

CRITICISMS OF THE $U(1)$ INVARIANT THEORY OF THE AHARONOV BOHM AND SAGNAC EFFECT AND ADVANTAGES OF AN $O(3)$ INVARIANT THEORY

by

Petar K. Anastasovski¹, T. E. Bearden², C. Ciubotariu³, W. T. Coffey⁴, L. B. Crowell⁵,
G. J. Evans⁶, M. W. Evans^{7,8}, R. Flower⁹, A. Labounsky¹⁰, B. Lehnert¹¹, M. Mészáros¹²,
P. R. Molnár¹², S. Roy¹³, and J. P. Vigiér¹⁴

Institute for Advanced Study, Alpha Foundation, Institute of Physics,
11 Rutafa Street, Building H, Budapest H-1165, Hungary

Also at:

¹Faculty of Technology and Metallurgy, Department of Physics,
University of Skopje, Republic of Macedonia

²CEO, CTEC Inc., 2311 Big Cove Road, Huntsville, AL 35801-1351.

³Institute for Information Technology, Stuttgart University, Stuttgart, Germany

⁴Department of Microelectronics and Electrical Engineering,
Trinity College, Dublin 2, Ireland.

⁵Department of Physics and Astronomy, University of New Mexico,
Albuquerque, New Mexico

⁶Ceredigion County Council, Aberaeron, Wales, Great Britain

⁷former Edward Davies Chemical Laboratories, University College of Wales,
Aberystwyth SY23 1NE, Wales, Great Britain

⁸sometime JRF, Wolfson College, Oxford, Great Britain

⁹CEO, Applied Science Associates, and Temple University,
Philadelphia, Pennsylvania, USA.

¹⁰The Boeing Company, Huntington Beach, California.

¹¹Alfvén Laboratory, Royal Institute of Technology, Stockholm, S-100 44, Sweden.

¹²Alpha Foundation, Institute of Physics, 11 Rutafa Street, Building H, H-1165, Hungary.

¹³Indian Statistical Institute, Calcutta, India.

¹⁴Labo de Gravitation et Cosmologie Relativistes, Université Pierre et Marie Curie,
Tour 22-12, 4 ème étage, BP 142, 4 Place Jussieu, 75252 Paris Cedex 05, France.

KEYWORDS: Criticisms of the $U(1)$ invariant theories of the Aharonov Bohm
and Sagnac effects; $O(3)$ invariant theories

ABSTRACT

The U(1) invariant electrodynamic theories of the Aharonov Bohm and Sagnac effects are criticised in several ways and the advantages of a novel O(3) invariant electrodynamic theory of the effects are detailed with reference to empirical data.

1 INTRODUCTION

It is well known that the Aharonov Bohm effect[1]-[3] is described conventionally in terms of a holonomy consisting of parallel transport around a closed loop assuming values in the Abelian Lie group U(1)[4], conventionally ascribed to electromagnetism[5]. In this paper the U(1) invariant theory of the Aharonov Bohm effect is criticised in several ways with reference to the well known test of the effect first verified empirically by Chambers[6] by placing a solenoid between the apertures of a Young interferometer. The criticisms of the U(1) invariant theory are given in section two. In section three a holonomy consisting of parallel transport with O(3) covariant derivatives is applied to the Aharonov Bohm effect, and it is shown that the self-inconsistencies detailed in section two are not present. In section four a similar analysis is carried out for the Sagnac effect, and the close similarities between the O(3) Aharonov Bohm and Sagnac effects revealed. Finally a discussion section suggests that all physical optical and interferometric effects are O(3) invariant.

2 CRITICISMS OF THE U(1) INVARIANT THEORY OF THE AHARONOV BOHM EFFECT

It is well known that the change in phase difference of two electron beams caused by the solenoid in the Aharonov Bohm effect is described in the conventional U(1) invariant theory by:

$$\Delta\delta = \frac{e}{\hbar} \int \nabla \times \mathbf{A} \cdot d\mathbf{S} = \frac{e}{\hbar} \int \mathbf{B} \cdot d\mathbf{S} \quad (1)$$

where the magnetic flux density \mathbf{B} of the solenoid is related to the vector potential \mathbf{A} by:

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (2)$$

Outside the solenoid however:

$$\mathbf{B} = \nabla \times \mathbf{A} = \mathbf{0} \quad (3)$$

which is self inconsistent with eqn.(1). In a U(1) invariant theory[5], an attempt is made to remedy this self-inconsistency by using the fact that \mathbf{A} is not zero outside the solenoid, and so can be represented by a function of the type:

$$\mathbf{A} = \nabla\chi \quad (4)$$

The Aharonov Bohm effect is then described by[5]:

$$\Delta\delta = \frac{e}{\hbar} \oint \nabla\chi \cdot d\mathbf{r} = \frac{e}{\hbar} [\chi]_0^{2\pi} = \frac{e}{\hbar} \int \mathbf{B} \cdot d\mathbf{S} \quad (5)$$

using the assertion that χ is not single valued. The analytical form of χ is:

$$\chi = \frac{BR^2}{2} \phi \quad (6)$$

where B is the magnitude of the flux density \mathbf{B} inside the solenoid, R is the radius of the solenoid, and ϕ is an angle, the ϕ component in cylindrical polar coordinates.

However, the interpretation[5] is self-inconsistent in several ways:

1. Outside the solenoid, $B = 0$, so $\chi = 0$, and there is no effect from eqn. (5), contradicting the empirical data[6].
2. For any function χ , a basic theorem of vector analysis states that:

$$\nabla \times (\nabla\chi) := \mathbf{0} \quad (7)$$

This theorem is also true for a periodic function, so outside the solenoid:

$$\mathbf{B} = \nabla \times \mathbf{A} = \mathbf{0} \quad (8)$$

for all χ , and from eqn. (1) the Aharonov Bohm effect again disappears. For example if we take the angle:

$$\chi = \sin^{-1} \frac{x}{a} \quad (|x| < a) \quad (9)$$

then:

$$\nabla\chi = (a^2 - x^2)^{-\frac{1}{2}} \mathbf{i} \quad (10)$$

and

$$\nabla \times (\nabla\chi) := \mathbf{0} \quad (11)$$

or if we take the periodic function:

$$\chi = \cos x; \quad \nabla\chi = -\sin x \mathbf{i} \quad (12)$$

then:

$$\nabla \times (\nabla\chi) := \mathbf{0} \quad (13)$$

Another criticism of eqn.(5) is that the empirical result is obtained only if $\chi \rightarrow \chi + 2\pi$, whereas in general $\chi \rightarrow \chi + 2n\pi$ for a periodic function. So the value of n has to be artificially restricted to $n = 1$ to obtain the correct analytical and empirical result.

The basic problem in a U(1) invariant theorem is that the field \mathbf{B} is zero outside the solenoid, so outside the solenoid $\nabla \times \mathbf{A}$ is zero whereas \mathbf{A} is not zero[5]. At the same time the U(1) Stokes Theorem states that:

$$\int \nabla \times \mathbf{A} \cdot d\mathbf{S} = \oint \mathbf{A} \cdot d\mathbf{r} \quad (14)$$

so that the holonomy $\oint \mathbf{A} \cdot d\mathbf{r}$ is zero and the effect again disappears for all \mathbf{A} outside the solenoid because the left hand side in eqn.(14) is zero.

3 O(3) INVARIANT EXPLANATION OF THE AHARONOV BOHM EFFECT

Recently an O(3) invariant theory of electrodynamics has been developed and shown[7]-[21] to succeed self-consistently in describing all physical optical and interferometric effects using a holonomy consisting of parallel transport with O(3) covariant derivatives as the phase factor of electrodynamics. In this theory, physical effects are rotations in the internal O(3) symmetry gauge space of the theory. The O(3) symmetry gauge space is the physical space of three dimensions, represented in a complex basis ((1), (2), (3))[20, 21], so a rotation in the internal gauge space is a physical rotation, and causes a gauge transformation. Such a gauge transformation has been shown[9, 10] to give a precise explanation of the extra phase shift observed in the Sagnac effect when the platform is rotated[22]. The O(3) invariant theory also gives the first self-consistent explanation of the Sagnac effect when the platform is stationary, and indeed gives the first self-consistent explanation of all interferometric and physical optical effects[20, 21]. The self-inconsistencies inherent in a U(1) covariant theory of these effects have also been shown in considerable detail[7]-[21].

The core of the O(3) invariant explanation of the Aharonov Bohm effect is that the Jacobi identity of covariant derivatives[5]:

$$\sum_{\sigma, \mu, \nu} [D_\sigma, [D_\mu, D_\nu]] := 0 \quad (15)$$

is identical for all gauge group symmetries with the non-Abelian Stokes Theorem:

$$\oint D_\mu dx^\mu + \frac{1}{2} \int [D_\mu, D_\nu] d\sigma^{\mu\nu} := 0 \quad (16)$$

The latter is therefore also an identity for all gauge group symmetries. The equivalence of the two identities (15) and (16) can be demonstrated most easily for the U(1) gauge group, where eqn.(15) becomes the homogeneous field equation:

$$\partial_\mu \tilde{F}^{\mu\nu} = \partial^\sigma F^{\mu\nu} + \partial^\mu F^{\nu\sigma} + \partial^\nu F^{\sigma\mu} := 0 \quad (17)$$

and eqn. (16) becomes the Stokes Theorem:

$$\oint A_\mu dx^\mu = -\frac{1}{2} \int F_{\mu\nu} d\sigma^{\mu\nu} \quad (18)$$

here $\tilde{F}^{\mu\nu}$ is the dual of the U(1) field tensor $F_{\sigma\rho}$, defined by:

$$F_{\sigma\rho} := \partial_\sigma A_\rho - \partial_\rho A_\sigma \quad (19)$$

where A_μ is the four potential.

The proof of the equivalence of eqns. (17) and (18) can be obtained by integrating eqn. (17) as follows:

$$\begin{aligned} -\frac{1}{2} \left(\int \partial^\sigma F^{\mu\nu} d\sigma_{\mu\nu} + \int \partial^\mu F^{\nu\sigma} d\sigma_{\nu\sigma} + \int \partial^\nu F^{\sigma\mu} d\sigma_{\sigma\mu} \right) &:= 0 \\ -\frac{1}{2} \left(\partial^\sigma \int F^{\mu\nu} d\sigma_{\mu\nu} + \partial^\mu \int F^{\nu\sigma} d\sigma_{\nu\sigma} + \partial^\nu \int F^{\sigma\mu} d\sigma_{\sigma\mu} \right) &:= 0 \\ \partial^\sigma \oint A_\mu dx^\mu + \partial^\mu \oint A_\nu dx^\nu + \partial^\nu \oint A_\sigma dx^\sigma &:= 0 \end{aligned} \quad (20)$$

However, it is known that

$$\begin{aligned} -\frac{1}{2} \int F^{\mu\nu} d\sigma_{\mu\nu} &= \oint A_\mu dx^\mu \\ -\frac{1}{2} \int F^{\nu\sigma} d\sigma_{\nu\sigma} &= \oint A_\nu dx^\nu \\ -\frac{1}{2} \int F^{\sigma\mu} d\sigma_{\sigma\mu} &= \oint A_\sigma dx^\sigma \end{aligned} \quad (21)$$

individually, so eqn.(17) is equivalent to any of eqns.(21). The U(1) covariant derivative[5] is:

$$D_\mu = \partial_\mu + igA_\mu \quad (22)$$

so the equivalence of eqn.(17) and eqn.(18) can be writtn more generally as the equivalence of eqns. (15) and (16). More generally, any covariant derivative can be written as[5]:

$$D_\mu = \partial_\mu - igA_\mu \quad (23)$$

so the equivalence of eqn.(15) and (16) is true for all gauge group symmetries. In eqn. (23) it is understood that

$$A_\mu = M^a A_\mu^a \quad (24)$$

where M^a is the group rotation generator. For the U(1) group we write[5] $M = -1$.

In a U(1) invariant theory of electrodynamics therefore eqn.(18) is the integral form of eqn.(17). More generally, eqn.(16) is the integral form of eqn.(15) for all gauge group symmetries. The space part of eqn.(18) is:

$$\oint \mathbf{A} \cdot d\mathbf{r} = \int \mathbf{B} \cdot d\mathbf{S} \quad (25)$$

and using the Stokes Theorem we obtain an identity as required:

$$\int \nabla \times \mathbf{A} \cdot d\mathbf{S} = \int \mathbf{B} \cdot d\mathbf{S} := \int \mathbf{B} \cdot d\mathbf{S} \quad (26)$$

Similarly it can be shown that the integral form of the U(1) Faraday Law:

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \quad (27)$$

is:

$$\oint \mathbf{E} \cdot d\mathbf{r} + \int \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{A}\mathbf{r} = 0 \quad (28)$$

In the novel O(3) invariant electrodynamics[7]-[21], which is much more accurate and self-consistent than the U(1) invariant form, the following three identities exist:

$$\oint \mathbf{A}^{(i)} \cdot d\mathbf{r} := \int \mathbf{B}^{(i)} \cdot d\mathbf{S} \quad (i = 1, 2, 3) \quad (29)$$

i.e. one for each of the three internal indices (1), (2) and (3). The quantities in the equations (29) are linked by the vacuum definition[7]-[21]:

$$\mathbf{B}^{(3)*} := -ig\mathbf{A}^{(1)} \times \mathbf{A}^{(2)} \quad (30)$$

where g is a topological charge defined by:

$$g = \frac{\kappa}{A^{(0)}} \quad (31)$$

and where κ is the wave-number and $A^{(0)}$ the magnitude of $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$ or $\mathbf{A}^{(3)}$. The vector potential $\mathbf{A}^{(3)}$ and the longitudinal magnetic flux density $\mathbf{B}^{(3)}$ are both phaseless[7]-[21], so eqn.(29) with $i = 3$ is the O(3) invariant equation needed for a description of the Aharonov Bohm effect, i.e.

$$\oint \mathbf{A}^{(3)} \cdot d\mathbf{r} := \int \mathbf{B}^{(3)} \cdot d\mathbf{S} \quad (32)$$

The Aharonov Bohm effect is therefore understood in O(3) invariant electrodynamics as a gauge transformation in a vacuum whose configuration space is O(3). The effect is a gauge transformation of eqn.(32) into the region outside the solenoid, because the left and right hand sides of eqn.(32)

exist only inside the solenoid. In general gauge field theory, a gauge transformation of the potential and of the the field are defined through the rotation operator:

$$S = \exp(iM^a\Lambda^a(x^\mu)) \quad (33)$$

where M^a are the group rotation generators and where Λ^a are angles which depend on the four vector x^μ . Under a general gauge field transformation:

$$A'_\mu = SA_\mu S^{-1} - \frac{i}{g}(\partial_\mu S)S^{-1} \quad (34)$$

$$G'_{\mu\nu} = SG_{\mu\nu}S^{-1} \quad (35)$$

In the O(3) invariant expression (32), the vector potential transforms according to:

$$\mathbf{A}^{(3)} \rightarrow \mathbf{A}^{(3)} + \frac{1}{g} \frac{\partial\alpha}{\partial Z} \mathbf{e}^{(3)} \quad (36)$$

and the magnetic field transforms as:

$$\mathbf{B}^{(3)} \rightarrow \mathbf{B}^{(3)} \quad (37)$$

At the point of contact with the electrons therefore, in the region outside the solenoid, the Aharonov Bohm effect is caused by:

$$\frac{1}{g} \oint \frac{\partial\alpha}{\partial Z} \mathbf{e}^{(3)} \cdot d\mathbf{r} = \int \mathbf{B}^{(3)} \cdot d\mathbf{S} \quad (38)$$

i.e. there is a magnetic field present at the point of contact with the electrons and the left hand side of eqn. (38) is physically significant. The reason for this is that the O(3) symmetry internal space of the theory[7]-[21] is the physical space, i.e. the vacuum with configuration space O(3), a non-simply connected configuration space.

Therefore none of the self-inconsistencies present in the U(1) theory of the Aharonov Bohm effect are present in the O(3) invariant electrodynamics recently developed by several workers[7]-[21]. There is a gauge transformed magnetic field present at the point of contact. Agreement with the empirical result by Chambers is obtained through the equation:

$$\Delta\delta = \frac{e}{\hbar} \int \mathbf{B}^{(3)} \cdot d\mathbf{S} = \frac{e}{\hbar} \frac{1}{g} \oint \frac{\partial\alpha}{\partial Z} \mathbf{e}^{(3)} \cdot d\mathbf{r} \quad (39)$$

and this analysis clearly demonstrates the simplicity with which the novel O(3) electrodynamics[7]-[21] removes the self-inconsistencies of the U(1) description of the Aharonov Bohm effect.

4 O(3) INVARIANT DESCRIPTION OF THE SAGNAC EFFECT

The O(3) invariant description of the Sagnac effect[20, 21] is closely similar to that of the Aharonov Bohm effect developed in Section 3, and relies on the same type of gauge transformation, eqn. (36). In the case of the Sagnac effect, however, the factor g is the topological charge $g = \frac{\kappa}{A^{(0)}}$, whereas in the Aharonov Bohm effect it is the factor $g = \frac{e}{\hbar}$. The U(1) invariant electrodynamics usually identified with Maxwell Heaviside electrodynamics[5] fails to describe the Sagnac effect, whereas the O(3) invariant electrodynamics describes it completely, both with platform at rest and when the platform is rotated[22]. The clockwise loop of the Sagnac interferometer is generated from the anticlockwise loop by motion reversal symmetry, T[22], under which the U(1) phase is invariant. So when the platform is at rest there is no phase shift in U(1) invariant electrodynamics, contrary to observation[22]. If one attempts to describe the Sagnac effect with a holonomy based on parallel transport with U(1) covariant derivatives the phase factor becomes:

$$\exp(ig \oint \mathbf{A}^{(i)} \cdot d\mathbf{r}) = \exp(ig \int \mathbf{B}^{(i)} \cdot d\mathbf{S}) \quad (i = 1, 2) \quad (40)$$

and there is no effect because $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are always perpendicular to the path \mathbf{r} [22]. Parallel transport with O(3) covariant derivatives is necessary to construct the holonomy used to describe the Sagnac effect[20]-[22], both when the platform is at rest and when it is rotating.

5 DISCUSSION

In a vacuum whose configuration space is described by the non-simply connected O(3) group it has been shown[20, 21] that eqn.(32) is the general equation for the phase factor in physical optics and interferometry. The equation is:

$$\exp(i \oint \kappa^{(3)} \cdot d\mathbf{r}) = \exp(ig \int \mathbf{B}^{(3)} \cdot d\mathbf{S}) \quad (41)$$

The left hand side can be recognized as a line integral over what is usually termed the dynamical phase. By definition, the line integral changes sign on traversing a closed loop from O to A to A to O, and this fundamental property is responsible for physical optics and interferometry[20, 21]. The $\mathbf{B}^{(3)}$ field appearing on the right hand side changes sign between left and right handed circularly polarized states, and is a superposition of two circularly polarized states. This gives rise to Pancharatnam's phase, which is due to polarization changes and also to the phase caused by cycling of the tip of the vector in a circularly polarized electromagnetic field[23]. Therefore we reach the important conclusion that the $\mathbf{B}^{(3)}$ field is an observable of the phase in all optics and electrodynamics. For example it provides an explanation of the Sagnac effect as argued already.

The U(1) phase factor on the other hand is well known to be

$$\gamma = \exp(i(\omega t - \kappa \cdot \mathbf{r} + \alpha)) \quad (42)$$

where α is an arbitrary number. So the phase factor (42) is defined only up to an arbitrary α , an unphysical result. If $\alpha = 0$ for the sake of argument the phase factor (42) is invariant under motion reversal symmetry (T) and parity inversion symmetry (P)[20, 21]. Since one loop of the Sagnac interferometer is generated from the other by T, it follows that the received phase factor (42) is invariant in the Sagnac effect with platform at rest and there is no phase shift, contrary to observation[22]. The phase factor (41) on the other hand changes sign under T and produces the observed Sagnac effect. The phase factor (42) is invariant under P and cannot explain Michelson interferometry or normal reflection[21]. The phase factor (41) changes sign under P and explains Michelson interferometry as observed[21]. The phase factor (41) also explains Young interferometry[21], which is the Aharonov Bohm effect without the solenoid placed between the apertures of the Young interferometer.

ACKNOWLEDGEMENTS

Funding for individual member laboratories of AIAS is acknowledged together with funding from the U. S. Department of Energy in the form of supercomputer time and the website:

<http://www.ott.doe.gov/electromagnetic/>

which contains material on papers and review articles produced by AIAS.

References

- [1] Y. Aharonov and D. Bohm, Phys. Rev., 115, 484 (1959).
- [2] R. P. Feynman, R. B. Leighton and M. Sands, “The Feynman Lectures on Physics”, vol. 2, section15.5.
- [3] T. T. Wu and C. N. Yang, Phys. Rev. Rev. D, 12, 3845 (1975).
- [4] B. Broda in T. W. Barrett and D. M. Grimes (eds.), “Advanced Electromagnetism” (World Scientific, Singapore, 1995).
- [5] L. H. Ryder, “Quantum Field Theory” (Cambridge, 1987, 2nd. Ed.).
- [6] R. G. Chambers, Phys. Rev. Lett., 5, 3 (1960).
- [7] M. W. Evans, Physica B, 182,227,237 (1992).
- [8] T. W. Barrett in A. Lakhtakia (ed.), “Essays on the Formal Aspects of Electromagnetic Theory” (World Scientific Singapore, 1993).
- [9] B. Lehnert and S. Roy, “Extended Electromagnetic Theory” (World Scientific, Singapore, 1998).
- [10] H. F. Harmuth and M. G. M. Hussain, “Propagation of Electromagnetic Signals” (World Scientific, Singapore, 1994).
- [11] H. F. Harmuth, “Information Theory Applied to Space-Time Physics” (World Scientific, Singapore, 1993)
- [12] M. W. Evans and S. Kielich (eds.), “Modern Nonlinear Optics”, a special topical issue in three parts of I. Prigogine and S. A. Rice (series eds.), “Advances in Chemical Physics” (Wiley, New York, 1992, 1993, 1997 (softback)), vol. 85.
- [13] M. W. Evans and A. A. Hasanein, “The Photomagnetron in Quantum Field Theory” (World Scientific, Singapore, 1994).
- [14] M. W. Evans, J. P. Vigier, S. Roy and S. Jeffers, “The Enigmatic Photon” (Kluwer, Dordrecht, 1994 to 1999) in five volumes.
- [15] M. W. Evans and L. B. Crowell, “Classical and Quantum Electrodynamics and the $\mathbf{B}^{(3)}$ Field” (World Scientific, Singapore, 2000).
- [16] M. W. Evans et al., AIAS group paper, Found. Phys. Lett., 12, 187, 579 (1999); L. B. Crowell and M. W. Evans, Found. Phys. Lett., 12, 373, 475 (1999); L. B. Crowell et al., AIAS Group paper, Found. Phys. Lett., in press (2000); M. W. Evans et al., AIAS Group paper, Found. Phys. Lett., in press (2000); *ibid.*, Found. Phys., in press (2000).
- [17] M. W. Evans et al., AIAS Group paper, Phys. Scripta, 61, 79, 287 and in press (2000).
- [18] M. W. Evans et al., AIAS Group paper Optik, 111, 53 (2000).
- [19] U. S. Department of Energy Website <http://www.ott.doe.gov/electromagnetic/>; special issue of J. New Energy (2000).
- [20] M. W. Evans (ed.), “Contemporary Optics and Electrodynamics” a special topical issue in three parts of I. Prigogine and S. A. Rice (series eds.), “Advances in Chemical Physics” (Wiley, New York, 2001, in prep.), vol. 114, second edition ref.[12].
- [21] M. W. Evans, S. Jeffers and J. P. Vigier in ref. [20], vol. 114(3).
- [22] M. W. Evans et al., AIAS group paper, Phys. Scripta, 61, 79 (2000).
- [23] T. W. Barrett in ref. [8].