Monopoles in Quantum Field Theory

Proceedings of the Monopole Meeting
December 1981

Edited by
N. S. CRAIGIE
P. GODDARD
W. NAHM

Monopoles in Quantum Field Theory

Proceedings of the Monopole Meeting
Trieste, Italy
December 1981

Edited by N.S. Craigie
P. Goddard
W. Nahm

World Scientific

World Scientific Publishing Co Pte Ltd
P.O. Box 128
Farrer Road
Singapore 9128

ISBN 9971-950-29-4 pbk 9971-950-28-6
Printed by Singapore National Printers (Pte) Ltd

F O R E W O R D

The International Centre for Theoretical Physics, Trieste, is an institution devoted to the promotion of research in pure and applied science throughout the world, with special emphasis on developing countries. On 11-15 December 1981, 125 physicists and mathematicians from 37 countries assembled here for a conference. The occasion was the 50th anniversary of the introduction, by Dirac, of the concept of the magnetic monopole.

The meeting, organized by the ICTP in collaboration with the Istituto Nazionale di Fisica Nucleare (INFN), consisted of four days of seminars and discussions on the numerous recent theoretical developments which have emerged out of Dirac's original monopole idea, and on the search for such entities in nature.

We hope these proceedings serve as a useful and comprehensive source of information on the various aspects of a subject which is having a remarkable impact on present-day quantum field theory.

ACKNOWLEDGMENTS

The organizers wish to thank all the speakers and participants for making the meeting an interesting and, we believe, a fruitful one, and the staff of the ICTP for all their help in organizing it. We would like to express our gratitude to L. O'Raifeartaigh for sending us, at our request, a written version of his delightful summary talk, despite his recent illness. Finally we wish to thank Sunil Mukhi for his help in putting the proceedings together.

N.S. Craigie
W. Nahm
P. Goddard

Department of Physics

The Florida State University
Tallahassee, Florida 32306

11 Nov 1981

Dear Abdus,

I am sorry I cannot come to your monopole conference. It would be too much of a dislocation for me at such short notice. It was very kind of you to invite me.

I am inclined now to believe that monopoles do not exist. So many years have gone by without any encouragement from the experimental side. It will be interesting to see if your conference can produce any new angles of attack on the problem.

With best wishes,

Yours sincerely

Paul Dirac

OPENING ADDRESS BY PROFESSOR ABDUS SALAM

The subject of magnetic monopoles began with Dirac's paper on Quantised singularities in the Electromagnetic field which was received by the Royal Society on 29th March 1931. Dirac's motivation in this paper was to find the reason for the existence of the smallest electric charge. Proposing a generalisation of the formalism of quantum mechanics as used for electromagnetism, Dirac allowed for wave functions with non-integrable phases. He showed that the non-integrability of the phases could be interpreted in terms of the presence of an electromagnetic field, with the possibility of there being singularities of the field. These corresponded to single magnetic poles with their strength restricted by the relation $eg/(4\pi) = (n/2)$. This relation provided Dirac with the basis for the explanation of electric charge quantisation. Already in this paper Dirac commented on the difficulty of creating pole pairs due to their strong binding.

This paper contained much else besides the notion of monopoles and the relation $eg/(4\pi) = (n/2)$. Dirac, in fact, started by commenting on a paper he had written a year before, where he had suggested the idea of filling the negative sea of states and identifying the holes in the negative sea as protons. In the 1931 paper he accepted the results of Weyl about the value of the mass of such a hole. Following a suggestion due to Oppenheimer, he concluded that such a hole should be "a new kind of particle unknown to experimental physics". This he called an anti-electron. Protons, he commented, are unconnected with holes in the electron sea but they should have their own anti-protons. One may recall that this was fully one year before Anderson's discovery of positive electrons, and at a time when unlike to-day theoretical physicists were very reluctant to postulate new particles, at the drop of a hat.

After 1931, and before 1948 when Dirac wrote on the subject again, the theory of monopoles and of bound states of monopoles and electric charges was worked on extensively by Tamm, Fierz, Banderet, Harish-Chandra, and others. In this connection - and I owe this history to a beautiful review by Amaldi and Cabibbo published in Dirac's 70th Birthday Volume - it is good to recall that already around 1895, Poincare and J.J. Thomson had published the result that the static fields generated by a charge $+e$ placed at a distance of r from the pole of charge $+g$ gives rise to an angular momentum $eg/(4\pi)$, with the direction of the angular momentum pointing from $+g$ to $+e$. Saha in 1936 used this result to remark that the Dirac relationship could be obtained by equating this angular momentum $eg/(4\pi)$ to an integral multiple of $(1/2)\hbar$. In modern parlance he was suggesting that a bound state of a spinless charge and a spinless monopole could carry half a unit of spin - an idea which has been advanced recently for motivating supersymmetry dynamically. Saha also suggested that the neutron might consist of bound pole pairs. This anticipated in some ways another idea which has been used by Schwinger, Barut and others to suggest that the basic entities of which all matter may be composed of may be monopoles carrying electric charge (dyons).

The next step in the development of monopole theory came with Dirac's paper of 1948. I can recall, as an undergraduate in Cambridge, a popular evening lecture given by Dirac where he spoke about the Hamiltonian for a system consisting of a fixed number of poles and charges. Introducing electromagnetic potentials, Dirac motivated the well known strings, one attached to each pole. The condition $eg/(4\pi) = (n/2)$ ensured that the coordinates describing the string were ignorable.

I recall that Dirac finished his lecture by passing around the packed hall a set of photographic plates sent to him, if my memory serves me correctly, by Ehrenheft who claimed to have discovered ionising tracks which might, he suggested, be monopoles. Dirac said he did not believe this explanation of the tracks seen.

For our meeting to-day, 50 years after his first paper, Dirac has sent us the following message:

"I am sorry I cannot come to your monopole conference. It would be too much of a dislocation for me at such short notice. It was very kind of you to invite me.

"I am inclined now to believe that monopoles do not exist. So many years have gone by without any encouragement from the experimental side. It will be interesting to see if your conference can produce any new angle of attack on the problem."

Dirac, who concluded his first 1931 paper with the remark "one would be surprised if Nature had not made use" of the monopole concept, is as concerned to-day with the lack of experimental evidence for monopoles as he was in 1948. However, in the meanwhile the theory of monopoles and the experimental prospects of discovering them have undergone a complete transformation since 1974 with the pioneering work of 't Hooft and Polyakov in the context of modern grand unifying gauge theories. There is indeed a "new angle of attack on the problem" which we are assembled to hear about in the next five days, and which posits that monopoles are likely to have been abundant in the early Universe and that if they have survived to the present epoch, we should be looking for objects with masses around 10^{-8} grams. This represents an experimental regime completely different than the one researched on hitherto. A second aspect of this new revolution is the deepening of contact between modern mathematics and modern particle physics, leading to enrichment of both disciplines. This would delight Dirac, since in his 1931 monopole paper his opening remark concerns the abstraction in mathematics to be expected as a consequence of interaction with physics: "... It seems likely that this process of increasing abstraction will continue in the future and that the advance in physics is to be associated with a continual modification and generalisation of the axioms at the base of the mathematics..."

As you know Dirac will be eighty on 8th August 1982. I am sure you will wish me to cable him the following message on behalf of all those here to-day:

"Around one hundred and twenty physicists and mathematicians from 37 countries assembled at the International Centre for Theoretical Physics at Trieste, for a symposium on developments in Monopole Theory and Experiment, wish to express their deepest appreciation to you on the fiftieth anniversary of your seminal paper on the subject and extend to you and to your family their warmest greetings."

Copy of Paul Dirac's original article published in Proceedings of the Royal Society of London, A, Vol.133 (1931).

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*

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

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

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

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,

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

where $\mathbf{ds}$ (a 4-vector) is an element of arc of the closed curve and $\mathbf{dS}$ (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 $\mathbf{p}$ and W, ψ_1 will satisfy the corresponding wave equation in which $\mathbf{p}$ 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,

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