Aharonov-Bohm effect and gauge invariance

1983 J. Phys. A: Math. Gen. 16 2173

(http://iopscience.iop.org/0305-4470/16/10/013)

J. Phys. A: Math. Gen. 16 (1983) 2173–2177. Printed in Great Britain

Aharonov–Bohm effect and gauge invariance

M Bawin† and A Burnel
Institut de Physique B5, Université de Liège, Sart Tilman, B-4000, Liège 1, Belgium

Received 3 December 1982, in final form 8 February 1983

Abstract. We study the Aharonov–Bohm effect in a gauge for which the vector potential vanishes wherever the magnetic field does. We then show how gauge invariance implies the existence of Aharonov–Bohm scattering and excludes solutions recently discussed in the literature.

1. Introduction

The Aharonov–Bohm (AB) effect, which purports to establish the physical importance of vector potentials in quantum mechanics (Aharonov and Bohm 1959) remains a controversial topic. In particular, Bocchieri and Loinger (1978) have denied the very existence of the AB effect, while Henneberger (1981) has questioned its original interpretation by Aharonov and Bohm (1959). Specifically, while Henneberger's solution allows for the existence of the AB effect, it does not lead to AB scattering, i.e. the corresponding cross section for scattering of a charged particle off an inaccessible solenoid is zero. An extensive list of references concerning the AB effect and its surrounding controversy can be found in the paper by Ruijsenaars (1981) who discusses in detail both Aharonov and Bohm's classical analysis and Henneberger's viewpoint.

In this paper we re-examine the problem of quantum mechanical scattering of a charged particle off an infinitely long and inaccessible solenoid. In addition to some pedagogical remarks concerning the discussion of constants of the motion in Aharonov and Bohm's and Henneberger's solutions, we wish to show in detail that this problem can be studied in a gauge where the vector potential $\boldsymbol{A}$ vanishes wherever the magnetic field $\boldsymbol{B}$ does. The unusual feature of the vector potential in our gauge is that it is a multivalued function which requires a cut in space and a fibre bundle description. Such a description together with gauge invariance implies the existence of the AB effect and allows us to reject Henneberger's solution. A similar gauge was formally used by Wilczek (1982a, b) in related problems.

We organise our work as follows. In § 2, we explicitly show how both the particle angular momentum L_Z and the canonical momentum p_θ commute with the Hamiltonian provided the magnetic field and its first derivative have no discontinuity at the boundary of the inaccessible region. Thus, both Aharonov and Bohm's solution (which quantises p_θ) and Henneberger's solution (which quantises L_Z) are *a priori* acceptable. In § 3 we introduce our gauge and discuss why it implies AB scattering. Section 4 contains our conclusions.

† Chercheur qualifié FNRS.

2. Aharonov–Bohm effect and constants of the motion

In order to study scattering from an inaccessible and infinite solenoid, we introduce the usual vector potential defined by

$$A_\theta = \frac{1}{r}\int_0^r B_Z(r')r'\,\mathrm{d}r', \qquad A_Z = A_r = 0, \tag{1}$$

where r, θ, Z are cylindrical coordinates.

Formula (1) describes the vector potential corresponding to a magnetic field B_Z along the Z axis, which is only a function of the radius r.

A form of $B_Z(r)$ often considered in the literature is

$$B_Z = B_0 H(a-r), \tag{2}$$

where B_0 is a constant and $H(x)$ is the usual step function. We shall not, however, consider B_Z to be restricted as in (2).

The Schrödinger equation for a charged particle of mass m in the field described by (1) is

$$-\frac{1}{2m}\left[\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}-\left(\frac{p_\theta - erA_\theta}{r}\right)^2\right]\psi = E\psi, \tag{3}$$

where we have neglected the trivial Z-dependence and used cylindrical coordinates and a system of units such that $\hbar = c = 1$. The field B_Z is assumed to be zero for $r > a$, where a denotes the radius of the inaccessible cylindrical region. The boundary condition at $r = a$ is then

$$\psi(a) = 0. \tag{4}$$

Using (3), we now find, for $r \in [a, \infty)$,

$$-(2mH, A_\theta] = \left[\frac{\mathrm{d}^2}{\mathrm{d}r^2}+\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}, rA_\theta\right] = 2B_Z + r\frac{\mathrm{d}B_Z}{\mathrm{d}r} + 2rB_Z\frac{\mathrm{d}}{\mathrm{d}r}. \tag{5}$$

From (5), we find that H and erA_θ commute on the whole space $\mathscr{H}$ of physical states, provided *both* B_Z and $\mathrm{d}B_Z/\mathrm{d}r$ vanish at the boundary of the inaccessible region. As discussed by Henneberger (1981) and Peshkin (1981), quantisation of p_θ would then imply the existence of AB scattering, while quantisation of $p_\theta - erA_\theta \equiv L_Z$ would imply its non-existence. Note that, strictly speaking, the magnetic field given by (2), being discontinuous, does not lead to a conserved L_Z and *implies* AB scattering.

3. Gauge invariance and AB effect

Instead of (1), let us now consider the vector potential defined by

$$\boldsymbol{A}' = -\theta B_Z \boldsymbol{r} \tag{6}$$

for which we also have

$$\boldsymbol{\nabla}\times\boldsymbol{A}' = B_Z\boldsymbol{e}_z. \tag{7}$$

It is clear that we can also obtain $\boldsymbol{A}'$ from (1) by the gauge transformation

$$\boldsymbol{A}' = \boldsymbol{A} - \boldsymbol{\nabla}(\Lambda\theta/2\pi) \tag{8}$$

where

$$\Lambda(r) = 2\pi \int_0^r r' B_Z \, dr' \tag{9}$$

is univocally fixed by the condition

$$\Lambda(r = 0) = 0. \tag{10}$$

Our potential (6), which vanishes where $\boldsymbol{B}$ does, is a multivalued function. In order to treat such functions properly, we must cut the space with a half-plane and work in the interval $0 < \theta < 2\pi$. This cut cannot play any physical role, and if we want to define a vector potential in the whole space, we must consider the space as the union of two regions $0 < \theta < 2\pi$ and $-\pi < \theta' < \pi$, for instance. In both regions, the potential is defined by (6), i.e. in the whole space, the potential is defined by a set of two different functions:

$$\boldsymbol{A}'_1 = -\boldsymbol{r} B_Z \theta, \qquad 0 < \theta < 2\pi, \tag{11}$$

$$\boldsymbol{A}'_2 = -\boldsymbol{r} B_Z \theta', \qquad -\pi < \theta' < \pi. \tag{12}$$

The two domains have two intersecting regions R_{I} and R_{II} defined respectively by $0 < \theta < \pi$ and $\pi < \theta < 2\pi$. In R_{I}, we have $\theta = \theta'$ and $\boldsymbol{A}'_1 = \boldsymbol{A}'_2$, while in R_{II}, we have $\theta = \theta' + 2\pi$ and

$$\boldsymbol{A}'_1 = \boldsymbol{A}'_2 - 2\pi \boldsymbol{r} B_Z = \boldsymbol{A}'_2 - \boldsymbol{\nabla}\Lambda \tag{13}$$

where Λ is as in (9) and (10).

Formula (13) shows that in the two intersecting regions, the potential is represented by two functions which only differ by a gradient. Such a construction corresponds to a definition of the potential that is more general than the usual one (a representation by a single function). It is a one-form connection on a principal fibre bundle which, in general, is represented by a set of functions defined on different domains such that in the intersection of two domains the functions only differ by a gradient. Further mathematical details concerning the existence and uniqueness of such one-form connections can be found in Daniel and Viallet (1980). In the case of the Dirac monopole, this construction is the only possible way to obtain a well defined vector potential (Wu and Yang 1975, 1976, 1977). In each domain, we have a different Schrödinger equation:

$$H_i \psi'_i = E \psi'_i, \qquad i = 1, 2, \tag{14}$$

with

$$H_i = (2m)^{-1} (p - e\boldsymbol{A}'_i)^2 \tag{15}$$

and we have

$$\begin{aligned} \psi'_1 &= \psi'_2 && \text{in } R_{\mathrm{I}}, \\ \psi'_1 &= \mathrm{e}^{-\mathrm{i}\Lambda} \psi'_2 && \text{in } R_{\mathrm{II}}, \end{aligned} \tag{16}$$

by virtue of gauge invariance. If we wish to solve univocally the Schrödinger equation we must impose boundary conditions on the domains $-\pi < \theta' < \pi$ or $0 < \theta < 2\pi$. Equivalently, we can extend the domain of θ-values to be such that we have

$$-\infty < \theta < \infty,$$

provided we impose periodic boundary conditions on θ. According to (16), this periodic boundary condition is

$$\psi'(\theta+2\pi)=\psi'(\theta)\,\mathrm{e}^{-\mathrm{i}\Lambda} \qquad (17)$$

where $\psi'(\theta)$ denotes a solution to the Schrödinger equation with the potential $\boldsymbol{A}'$ given by (6). The corresponding boundary conditions for $\psi_1'(\theta)$ and $\psi_2'(\theta)$ would read

$$\psi_2'(\pi)=\psi_2'(-\pi)\,\mathrm{e}^{-\mathrm{i}\Lambda} \qquad (17')$$

and

$$\psi_1'(2\pi)=\psi_1'(0)\,\mathrm{e}^{-\mathrm{i}\Lambda}. \qquad (17'')$$

Note that Λ in (16), (17) is nothing but the total flux of the magnetic field, as ψ is non-zero only for $r>a$.

Equation (17) implies

$$\psi'(m,\theta)\propto\exp[\mathrm{i}(m-\Lambda/2\pi)\theta], \qquad (18)$$

where m is an arbitrary integer and $\propto$ means 'proportional to'. In each domain $\boldsymbol{R}_{\mathrm{I}}$ and $\boldsymbol{R}_{\mathrm{II}}$, we also have

$$\psi_i'(m,\theta)\propto\exp[\mathrm{i}(m-\Lambda/2\pi)\theta] \qquad (i=1,2). \qquad (19)$$

If we now come back to the usual gauge (1) through the inverse transformation of that given in (8), we obtain for the wavefunctions

$$\psi_i(m,\theta)\propto\mathrm{e}^{\mathrm{i}m\theta} \qquad (i=1,2). \qquad (20)$$

In $\boldsymbol{R}_{\mathrm{I}}$, we have $\psi_1(m,\theta)=\psi_2(m,\theta)$, while in $\boldsymbol{R}_{\mathrm{II}}$ we have $\psi_1(m,\theta)=\mathrm{e}^{2\mathrm{i}\pi m}\psi_2(m,\theta)=\psi_2(m,\theta)$, since m is an integer. The above manipulations consist in effect in making a *global* gauge transformation (Daniel and Viallet 1980). From (18) we now get

$$\psi(m,\theta)\propto\mathrm{e}^{\mathrm{i}m\theta} \qquad (21)$$

where $\psi(m,\theta)$ satisfies equation (3) in gauge (1). In other words, the periodic boundary condition (17) uniquely selects the angular dependence in $\psi'(m,\theta)$ and hence in $\psi(m,\theta)$, obtained from $\psi'(m,\theta)$ through a gauge transformation. The angular dependence in formula (21) is indeed the one found by Aharonov and Bohm (1959), so that in our approach, the solution $\tilde{\psi}_m$ suggested by Henneberger (1981), i.e.

$$\tilde{\psi}_m(\theta)\propto\exp[\mathrm{i}(m-\Lambda/2\pi)\theta], \qquad (22)$$

is ruled out by gauge invariance. That such a solution must be rejected can also be seen when we work with the usual vector potential (1). In that case, due to the assumed non-periodicity of the wavefunction (see formula (22)), a fibre bundle formulation is also needed. Since now $\boldsymbol{A}_1=\boldsymbol{A}_2$ everywhere, we must have $\psi_1=\psi_2$ everywhere, and the wavefunction is periodic, in contradiction with (22).

A remark with respect to Stokes' theorem is in order. Clearly, our potential A' in (6) violates Stokes' theorem for a curve winding around the Z axis. However, such a curve cannot be drawn as it crosses the cut. Stokes' theorem only (trivially) applies to curves which do not cross the cut.

4. Conclusions

We have discussed in detail why, for magnetic fields that are continuous at the boundary of the inaccessible region, together with their first derivatives, the Schrödinger equation does not lead unambiguously to AB scattering. We then showed, within fibre bundle theory, that gauge invariance *does* lead to the original AB solution and excludes the solution proposed by Henneberger (1981).

References

Aharonov Y and Bohm D 1959 *Phys. Rev.* **115** 485
Bocchieri P and Loinger A 1978 *Nuovo Cimento* **47A** 475
Daniel M and Viallet C M 1980 *Rev. Mod. Phys.* **52** 175
Henneberger W C 1981 *J. Math. Phys.* **22** 116
Peshkin M 1981 *Phys. Rep.* C **80** 375
Ruijsenaars S N M 1981 *The Aharonov–Bohm Effect and Scattering Theory, preprint*
Wilczek F 1982a *Phys. Rev. Lett.* **48** 1144
—— 1982b *Phys. Rev. Lett.* **49** 957
Wu and Yang C N 1975 *Phys. Rev.* D **12** 3845
—— 1976 *Nucl. Phys.* B **107** 365
—— 1977 *Phys. Rev.* D **16** 1018