac{\partial}{\partial x}\psi = e^{i\beta}\left(-ih\frac{\partial}{\partial x} + h\kappa_x\right)\psi_1, \tag{5}$$

with similar relations for the y, z and t derivatives. It follows that if ψ satisfies any wave equation, involving the momentum and energy operators **p** and W, ψ_1 will satisfy the corresponding wave equation in which **p** and W have been replaced by $\mathbf{p} + h\boldsymbol{\kappa}$ and $\mathrm{W} - h\kappa_0$ respectively.

Let us assume that ψ satisfies the usual wave equation for a free particle in the absence of any field. Then ψ_1 will satisfy the usual wave equation for a particle with charge $-e$ moving in an electromagnetic field whose potentials are

$$\mathbf{A} = hc/e \,.\, \boldsymbol{\kappa}, \qquad \mathrm{A}_0 = -h/e \,.\, \kappa_0. \tag{6}$$

Thus, since ψ_1 is just an ordinary wave function with a definite phase, our theory reverts to the usual one for the motion of an electron in an electromagnetic field. This gives a physical meaning to our non-integrability of phase. We see that we must have the wave function ψ always satisfying the same wave equation, whether there is a field or not, and the whole effect of the field when there is one is in making the phase non-integrable.

The components of the 6-vector curl $\boldsymbol{\kappa}$ appearing in (4) are, apart from numerical coefficients, equal to the components of the electric and magnetic fields **E** and **H**. They are, written in three-dimensional vector-notation,

$$\mathrm{curl}\,\boldsymbol{\kappa} = \frac{e}{hc}\mathbf{H}, \qquad \mathrm{grad}\,\kappa_0 - \frac{\partial\boldsymbol{\kappa}}{\partial t} = \frac{e}{h}\mathbf{E}. \tag{7}$$

The connection between non-integrability of phase and the electromagnetic field given in this section is not new, being essentially just Weyl's Principle of Gauge Invariance in its modern form.* It is also contained in the work of Iwanenko and Fock,† who consider a more general kind of non-integrability based on a general theory of parallel displacement of half-vectors. The present treatment is given in order to emphasise that non-integrable phases are perfectly compatible with all the general principles of quantum mechanics and do not in any way restrict their physical interpretation.

§ 3. *Nodal Singularities.*

We have seen in the preceding section how the non-integrable derivatives κ of the phase of the wave function receive a natural interpretation in terms of the potentials of the electromagnetic field, as the result of which our theory becomes mathematically equivalent to the usual one for the motion of an electron in an electromagnetic field and gives us nothing new. There is, however, one further fact which must now be taken into account, namely, that a phase is always undetermined to the extent of an arbitrary integral multiple of 2π. This requires a reconsideration of the connection between the κ's and the potentials and leads to a new physical phenomenon.

The condition for an unambiguous physical interpretation of the theory was that the change in phase round a closed curve should be the same for all wave functions. This change was then interpreted, by equations (4) and (7), as equal to (apart from numerical factors) the total flux through the closed curve of the 6-vector $\mathbf{E}$, $\mathbf{H}$ describing the electromagnetic field. Evidently these conditions must now be relaxed. The change in phase round a closed curve may be different for different wave functions by arbitrary multiples of 2π and is thus not sufficiently definite to be interpreted immediately in terms of the electromagnetic field.

To examine this question, let us consider first a very small closed curve. Now the wave equation requires the wave function to be continuous (except in very special circumstances which can be disregarded here) and hence the change in phase round a small closed curve must be small. Thus this change cannot now be different by multiples of 2π for different wave functions. It must have one definite value and may therefore be interpreted without

* H. Weyl, 'Z. Physik,' vol. 56, p. 330 (1929).

† D. Iwanenko and V. Fock, 'C. R.,' vol. 188, p. 1470 (1929); V. Fock, 'Z. Physik,' vol. 57, p. 261 (1929). The more general kind of non-integrability considered by these authors does not seem to have any physical application.

ambiguity in terms of the flux of the 6-vector **E**, **H** through the small closed curve, which flux must also be small.

There is an exceptional case, however, occurring when the wave function vanishes, since then its phase does not have a meaning. As the wave function is complex, its vanishing will require two conditions, so that in general the points at which it vanishes will lie along a line.* We call such a line a *nodal line.* If we now take a wave function having a nodal line passing through our small closed curve, considerations of continuity will no longer enable us to infer that the change in phase round the small closed curve must be small. All we shall be able to say is that the change in phase will be close to $2\pi n$ where n is some integer, positive or negative. This integer will be a characteristic of the nodal line. Its sign will be associated with a direction encircling the nodal line, which in turn may be associated with a direction along the nodal line.

The difference between the change in phase round the small closed curve and the nearest $2\pi n$ must now be the same as the change in phase round the closed curve for a wave function with no nodal line through it. It is therefore this difference that must be interpreted in terms of the flux of the 6-vector **E**, **H** through the closed curve. For a closed curve in three-dimensional space, only magnetic flux will come into play and hence we obtain for the change in phase round the small closed curve

$$2\pi n + e/hc \,.\, \int (\mathbf{H},\, \mathbf{dS}).$$

We can now treat a large closed curve by dividing it up into a network of small closed curves lying in a surface whose boundary is the large closed curve. The total change in phase round the large closed curve will equal the sum of all the changes round the small closed curves and will therefore be

$$2\pi \Sigma n + e/hc \,.\, \int (\mathbf{H},\, \mathbf{dS}), \tag{8}$$

the integration being taken over the surface and the summation over all nodal lines that pass through it, the proper sign being given to each term in the sum. This expression consists of two parts, a part $e/hc \,.\, \int (\mathbf{H},\, \mathbf{dS})$ which must be the same for all wave functions and a part $2\pi \Sigma n$ which may be different for different wave functions.

* We are here considering, for simplicity in explanation, that the wave function is in three dimensions. The passage to four dimensions makes no essential change in the theory. The nodal lines then become two-dimensional nodal surfaces, which can be encircled by curves in the same way as lines are in three dimensions.

Expression (8) applied to any surface is equal to the change in phase round the boundary of the surface. Hence expression (8) applied to a closed surface must vanish. It follows that Σn, summed for all nodal lines crossing a closed surface, must be the same for all wave functions and must equal $-e/2\pi hc$ times the total magnetic flux crossing the surface.

If Σn does not vanish, some nodal lines must have end points inside the closed surface, since a nodal line without such end point must cross the surface twice (at least) and will contribute equal and opposite amounts to Σn at the two points of crossing. The value of Σn for the closed surface will thus equal the sum of the values of n for all nodal lines having end points inside the surface. This sum must be the same for all wave functions. Since this result applies to *any* closed surface, it follows that *the end points of nodal lines must be the same for all wave functions. These end points are then points of singularity in the electromagnetic field.* The total flux of magnetic field crossing a small closed surface surrounding one of these points is

$$4\pi\mu = 2\pi nhc/e,$$

where n is the characteristic of the nodal line that ends there, or the sum of the characteristics of all nodal lines ending there when there is more than one. Thus at the end point there will be a magnetic pole of strength

$$\mu = \tfrac{1}{2}nhc/e.$$

Our theory thus allows isolated magnetic poles, but the strength of such poles must be quantised, the quantum μ_0 being connected with the electronic charge e by

$$hc/e\mu_0 = 2. \tag{9}$$

This equation is to be compared with (1). The theory also requires a quantisation of electric charge, since any charged particle moving in the field of a pole of strength μ_0 must have for its charge some integral multiple (positive or negative) of e, in order that wave functions describing the motion may exist.

§ 4. *Electron in Field of One-Quantum Pole.*

The wave functions discussed in the preceding section, having nodal lines ending on magnetic poles, are quite proper and amenable to analytic treatment by methods parallel to the usual ones of quantum mechanics. It will perhaps help the reader to realise this if a simple example is discussed more explicitly.

Let us consider the motion of an electron in the magnetic field of a one-

quantum pole when there is no electric field present. We take polar co-ordinates r, θ, ϕ with the magnetic pole as origin. Every wave function must now have a nodal line radiating out from the origin.

We express our wave function ψ in the form (3), where β is some non-integrable phase having derivatives κ that are connected with the known electromagnetic field by equations (6). It will not, however, be possible to obtain κ's satisfying these equations all round the magnetic pole. There must be some singular line radiating out from the pole along which these equations are not satisfied, but this line may be chosen arbitrarily. We may choose it to be the same as the nodal line for the wave function under consideration, which would result in ψ_1 being continuous. This choice, however, would mean different κ's for different wave functions (the difference between any two being, of course, the four-dimensional gradient of a scalar, except on the singular lines). This would perhaps be inconvenient and is not really necessary. We may express all our wave functions in the form (3) with the same $e^{i\beta}$, and then those wave functions whose nodal lines do not coincide with the singular line for the κ's will correspond to ψ_1's having a certain kind of discontinuity on this singular line, namely, a discontinuity just cancelling with the discontinuity in $e^{i\beta}$ here to give a continuous product.

The magnetic field $\mathbf{H}$, lies along the radial direction and is of magnitude μ_0/r^2, which by (9) equals $\frac{1}{2}hc/er^2$. Hence, from equations (7), curl $\boldsymbol{\kappa}$ is radial and of magnitude $1/2r^2$. It may now easily be verified that a solution of the whole of equations (7) is

$$\kappa_0 = 0, \qquad \kappa_r = \kappa_\theta = 0, \qquad \kappa_\phi = 1/2r \,.\, \tan \tfrac{1}{2}\theta, \tag{10}$$

where κ_r, κ_θ, κ_ϕ are the components of $\boldsymbol{\kappa}$ referred to the polar co-ordinates. This solution is valid at all points except along the line $\theta = \pi$, where κ_ϕ becomes infinite in such a way that $\int (\boldsymbol{\kappa}, \mathbf{ds})$ round a small curve encircling this line is 2π. We may refer all our wave functions to this set of κ's.

Let us consider a stationary state of the electron with energy W. Written non-relativistically, the wave equation is

$$-h^2/2m \,.\, \nabla^2\psi = \mathrm{W}\psi.$$

If we apply the rule expressed by equation (5), we get as the wave equation for ψ_1

$$-h^2/2m \,.\, \{\nabla^2 + i\,(\boldsymbol{\kappa}, \boldsymbol{\nabla}) + i\,(\dot{\boldsymbol{\nabla}}, \boldsymbol{\kappa}) - \kappa^2\}\,\psi_1 = \mathrm{W}\psi_1. \tag{11}$$

The values (10) for the κ's give

$$(\boldsymbol{\kappa}, \boldsymbol{\nabla}) = (\boldsymbol{\nabla}, \boldsymbol{\kappa}) = \kappa_\phi \frac{1}{r \sin\theta} \frac{\partial}{\partial \phi} = \frac{1}{4r^2} \sec^2 \tfrac{1}{2}\theta \frac{\partial}{\partial \phi}$$

$$\boldsymbol{\kappa}^2 = \kappa_\phi{}^2 = \frac{1}{4r^2} \tan^2 \tfrac{1}{2}\theta,$$

so that equation (11) becomes

$$-\frac{h^2}{2m}\left\{\nabla^2 + \frac{i}{2r^2} \sec^2 \tfrac{1}{2}\theta \frac{\partial}{\partial \phi} - \frac{1}{4r^2} \tan^2 \tfrac{1}{2}\theta\right\} \psi_1 = \mathrm{W}\psi_1.$$

We now suppose ψ_1 to be of the form of a function f of r only multiplied by a function S of θ and ϕ only, *i.e.*,

$$\psi_1 = f(r)\, \mathrm{S}\,(\theta\phi).$$

This requires

$$\left\{\frac{d^2}{dr^2} + \frac{2}{r}\frac{d}{dr} - \frac{\lambda}{r^2}\right\} f = -\frac{2m\mathrm{W}}{h^2} f, \tag{12}$$

$$\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} + \tfrac{1}{2}i \sec^2\tfrac{1}{2}\theta \frac{\partial}{\partial\phi} - \tfrac{1}{4}\tan^2\tfrac{1}{2}\theta\right\} \mathrm{S} = -\lambda \mathrm{S}, \tag{13}$$

where λ is a number.

From equation (12) it is evident that there can be no stable states for which the electron is bound to the magnetic pole, because the operator on the left-hand side contains no constant with the dimensions of a length. This result is what one would expect from analogy with the classical theory. Equation (13) determines the dependence of the wave function on angle. It may be considered as a generalisation of the ordinary equation for spherical harmonies.

The lowest eigenvalue of (13) is $\lambda = \frac{1}{2}$, corresponding to which there are two independent wave functions

$$\mathrm{S}_a = \cos\tfrac{1}{2}\theta, \quad \mathrm{S}_b = \sin\tfrac{1}{2}\theta\, e^{i\phi},$$

as may easily be verified by direct substitution. The nodal line for S_a is $\theta = \pi$, that for S_b is $\theta = 0$. It should be observed that S_a is continuous everywhere, while S_b is discontinuous for $\theta = \pi$, its phase changing by 2π when one goes round a small curve encircling the line $\theta = \pi$. This is just what is necessary in order that both S_a and S_b, when multiplied by the $e^{i\beta}$ factor, may give continuous wave functions ψ. The two ψ's that we get in this way are both on the same footing and the difference in behaviour of S_a and S_b is due to our having chosen κ's with a singularity at $\theta = \pi$.

The general eigenvalue of (13) is $\lambda = n^2 + 2n + \frac{1}{2}$. The general solution of this wave equation has been worked out by I. Tamm.*

* Appearing probably in 'Z. Physik.'

§ 5. *Conclusion.*

Elementary classical theory allows us to formulate equations of motion for an electron in the field produced by an arbitrary distribution of electric charges and magnetic poles. If we wish to put the equations of motion in the Hamiltonian form, however, we have to introduce the electromagnetic potentials, and this is possible only when there are no isolated magnetic poles. Quantum mechanics, as it is usually established, is derived from the Hamiltonian form of the classical theory and therefore is applicable only when there are no isolated magnetic poles.

The object of the present paper is to show that quantum mechanics does not really preclude the existence of isolated magnetic poles. On the contrary, the present formalism of quantum mechanics, when developed naturally without the imposition of arbitrary restrictions, leads inevitably to wave equations whose only physical interpretation is the motion of an electron in the field of a single pole. This new development requires *no change whatever* in the formalism when expressed in terms of abstract symbols denoting states and observables, but is merely a generalisation of the possibilities of representation of these abstract symbols by wave functions and matrices. Under these circumstances one would be surprised if Nature had made no use of it.

The theory leads to a connection, namely, equation (9), between the quantum of magnetic pole and the electronic charge. It is rather disappointing to find this reciprocity between electricity and magnetism, instead of a purely electronic quantum condition, such as (1). However, there appears to be no possibility of modifying the theory, as it contains no arbitrary features, so presumably the explanation of (1) will require some entirely new idea.

The theoretical reciprocity between electricity and magnetism is perfect. Instead of discussing the motion of an electron in the field of a fixed magnetic pole, as we did in § 4, we could equally well consider the motion of a pole in the field of fixed charge. This would require the introduction of the electromagnetic potentials B satisfying

$$\mathbf{E} = \text{curl } \mathbf{B}, \quad \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} + \text{grad } \mathrm{B}_0,$$

to be used instead of the A's in equations (6). The theory would now run quite parallel and would lead to the same condition (9) connecting the smallest pole with the smallest charge.

There remains to be discussed the question of why isolated magnetic poles are not observed. The experimental result (1) shows that there must be some

cause of dissimilarity between electricity and magnetism (possibly connected with the cause of dissimilarity between electrons and protons) as the result of which we have, not $\mu_0 = e$, but $\mu_0 = 137/2 \, . \, e$. This means that the attractive force between two one-quantum poles of opposite sign is $(137/2)^2 = 4692\frac{1}{4}$ times that between electron and proton. This very large force may perhaps account for why poles of opposite sign have never yet been separated.

Estimation of Metals in Solution by Means of their Spark Spectra.

By F. Twyman, F.Inst.P., F.R.S., and C. Stansfield Hitchen, A.R.C.S., Ph.D.

(Received May 29, 1931.—Revised August 5, 1931.)

[Plates 5, 6.]

Introduction.

This paper records experiments undertaken to extend to liquids the accuracy of quantitative analysis recently attained by the spectrography of alloys. How to produce from a solution a spectrum which shall truly represent the solution, and what effect may be expected from the presence of other metals than the one under determination, are among the questions dealt with, for they must come into consideration in applying any method of quantitative spectrography of solutions.

The Production of the Spark.

(*a*) *The Sparking Vessel for Liquids.*—At the outset of the investigation it became apparent that the older forms of sparking apparatus, used by Hartley, Pollok and Leonard, and others, were unsuitable for quantitative work, for it was found that, owing to incrustation of the electrodes and to decomposition of the solution around them, the spark soon became unrepresentative of the bulk of the solution.

The apparatus finally devised embodies two principles which a number of trials showed to be necessary :—

(i) The spark takes place from liquid to liquid.

(ii) There is a steady feed of fresh liquid, any scum being carried away.