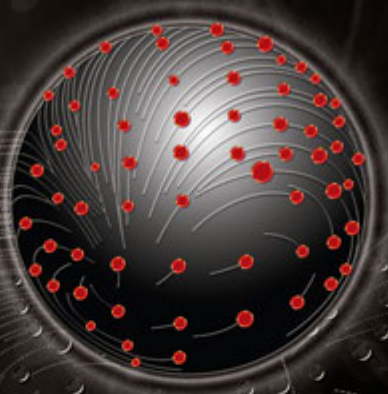




World Scientific Series in Contemporary Chemical Physics – Vol. 26



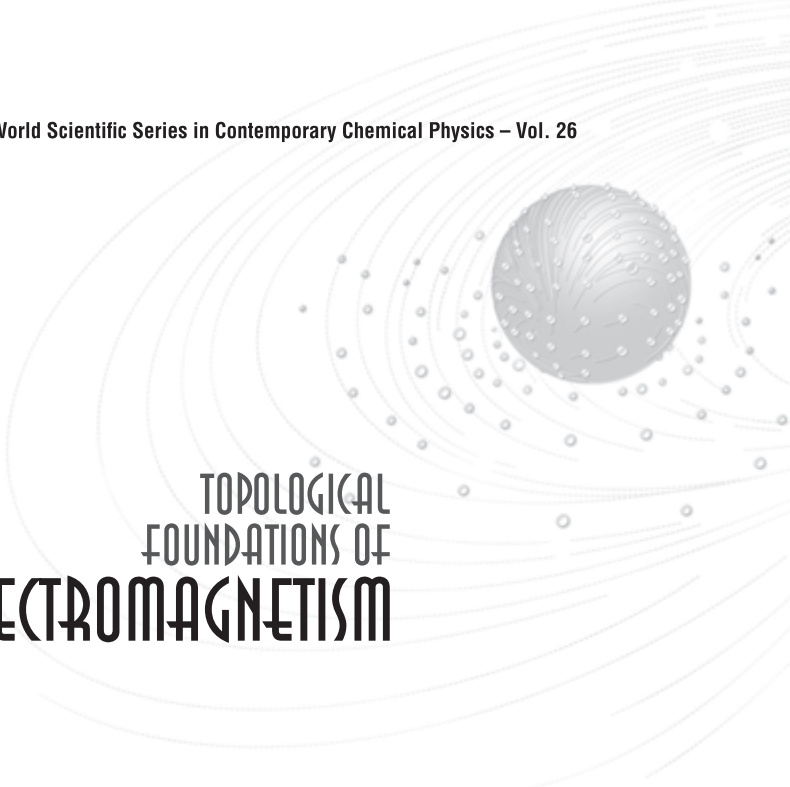
TOPOLOGICAL FOUNDATIONS OF ELECTROMAGNETISM

Terence W Barrett

 World Scientific



World Scientific Series in Contemporary Chemical Physics – Vol. 26



TOPOLOGICAL FOUNDATIONS OF ELECTROMAGNETISM

10.1142/9789810924444



Documents

Document Sharing Hub

World Scientific Series in Contemporary Chemical Physics

Editor-in-Chief: M. W. Evans (*AIAS, Institute of Physics, Budapest, Hungary*)

Associate Editors: S. Jeffers (*York University, Toronto*)
D. Leporini (*University of Pisa, Italy*)
J. Moscicki (*Jagellonian University, Poland*)
L. Pozhar (*The Ukrainian Academy of Sciences*)
S. Roy (*The Indian Statistical Institute*)

- Vol. 9 Water in Biology, Chemistry and Physics: Experimental Overviews and Computational Methodologies
by G. W. Robinson, S. B. Zhu, S. Singh and M. W. Evans
- Vol. 10 The Langevin Equation: With Applications in Physics, Chemistry and Electrical Engineering
by W. T. Coffey, Yu P. Kalmykov and J. T. Waldron
- Vol. 11 Structure and Properties in Organised Polymeric Materials
eds. E. Chiellini, M. Giordano and D. Leporini
- Vol. 12 Proceedings of the Euroconference on Non-Equilibrium Phenomena in Supercooled Fluids, Glasses and Amorphous Materials
eds. M. Giordano, D. Leporini and M. P. Tosi
- Vol. 13 Electronic Structure and Chemical Bonding
by J.-R. Lalanne
- Vol. 14 The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering, 2nd Edition
by W. T. Coffey, Yu. P. Kalmykov and J. T. Waldron
- Vol. 15 Phase in Optics
by V. Perinova, A. Luks and J. Perina
- Vol. 16 Extended Electromagnetic Theory: Space Charge in Vacuo and the Rest Mass of the Photon
by S. Roy and B. Lehnert
- Vol. 17 Optical Spectroscopies of Electronic Absorption
by J.-R. Lalanne, F. Carmona and L. Servant
- Vol. 18 Classical and Quantum Electrodynamics and the B(3) Field
by M. W. Evans and L. B. Crowell
- Vol. 19 Modified Maxwell Equations in Quantum Electrodynamics
by H. F. Harmuth, T. W. Barrett and B. Meffert
- Vol. 20 Towards a Nonlinear Quantum Physics
by J. R. Croca
- Vol. 21 Advanced Electromagnetism and Vacuum Physics
by P. Cornille
- Vol. 22 Energy and Geometry: An Introduction to Deformed Special Relativity
by F. Cardone and R. Mignani
- Vol. 23 Liquid Crystals, Laptops and Life
by M. R. Fisch
- Vol. 24 Dynamics of Particles and the Electromagnetic Field
by S. D. Bosanac
- Vol. 25 Quantum Fluctuations of Spacetime
by L. B. Crowell



World Scientific Series in Contemporary Chemical Physics – Vol. 26



TOPOLOGICAL FOUNDATIONS OF ELECTROMAGNETISM

Terence W Barrett

BSEI, Vienna, Virginia, USA



World Scientific



Documents

Document Sharing Hub

Copyright World Scientific Publishing Co.
Provided by IHS Market under license with WILEY

No reproduction or networking permitted without license from IHS

Licensee: User

Not for Resale,

Full document details: WSPC TOPOL FNDTN ELEC 2008

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

TOPOLOGICAL FOUNDATIONS OF ELECTROMAGNETISM
World Scientific Series in Contemporary Chemical Physics — Vol. 26

Copyright © 2008 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN-13 978-981-277-996-0

ISBN-10 981-277-996-5

Typeset by Stallion Press

Email: enquiries@stallionpress.com

Printed in Singapore.

Preface

Maxwell's equations are foundational to electromagnetic theory. They are the cornerstone of a myriad of technologies and are basic to the understanding of innumerable effects. Yet there are a few effects or phenomena that cannot be explained by the conventional Maxwell theory. This book examines those anomalous effects and shows that they can be interpreted by a Maxwell theory that is subsumed under gauge theory. Moreover, in the case of these few anomalous effects, and when Maxwell's theory finds its place in gauge theory, the conventional Maxwell theory must be extended, or generalized, to a non-Abelian form.

The tried-and-tested conventional Maxwell theory is of Abelian form. It is correctly and appropriately applied to, and explains, the great majority of cases in electromagnetism. What, then, distinguishes these cases from the aforementioned anomalous phenomena? It is the thesis of this book that it is the topology of the spatiotemporal situation that distinguishes the two classes of effects or phenomena, and the topology that is the final arbiter of the correct choice of group algebra — Abelian or non-Abelian — to use in describing an effect.

Therefore, the most basic explanation of electromagnetic phenomena and their physical models lies not in differential calculus or group theory, useful as they are, but in the topological description of the (spatiotemporal) situation. Thus, this book shows that only after the topological description is provided can understanding move to an appropriate and now-justified application of differential calculus or group theory.

Terence W. Barrett

This page intentionally left blank

Contents

<i>Preface</i>	v
Chapter 1: Electromagnetic Phenomena Not Explained by Maxwell's Equations	1
Overview	1
Prolegomena A: Physical Effects Challenging a Maxwell Interpretation	3
Prolegomena B: Interpretation of Maxwell's Original Formulation	6
B.1. The Faraday–Maxwell formulation	6
B.2. The British Maxwellians and the Maxwell–Heaviside formulation	7
B.3. The Hertzian and current classical formulation	9
1. Introduction	12
2. What is a Gauge?	16
3. Empirical Reasons for Questioning the Completeness of Maxwell's Theory	18
3.1. Aharonov–Bohm (AB) and Altshuler–Aronov–Spivak (AAS) effects	18
3.2. Topological phases: Berry, Aharonov–Anandan, Pancharatnam and Chiao–Wu phase rotation effects	27
3.3. Stokes' theorem re-examined	36
3.4. Properties of bulk condensed matter — Ehrenberg and Siday's observation	38
3.5. The Josephson effect	39

3.6. The quantized Hall effect	42
3.7. The de Haas–van Alphen effect	45
3.8. The Sagnac effect	46
3.9. Summary	49
4. Theoretical Reasons for Questioning the Completeness of Maxwell’s Theory	50
5. Pragmatic Reasons for Questioning the Completeness of Maxwell’s Theory	56
5.1 Harmuth’s ansatz	56
5.2 Conditioning the electromagnetic field into altered symmetry: Stokes’ interferometers and Lie algebras	60
5.3 Non-Abelian Maxwell equations	70
6. Discussion	74
References	76

Chapter 2: The Sagnac Effect: A Consequence of Conservation of Action Due to Gauge Field Global Conformal Invariance in a Multiply Joined Topology of Coherent Fields 95

Overview	95
1. Sagnac Effect Phenomenology	96
1.1. The kinematic description	98
1.2. The physical–optical description	101
1.3. The dielectric metaphor description	105
1.4. The gauge field explanation	106
2. The Lorentz Group and the Lorenz Gauge Condition	115
3. The Phase Factor Concept	116
3.1. SU(2) group algebra	118
3.2. A short primer on topological concepts	122
4. Minkowski Space–Time Versus Cartan–Weyl Form	129
5. Discussion	134
References	137

Chapter 3: Topological Approaches to Electromagnetism **141**

Overview 141

1. Solitons 149

2. Instantons 154

3. Polarization Modulation Over a Set Sampling Interval . . 156

4. The Aharonov–Bohm Effect 168

5. Discussion 181

This page intentionally left blank

Downloaded from https://www.cambridge.org/core. University of Cambridge, on 02 Jun 2020 at 10:10:10, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781108753238.001

Chapter

1

Electromagnetic Phenomena Not Explained by Maxwell's Equations²⁶⁰

Overview

The conventional Maxwell theory is a classical linear theory in which the scalar and vector potentials appear to be arbitrary and defined by boundary conditions and choice of gauge. The conventional wisdom in engineering is that potentials have only mathematical, not physical, significance. However, besides the case of quantum theory, in which it is well known that the potentials are physical constructs, there are a number of physical phenomena — *both classical and quantum-mechanical* — which indicate that the A_μ fields, $\mu = 0, 1, 2, 3$, do possess physical significance as global-to-local operators or gauge fields, in precisely constrained topologies.

Maxwell's linear theory is of $U(1)$ symmetry form, with Abelian commutation relations. It can be extended to include physically meaningful A_μ effects by its reformulation in $SU(2)$ and higher symmetry forms. The commutation relations of the conventional classical Maxwell theory are Abelian. When extended to $SU(2)$ or higher

symmetry forms, Maxwell's theory possesses non-Abelian commutation relations, and addresses global, i.e. nonlocal in space, as well as local phenomena with the potentials used as local-to-global operators.

An adapted Yang–Mills interpretation of *low energy fields* is applied in the following pages — an adaptation previously applied only to *high energy fields*. This adaptation is permitted by precise definition of the topological boundary conditions of those low energy electromagnetic fields. The Wu–Yang interpretation of Maxwell's theory implies the existence of magnetic monopoles and magnetic charge. As the classical fields considered here are low energy fields, *these theoretical constructs are pseudoparticle, or instanton, low energy monopoles and charges*, rather than high energy monopoles and magnetic charge (cf. Refs. 1 and 2).

Although the term “classical Maxwell theory” has a conventional meaning, this meaning actually refers to the interpretations of Maxwell's original writings by Heaviside, Fitzgerald, Lodge and Hertz. These *later interpretations of Maxwell* actually depart in a number of significant ways from Maxwell's original intention. In Maxwell's original formulation, *Faraday's electrotonic state, the A field, was central*, making this prior-to-interpretation, original Maxwell formulation compatible with Yang–Mills theory, and naturally extendable.

The *interpreted classical Maxwell theory* is, as stated, a linear theory of U(1) gauge symmetry. The mathematical dynamic entities called solitons can be either classical or quantum-mechanical, *linear or nonlinear* (cf. Refs. 3 and 4), and describe electromagnetic waves propagating through media. However, solitons are of SU(2) symmetry form.²⁵⁹ In order for the conventional *interpreted classical Maxwell theory* of U(1) symmetry to describe such entities, the theory must be extended to SU(2) form.

This recent extension of soliton theory to linear equations of motion, together with the recent demonstration that the nonlinear Schrödinger equation and the Korteweg–de-Vries equation — equations of motion with soliton solutions — are reductions of the self-dual Yang–Mills equation (SDYM),⁵ are pivotal in understanding the extension of Maxwell's U(1) theory to higher order symmetry forms

such as $SU(2)$. Instantons are solutions to SDYM equations which have minimum action. The use of Ward's SDYM twistor correspondence for universal integrable systems means that instantons, twistor forms, magnetic monopole constructs and soliton forms all have a pseudoparticle $SU(2)$ correspondence.

Prolegomena A: Physical Effects Challenging a Maxwell Interpretation

A number of physical effects strongly suggest that the Maxwell field theory of electromagnetism is incomplete. Representing the influence of the independent variable, x , and the dependent variable, y , as $x \rightarrow y$, these effects address: field(s) $\rightarrow$ free electron ($F \rightarrow FE$), field(s) $\rightarrow$ conducting electron ($F \rightarrow CE$), field(s) $\rightarrow$ particle ($F \rightarrow P$), wave guide $\rightarrow$ field ($WG \rightarrow F$), conducting electron $\rightarrow$ field(s) ($CE \rightarrow F$) and rotating frame $\rightarrow$ field(s) ($RF \rightarrow F$) interactions. A nonexhaustive list of these experimentally observed effects, all of which involve the A_μ four-potentials (vector and scalar potentials) in a physically effective role, includes:

1. *The Aharonov–Bohm and Altshuler–Aronov–Spivak effects* ($F \rightarrow FE$ and $F \rightarrow CE$). Ehrenberg and Siday, Aharonov and Bohm, and Altshuler, Aronov and Spivak predicted experimental results by attributing physical effects to the A_μ potentials. Most commentaries in classical field theory still show these potentials as mathematical conveniences without gauge invariance and with no physical significance.
2. *The topological phase effects of Berry, Aharonov, Anandan, Pancharatnam, Chiao and Wu* ($WG \rightarrow F$ and $F \rightarrow P$). In the $WG \rightarrow F$ version, the polarization of light is changed by changing the spatial trajectory adiabatically. The Berry–Aharonov–Anandan phase has also been demonstrated at the quantum as well as the classical level. This phase effect in parameter (momentum) space is the correlate of the Aharonov–Bohm effect in metric (ordinary) space, both involving adiabatic transport.

3. *The Josephson effect* ($CE \rightarrow F$). At both the quantum and the macrophysical level, the free energy of the barrier is defined with respect to an A_μ potential variable (phase).
4. *The quantum Hall effect* ($F \rightarrow CE$). Gauge invariance of the A_μ vector potential, being an exact symmetry, forces the addition of a flux quantum to result in an excitation without dependence on the electron density.
5. *The De Haas–Van Alphen effect* ($F \rightarrow CE$). The periodicity of oscillations in this effect is determined by A_μ potential dependency and gauge invariance.
6. *The Sagnac effect* ($RF \rightarrow F$). Exhibited in the well-known and well-used ring laser gyro, this effect demonstrates that the Maxwell theory, as presently formulated, does not make explicit the constitutive relations of free space, and does not have a built-in Lorentz invariance as its field equations are independent of the metric.

The A_μ potentials have been demonstrated to be physically meaningful constructs at the quantum level (effects 1–5), at the classical level (effects 2, 3 and 6), and at a relatively long range in the case of effect 2. In the $F \rightarrow CE$ and $CE \rightarrow F$ cases (effects 1, 3–5), the effect is limited by the temperature-dependent electron coherence length with respect to the device/antenna length.

The Wu–Yang theory attempted the completion of Maxwell’s theory of electromagnetism by the introduction of a nonintegrable (path-dependent) phase factor (NIP) as a physically meaningful quantity. The introduction of this construct permitted the demonstration of A_μ potential gauge invariance and gave an explanation of the Aharonov–Bohm effect. The NIP is implied by the magnetic monopole and magnetic charge theoretical constructs viewed as *pseudoparticles* or *instantons*.^a

The recently formulated Harmuth ansatz also addresses the incompleteness of Maxwell’s theory: an amended version of Maxwell’s

^aThe term “instanton” or “pseudoparticle” is defined as the minimum action solutions of $SU(2)$ Yang–Mills fields in Euclidean four-space R^4 .³²

equations can be used to calculate e.m. signal velocities provided that (a) a magnetic current density and (b) a magnetic monopole theoretical construct are assumed.

Formerly, treatment of the $\mathbf{A}_\mu$ potentials as anything more than mathematical conveniences was prevented by their obvious lack of gauge invariance.^{251,252} However, gauge invariance for the $\mathbf{A}_\mu$ potentials results from situations in which fields, firstly, have a history of separate spatiotemporal conditioning and, secondly, are mapped in a many-to-one, or global-to-local, fashion (in holonomy). Such conditions are satisfied by $\mathbf{A}_\mu$ potentials with boundary conditions, i.e. the usual empirically encountered situation. Thus, with the correct geometry and topology (i.e. with stated boundary conditions) the $\mathbf{A}_\mu$ potentials always have physical meaning. This indicates that Maxwell's theory can be extended by the appropriate use of topological and gauge-symmetrical concepts.

The $\mathbf{A}_\mu$ potentials are local operators mapping global spatiotemporal conditions onto the local e.m. fields. The effect of this operation is measurable as a phase change, if there is a second, comparative mapping of differentially conditioned fields in a many-to-one (global-to-local) summation. With coherent fields, the possibility of measurement (detection) after the second mapping is maximized.

The conventional Maxwell theory is incomplete due to the neglect of (1) a definition of the $\mathbf{A}_\mu$ potentials as operators on the local intensity fields dependent on gauge, topology, geometry and global boundary conditions; and of (2) a definition of the constitutive relations between medium-independent fields and the topology of the medium.^b Addressing these issues extends the conventional Maxwell theory to cover physical phenomena which cannot be presently explained by that theory.

^bThe paper by Konopinski²⁵³ provides a notable exception to the general lack of appreciation of the central role of the $\mathbf{A}$ potentials in electromagnetism. Konopinski shows that the equations from which the Lorentz potentials $\mathbf{A}_\nu$ ($\mathbf{A}, \phi$) arising from the sources $\mathbf{j}_n$ ($\mathbf{j}, \rho$) are derived are $\partial_\mu^2 \mathbf{A}_\nu = -4\pi \mathbf{j}_\nu / c$, $\partial_\mu \mathbf{A}_\mu = 0$, and can displace the Maxwell equations as the basis of electromagnetic theory. The Maxwell equations follow from these equations whenever the antisymmetric field tensor $F_{\mu\nu}(\mathbf{E}, \mathbf{B}) = \partial_\mu \mathbf{A}_\nu - \partial_\nu \mathbf{A}_\mu$ is defined.

Prolegomena B: Interpretation of Maxwell's Original Formulation

B.1. *The Faraday–Maxwell formulation*

Central to the Maxwell formulation of electromagnetism^{6–16} was the Faraday concept of the *electrotonic state* (from the new Latin *tonicus*, “of tension or tone”; from the Greek *tonos*, “a stretching”). Maxwell defined this state as the “fundamental quantity in the theory of electromagnetism” on the *changes of which* (not on its absolute magnitude) the induction current depends (Ref. 8: vol. 2, p. 540). Faraday had clearly indicated the fundamental role of this state in his two circuit experiments (Ref. 17: vol. 1, series I, 60); and Maxwell endorsed its importance: “The scientific value of Faraday’s conception of an electrotonic state consists in its directing the mind to lay hold of a certain quantity, on the changes of which the actual phenomena depend.” (Ref. 8, vol. 2, p. 541).

The continental European views of the time concerning propagation (e.g. those of Weber, Biot, Savart and Neumann), which Maxwell opposed, were based on the concept of action-at-a-distance. In place of action-at-a-distance, Maxwell offered *a medium characterized by polarization and strain* through which radiation propagated from one local region to another local region. Furthermore, instead of force residing in the medium, Maxwell adopted another Faraday concept: force fields, or *magnetic lines of force independent of matter or magnet*. In contrast, Weber’s position was that force is dependent on relative velocity and acceleration. However, Maxwell’s response was not made in rebuttal of Weber’s. Rather, Maxwell believed that there are two ways of looking at the subject¹⁴: his own and Weber’s. Certainly, Helmholtz achieved a form of synthesis of the two pictures.¹⁸

For Maxwell, the distinction between *quantity* and *intensity* was also central. For example, magnetic intensity was represented by a line integral and referred to the magnetic polarization of the medium. Magnetic quantity was represented by a surface integral and referred to the magnetic induction in the medium. In all cases, a medium was required and *the medium was the seat of electromagnetic phenomena*. The electromagnetic and luminiferous medium were identified.

Faraday's lines of force were said to indicate the direction of minimum pressure at every point in the medium and constituted the field concept. Thus, for Maxwell, the electromagnetic field did not exist, *sui generis*, but as a state of the medium, and the mechanical cause of differences in pressure in the medium was accounted for by the hypothesis of vortices, i.e. polarization vectors. The medium was also restricted to be one in which there was only a displacement current (with no conduction currents) and in which there were no electrical or magnetic sources. Furthermore, rather than electricity *producing* a disturbance in the medium (which was W. Thomson's view), Maxwell's field theory described the presence of electricity as a disturbance, i.e. the electricity *was* the disturbance.

Maxwell conceived of electric current as a moving system with forces communicating the motion from one part of the system to another. The nature of the forces was undefined, and did not need to be defined. As for Maxwell, the forces could be eliminated from the equations of motion by Lagrangian methods for any connected system. That is, *the equations of motion were defined only locally*. Thus, Maxwell dispensed with the dynamic forces permitting propagation through the medium; only the beginning and end of the propagating process was examined and that only locally. Therefore, only the local state of the medium, or the electrotonic intensity or state, was primary, and that corresponds to Maxwell's "F" or " α_o " — or, in modern symbols, the **A** field. This is where matters stood until 1873, with the vector potential playing a pivotal *physical* role in Maxwell's theory.

B.2. *The British Maxwellians and the Maxwell–Heaviside formulation*

Subsequently, the **A** field was *banished from playing the central role in Maxwell's theory and relegated to being a mathematical (but not physical) auxiliary*. This banishment took place during the interpretation of Maxwell's theory by the Maxwellians,¹² i.e. chiefly by Heaviside, Fitzgerald, Lodge and Hertz. The "Maxwell theory" and "Maxwell's equations" we know today are really the interpretation of

Maxwell by these Maxwellians.^{12,16} It was Heaviside who “murdered the $\mathbf{A}$ field” (Heaviside’s description) and whose work influenced the crucial discussion which took place at the 1888 Bath meeting of the British Association (although Heaviside was not present). The “Maxwell’s equations” of today are due to Heaviside’s “redressing” of Maxwell’s work, and should, more accurately, be known as the “Maxwell–Heaviside equations.” Essentially, Heaviside took the twenty equations of Maxwell and reduced them to the four now known as “Maxwell’s equations.”

The British Maxwellians, Heaviside, Fitzgerald and Lodge, may have banished scalar and vector potentials from the propagation equations, but the center of concern for them remained the *dynamic state of the medium, or ether*. The banishment of the potentials can today be justified in the light of the discussion to follow in that the Maxwell theory focused on local phenomena, and the $\mathbf{A}$ field, as we shall see, addresses global connectivity. Therefore, in order for the theory to progress, it was perhaps better that the $\mathbf{A}$ field was put aside, or at least assigned an auxiliary role, *at that time*. Heaviside’s comment that the electrostatic potential was a “physical inanity” was probably correct for the 19th century but, as will be shown below, the potential regained its sanity in the 20th century — starting with the work of Hermann Weyl.

But it should be emphasized that the British Maxwellians retained the focus of theory *on the medium*. Both Heaviside and Poynting agreed that the function of a wire is as a sink into which energy passes from the medium (ether) and is convected into heat. For them, wires conduct electricity with the Poynting vector pointing at right angles to the conducting wire (cf. Ref. 19, Sec. 27-5). The modern conventional view on conduction in wires is similar, but modern theory is not straightforward about where this energy goes, yet still retains Poynting’s theorem. The energy flows, not through a current-carrying wire itself, but through the medium (ether) around it — or, rather, through whatever energy-storing substance a modern theorist imagines exists in the absence of the ether. Nonetheless, Heaviside was probably correct to banish scalar and vector potentials from propagation equations due to the fact that the notion of gauge invariance

(*Mass-stabinvarianz* — see below) was not yet conceived and thus not known to the Maxwellians. However, these $\mathbf{A}$ fields still remained as a repository of energy in the electrotonic state of the medium and the redressed and interpreted Maxwell theory of the British Maxwellians remained a *true dynamic* theory of electromagnetism.

B.3. *The Hertzian and current classical formulation*

But all dynamics were banished by Hertz. Hertz banished even the stresses and strains of the medium (ether) and was vigorously opposed in this by the British Maxwellians.¹² Hertz even went far beyond his mentor, Helmholtz, in his austere operational formulation. Nonetheless, the Hertz orientation finally prevailed, and the modern “Maxwell theory” is today a system of equations describing electrodynamics *which has lost its dynamical basis.*

Another significant reinterpretation of Maxwell took place in which Heaviside was involved. The 19th century battle between Heaviside and Tait concerning the use of quaternions and culminating in the victory of Heaviside and vector analysis may also be reassessed in the light of modern developments. *Without* the concepts of gauge, global (as opposed to local) fields, nonintegrable phase factors (see below) and topological connections, the use of quaternions was getting in the way of progress. That is not to say that either quaternionic algebra or the potentials were, or are, unphysical or unimportant. It is to say, rather, that the potentials could not be understood *then* with the limited theory and mathematical tools available *then*. Certainly it is *now* realized that the algebraic formulation of electromagnetism is more complicated than to be described completely even by quaternionic algebra, and certainly more complicated than to be described by simple vector analysis (cf. Ref. 20).

But to return to Maxwell's original formulation: Maxwell *did* place the $\mathbf{A}$ field center stage and *did* use quaternionic algebra to dress his theory. We know now that quaternionic algebra is described by the $SU(2)$ group of transformations, and vector algebra by the $U(1)$ group of transformations. As such modern propagation phenomena such as solitons are of $SU(2)$ form, we might even view the original Maxwell

formulation as more comprehensive than that offered by the British Maxwellian interpretation, and certainly more of a dynamic theory than the physically unintuitive local theory finally adopted by Hertz. That said, when it is stated, below, that the “Maxwell equations” need extension, it is really the modern Heaviside–Hertz interpretation that is meant. The *original* Maxwell theory could have been easily extended into Yang–Mills form.

Between the time of Hertz’s interpretation of Maxwell’s theory and the appearance of the gauge field concepts of Hermann Weyl, there appeared Whittaker’s notable *mathematical* statements^{21,22} that (i) the force potential can be defined in terms of both standing waves and propagating waves and (ii) any electromagnetic field e.g. dielectric displacement and magnetic force, can be expressed in terms of the derivatives of two scalar potential functions, and also be related to an inverse square law of attraction.

Whittaker²¹ commenced his statement with Laplace’s equation:

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = 0, \quad (\text{P.1})$$

which is satisfied by the potential of any distribution of matter which attracts according to Newton’s law. The potential at any point (x, y, z) of any distribution of matter of mass m , situated at the point (a, b, c) which attracts according to this law, is

$$\int_0^{2\pi} f(z + ix \cos u + iy \sin u, u) du, \quad (\text{P.2})$$

where u is a periodic argument. The most general solution to Laplace’s equation using this expression is

$$V = \int_0^{2\pi} f(z + ix \cos u + iy \sin u, u) du, \quad (\text{P.3})$$

where f is an arbitrary function of the two arguments:

$$z + ix \cos u + iy \sin u \quad \text{and} \quad u. \quad (\text{P.4})$$

In order to express this solution as a series of harmonic terms, Whittaker²¹ showed that it is only necessary to expand the function

~~~~~

$f$  as a *Taylor series* with respect to the first argument  $z + ix \cos u + iy \sin u$ , and as a *Fourier series* with respect to the second argument,  $u$ .

Whittaker also showed that the general solution to the partial differential wave equation,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = \frac{k^2 \partial^2 V}{\partial t^2}, \quad (\text{P.5})$$

is

$$V = \int_0^{2\pi} \int_0^\pi f \left( x \sin u \cos v + y \sin u \sin v + z \cos u + \frac{t}{k}, u, v \right) du dv, \quad (\text{P.6})$$

where  $f$  is now an arbitrary function of the three arguments:

$$x \sin u \cos v + y \sin u \sin v + z \cos u + \frac{t}{k}, u \quad \text{and} \quad v, \quad (\text{P.7})$$

and can be analyzed into a simple plane wave solution. Therefore, for any force varying as the inverse square of the distance, the potential of such a force satisfies *both* Laplace's equation *and* the wave equation, and can be analyzed into simple plane waves propagating with constant velocity. The *sum of these waves*, however, does not vary with time, i.e. *they are standing waves*. Therefore, the force potential can be defined in terms of both standing waves, i.e. by a global, or nonlocal, solution, and by propagating waves, i.e. by a local solution changing in time.

Furthermore, Whittaker<sup>22</sup> demonstrated that any electromagnetic field, e.g. dielectric displacement and magnetic force, can be expressed in terms of the derivatives of *two* scalar potential functions,  $F$  and  $G$ , satisfying

$$\begin{aligned} \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2} - \left[ \frac{1}{c^2} \right] \frac{\partial^2 F}{\partial t^2} &= 0, \\ \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2} + \frac{\partial^2 G}{\partial z^2} - \left[ \frac{1}{c^2} \right] \frac{\partial^2 G}{\partial t^2} &= 0. \end{aligned} \quad (\text{P.8})$$

Thus, Whittaker's mathematical statement related the inverse square law of force to the force potential defined in terms of both standing wave (i.e. global) and propagating wave (i.e. local) solutions. The analysis also showed that the electromagnetic force fields could be defined

in terms of the derivatives of two scalar potentials. This was the state of affairs prior to Weyl's introduction of gauge fields.

The landmark work of Weyl<sup>23–26</sup> and Yang and Mills<sup>27</sup> has been matched by the conception of pseudoparticle or minimum action solutions to the Yang–Mills equations,<sup>1</sup> i.e. instantons. Such phenomena and the appearance of gauge structure are found in simple dynamical, or classical, systems,<sup>28</sup> and the concept of instanton has been the focus of intense activity in recent years (cf. Refs. 2, 29–33).

The demonstration that the nonlinear Schrödinger equation and the Korteweg–de-Vries equation — equations with soliton solutions — are reductions of the self-dual Yang–Mills equations<sup>5</sup> with correspondences to twistor formulations<sup>34</sup> has provided additional evidence concerning the direction that Maxwell's theory must take. These reductions of self-dual Yang–Mills equations are known to apply to various classical systems, depending on the choice of Lie algebra associated with the self-dual fields.<sup>35</sup>

It is also relevant that the soliton mathematical concept need not result only from nonlinear equations. Recently, Barut<sup>3</sup> and Shaaarawi and Besieris<sup>4</sup> have demonstrated that soliton solutions are possible in the case of *linear* de-Broglie-like wave equations. Localized oscillating finite energy solutions to the massless wave equation are derived which move like massive relativistic particles with energy  $E = \lambda\omega$  and momentum  $\mathbf{p} = \lambda\mathbf{k}$  ( $\lambda = \text{const.}$ ). Such soliton solutions to linear wave equations do not spread and have a finite energy field.

## 1. Introduction

There are a number of reasons for questioning the completeness of the conventionally interpreted Maxwell theory of electromagnetism. It is well known that there is an arbitrariness in the definition of the  $\mathbf{A}$  vector and scalar potentials, which, nevertheless, have been found very useful when used in calculations with boundary conditions known.<sup>2,53</sup> The reasons for questioning completeness are due to experimental evidence (Sec. 3), theoretical (Sec. 4) and pragmatic (Sec. 5).

An examination of the Maxwell theory may begin with the well-known Maxwell equations<sup>c</sup>:

Coulomb's law:

$$\nabla \cdot \mathbf{D} = 4\pi\rho; \quad (1.1)$$

Maxwell's generalization of Ampere's Law:

$$\nabla \times \mathbf{H} = \left(\frac{4\pi}{c}\right)\mathbf{J} + \left(\frac{1}{c}\right)\frac{\partial \mathbf{D}}{\partial t}; \quad (1.2)$$

the postulate of an absence of free *local* magnetic poles or the differential form of Gauss' law:

$$\nabla \cdot \mathbf{B} = 0; \quad (1.3)$$

and Faraday's Law:

$$\nabla \times \mathbf{E} + \left(\frac{1}{c}\right)\frac{\partial \mathbf{B}}{\partial t} = 0. \quad (1.4)$$

The constitutive relations of the medium-independent fields to matter are well known to be

$$\mathbf{D} = \varepsilon\mathbf{E}, \quad (1.5)$$

$$\mathbf{J} = \sigma\mathbf{E}, \quad (1.6)$$

$$\mathbf{B} = \mu\mathbf{H}. \quad (1.7)$$

Because of the postulate of an absence of free local magnetic monopoles [Eq. (1.3)], the following relation is permitted:

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (1.8)$$

but the vector potential,  $\mathbf{A}$ , is thus always arbitrarily defined, because the gradient of some scalar function,  $L$ , can be added leaving  $\mathbf{B}$  unchanged, i.e.  $\mathbf{B}$  is unchanged by the gauge transformations:

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla\Lambda; \quad \Phi \rightarrow \Phi' = \Phi - \left(\frac{1}{c}\right)\frac{\partial \Lambda}{\partial t}. \quad (1.9)$$

This arbitrary definition of the potentials means that *any* gauge chosen is arbitrary, or, an appeal must be made to boundary conditions for any choice.

<sup>c</sup>The equations are in Gaussian or cgs units: centimeter, gram and second. The Système International (SI) or mks units are: meter, kilogram and second.

Now Eq. (1.8) permits a redefinition of Eq. (1.4):

$$\nabla \times \left( \mathbf{E} + \left( \frac{1}{c} \right) \frac{\partial \mathbf{A}}{\partial t} \right) = 0 \quad (\text{Faraday's law rewritten}), \quad (1.10)$$

which means that the quantity in brackets is the gradient of a scalar function,  $\Phi$ , and so

$$\mathbf{E} + \left( \frac{1}{c} \right) \frac{\partial \mathbf{A}}{\partial t} = -\nabla \Phi, \quad \text{or} \quad \mathbf{E} = -\nabla \Phi - \left( \frac{1}{c} \right) \frac{\partial \mathbf{A}}{\partial t}, \quad (1.11)$$

and the Maxwell equations (1.3) and (1.4) can be redefined by the use of Eqs. (1.8) and (1.11).

The Maxwell equations (1.1) and (1.2) can also be written as

$$\partial^2 \Phi + \left( \frac{1}{c} \right) \frac{\partial (\nabla \cdot \mathbf{A})}{\partial t} = -4\pi\rho, \quad (1.12)$$

$$\nabla^2 \mathbf{A} - \left( \frac{1}{c^2} \right) \left( \frac{\partial^2 \mathbf{A}}{\partial t^2} \right) - \nabla \left( \nabla \cdot \mathbf{A} + \left( \frac{1}{c} \right) \frac{\partial \Phi}{\partial t} \right) = - \left( \frac{4\pi}{c} \right) \mathbf{J}. \quad (1.13)$$

Since the gauge conditions (1.9) are arbitrary, a set of potentials ( $\mathbf{A}$ ,  $\Phi$ ) can be chosen so that

$$\nabla \cdot \mathbf{A} + \left( \left( \frac{1}{c} \right) \frac{\partial \Phi}{\partial t} \right) = 0. \quad (1.14)$$

This choice is called the Lorentz condition or the Lorentz gauge. Equations (1.12) and (1.13) can then be decoupled to obtain

$$\nabla^2 \Phi + \left( \frac{1}{c^2} \right) \frac{\partial^2 \Phi}{\partial t^2} = -4\pi\rho, \quad (1.15)$$

$$\nabla^2 \mathbf{A} - \left( \frac{1}{c^2} \right) \frac{\partial^2 \mathbf{A}}{\partial t^2} = - \left( \frac{4\pi}{c} \right) \mathbf{J}, \quad (1.16)$$

which is useful, because the Maxwell equations are then independent of the coordinate system chosen. Nonetheless, as  $\mathbf{A}$  and  $\Phi$  are not gauge-invariant, the original choice of the Lorentz gauge is arbitrary — a choice which is not an inevitable consequence of the Maxwell

theory — and the resultant from that choice, namely Eqs. (1.15) and (1.16), is equally arbitrary.<sup>d</sup>

Then again, the arbitrariness of Eq. (1.9) is useful because it permits the choice

$$\nabla \cdot \mathbf{A} = 0. \quad (1.17)$$

Equation (1.12), which is the Maxwell equation (1.1), then permits

$$\nabla^2 \Phi = -4\pi\rho, \quad (1.18)$$

which is the instantaneous Coulomb potential, and hence the condition (1.17) is called the Coulomb or transverse gauge, because the wave equation for  $\mathbf{A}$  can be expressed in terms of the transverse current:

$$\nabla^2 \mathbf{A} - \left( \frac{1}{c^2} \right) \frac{\partial^2 \mathbf{A}}{\partial t^2} = - \left( \frac{4\pi}{c} \right) \mathbf{J}_t, \quad (1.19)$$

where  $\mathbf{J}_t = \mathbf{J} - \mathbf{J}_l$ , with  $\mathbf{J}_l$  being the longitudinal current. This is a useful thing to do when no sources are present, but, again, as  $\mathbf{A}$  and  $\Phi$  are not gauge-invariant, i.e. considered to have no physical meaning, the original choice of the Coulomb gauge is arbitrary, and so is the resultant from that choice, namely Eq. (1.19).

For all that, the absence of gauge invariance (physical meaning) of the  $\mathbf{A}$  vector potential and the  $\Phi$  scalar potential may seem a fortunate circumstance to those using the Maxwell theory to calculate predictions. These potentials have long been considered a fortunate mathematical convenience, but *just* a mathematical convenience, with no physical meaning. These constructs lack gauge invariance, a defining characteristic of physical, rather than merely mathematical, constructs. What then is meant by a gauge and gauge invariance?

<sup>d</sup>The above account applies to the Hertzian potential and the Hertz vector which are related to  $\mathbf{A}$  and  $\phi$ . However, the Hertzian vector obeys an inhomogeneous wave equation with the polarization vector as source, whereas  $\mathbf{A}$  and  $\phi$  obey their respective wave equations with electric current and charge as source. Furthermore, the Hertzian potential is a three-component potential, whereas  $\mathbf{A}$  and  $\phi$  amount to a four-potential (cf. Ref. 252, p. 254).

## 2. What is a Gauge?

In 1918 Weyl<sup>23</sup> (see also Ref. 36) treated Einstein's general theory of relativity as if the Lorentz symmetry were an example of *global* symmetry but with only *local* coordinates defineable, i.e. the general theory was considered as a *local* theory. A consequence of Weyl's theory is that the absolute magnitude or norm of a physical vector is not treated as an absolute quantity but depends on its location in space–time. This notion was called scale (*Mass-stab*) or gauge invariance.

This concept can be understood as follows. Consider a vector at position  $x$  with norm given by  $f(x)$ . If the coordinates are transformed, so that the vector is now at  $x + dx$ , the norm is  $f(x + dx)$ . Using the abbreviation  $\partial/\partial^\mu$ ,  $\mu = 0, 1, 2, 3$ , expanding to first order, and using Einstein's summation convention,

$$f(x + dx) = f(x) + \partial_\mu f dx^\mu. \quad (2.1)$$

If a gauge change is introduced by a multiplicative scaling factor,  $S(x)$ , which equals unity at  $x$ , then

$$S(x + dx) = 1 + \partial_\mu S dx^\mu. \quad (2.2)$$

If a vector is to be constant under change of location, then

$$Sf = f + [\partial_\mu S]f dx^\mu + [\partial_\mu f]dx^\mu \quad (2.3)$$

and, on moving, the norm changes by the amount

$$[\partial_\mu + \partial_\mu S]f dx^\mu. \quad (2.4)$$

Weyl identified  $\partial_\mu S$  with the electromagnetic potential  $A_\mu$ .

However, this suggestion was rejected (by Einstein) because the natural scale for matter is the Compton wavelength,  $\lambda$ , and as the wave description of matter is  $\lambda = \hbar/mc\lambda$  ( $\hbar$  is Planck's constant and  $c$  is the speed of light), then if, as is always assumed, the wavelength is determined by the particle's mass,  $m$ , and with  $\hbar$  and  $c$  constant (according to the special theory of relativity),  $\lambda$  cannot depend on position without violating the special theory. When made aware of this reasoning, Weyl abandoned his proposal. So the term “gauge change” originally meant “change in length,” and was withdrawn from consideration for this particular metric connotation shortly after its introduction.

But the term did not die. “Gauge invariance” managed to survive in classical mechanics because, with the potentials arbitrary, Maxwell’s equations for the  $\mathbf{E}$ ,  $\mathbf{B}$ ,  $\mathbf{H}$  and  $\mathbf{D}$  fields have a built-in symmetry and such arbitrary potentials became a useful mathematical device for simplifying many calculations in electrodynamics, as we have seen. Nevertheless, the gauge invariance in electromagnetism for the  $\mathbf{E}$ ,  $\mathbf{B}$ ,  $\mathbf{H}$  and  $\mathbf{D}$  fields was regarded as only an “accidental” symmetry, and the lack of gauge invariance of the electromagnetic vector and scalar potentials was interpreted as an example of the well-known arbitrariness of the concept of the potential in classical mechanics.

But this arbitrariness of the concept of the potential did not, and does not, exist in quantum mechanics. The electromagnetic vector and scalar potentials were viewed in quantum mechanics in yet another way. Upon the development of quantum mechanics, Weyl and others realized that the original gauge theory could be given a new meaning. They realized that the phase of a wave function could be a new *local* variable. Instead of a change of scale or metric, for which it was originally introduced, a gauge transformation was reinterpreted as a change in the phase of the wave function,

$$\Psi \rightarrow \Psi \exp[-ie^\lambda], \quad (2.5)$$

and the gauge transformation for the potential  $\mathbf{A}_\mu$  became

$$\mathbf{A}_\mu \rightarrow \mathbf{A}_\mu - \frac{\partial \lambda}{\partial x_\mu}. \quad (2.6)$$

Equations (2.5) and (2.6) together ensure that the Schrödinger formulation for a single charged particle in an electromagnetic field remains invariant to phase changes because they self-cancel.<sup>e</sup> Thus any change in location, for that *single* charged particle, which produces a change in the phase [Eq. (2.5)] is compensated for by a corresponding change in the potential [Eq. (2.6)]. Therefore Weyl’s original idea, reinterpreted, was accepted, and the potential in quantum mechanics was viewed as a *connection* which relates phases at different locations. Nevertheless, this use and interpretation did not carry

<sup>e</sup>Wignall<sup>255</sup> has shown that no phase change occurs for de Broglie waves under the low velocity limit of the Lorentz transformation. However, the phase of de Broglie waves is not invariant under a change of frame described by a Galilean transformation.

over into classical mechanics and a schizoid attitude has existed to this day regarding the physical meaning of the potentials in classical and quantum mechanics. In classical mechanics the potentials were, up until recently, viewed as having only an arbitrary mathematical, not physical, meaning, as they seemed to lack gauge invariance. In quantum mechanics, however, they *are* viewed as gauge-invariant and *do* possess a physical meaning. It is an aim of this book to show that in classical mechanics the potentials can also be taken to have, under special circumstances, a physical meaning, and possess gauge invariance under certain (topological) circumstances, and gauge covariance under others.

A major impetus to rethink the physical meaning of the potentials in classical mechanics came about from the experiments examined in the next section.

### 3. Empirical Reasons for Questioning the Completeness of Maxwell's Theory

#### 3.1. *Aharonov–Bohm (AB) and Altshuler–Aronov–Spivak (AAS) effects*

Beginning in 1959 Aharonov and Bohm<sup>40</sup> challenged the view that the classical vector potential produces no observable physical effects by proposing two experiments. The one which is most discussed is shown in Fig. 3.1.1. A beam of monoenergetic electrons exits from the source at X and is diffracted into two beams by the two slits in the wall at Y1 and Y2. The two beams produce an interference pattern at Z which is measured. Behind the wall is a solenoid, S, the **B** field of which points out of the paper. The postulate of the absence of a free local magnetic monopole [Maxwell equation (3) above] predicts the magnetic field outside the solenoid to be zero. Before the current is turned on in the solenoid, there should be the usually expected interference patterns seen at Z. Aharonov and Bohm predicted that if the current is turned on and due to the differently directed **A** fields in paths 1 and 2 indicated by the arrows in Fig. 3.1.1, additional phase shifts should be discernible at Z. This prediction was confirmed experimentally<sup>41–48</sup> and the evidence has been extensively reviewed.<sup>49–53</sup> Aharonov and Casher<sup>54</sup> have extended the theoretical treatment of the AB effect to neutral particles with

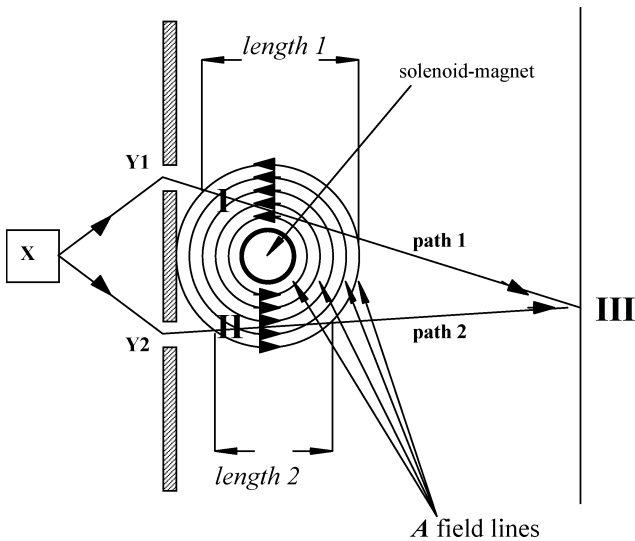


**Fig. 3.1.1.** Two-slit diffraction experiment of the Aharonov–Bohm effect. Electrons are produced by a source at X, diffracted by the slits at Y1 and Y2, and their diffraction pattern is detected at Z. The solenoid is between the slits and directed out of the page. The different orientations of the  $\mathbf{A}$  field at the points of interaction with the two paths are indicated by the arrows  $>$  and  $<$  following the right hand rule.

a magnetic moment; and Botelho and de Mello<sup>55</sup> have analyzed a non-Abelian AB effect in the framework of pseudoclassical mechanics.

One explanation of the effect is as follows. Let  $\psi_0$  be the wave function when there is no current in the solenoid. After the current is turned on, the Hamiltonian is

$$H = \frac{1}{2m} (-i\hbar\nabla\psi - e\mathbf{A})^2, \quad (3.1.1)$$

and the new wave function is

$$\psi = \psi_0 \exp \left[ \frac{-ieS}{\hbar} \right], \quad (3.1.2)$$

where,  $S$ , the flux, is defined as

$$S = \oint \mathbf{A} \cdot d\mathbf{x}, \quad (3.1.3)$$

which is the quantum analog of the classical action evaluated along the paths 1 and 2. At point Z, the wave functions of the two electrons

beams are

$$\begin{aligned}\psi_1 &= \psi_0 \exp \left[ \frac{-eS_1}{\hbar} \right], \\ \psi_2 &= \psi_0 \exp \left[ \frac{-eS_2}{\hbar} \right],\end{aligned}\tag{3.1.4}$$

and the phase difference is

$$\left( \frac{e}{\hbar} \right) (S_1 - S_2) = \left( \frac{e}{\hbar} \right) \left( \int_1 \mathbf{A} \cdot d\mathbf{x} - \int_2 \mathbf{A} \cdot d\mathbf{x} \right) = 2\pi \left( \frac{e}{\hbar} \right) \Phi. \tag{3.1.5}$$

By Stokes' theorem, this is directly proportional to the magnetic flux,  $\Phi = \oint \mathbf{A} \cdot d\mathbf{x}$ , in the solenoid.

However, the phase difference given by Eq. (3.1.5) is not single-valued. Therefore, the value of the phase change will only be determined to within an arbitrary multiple,  $n$ , of  $2\pi e\Phi/\hbar$ , where  $n$  is the number of times the measured charge circulated the solenoid. The topological feature of the background space of the AB effect is its *multiple connectedness*.<sup>56</sup> Therefore, the mathematical object to be computed in this framework is a propagator expressed as a path integral in the covering space of the background physical space (cf. Ref. 57). This means that for a *simply connected* space, all paths between two points are in the same homotopy class, and the effect of the potential,  $A_\mu$ , is as a multiplier of the free-particle propagator with a single gauge phase factor. *In this case, the potential has no physically discernible effect.* However, for a *multiply connected* manifold, the potential *can have a physically discernible effect* because the gauge factors can be different for different homotopy classes.<sup>58,59</sup>

The AB effect was confirmed experimentally in the originally proposed field  $\rightarrow$  free electron (F  $\rightarrow$  FE) situation (cf. Ref. 41). That is, the effect predicted by Aharonov and Bohm refers to the influence of the  $\mathbf{A}$  vector potential on electrons confined to a *multiply connected region*, but within which the magnetic field is zero. As a consequence of the gauge invariance, the energy levels of the electrons have a period  $\hbar/e$  of the enclosed flux. More recent experiments have addressed the appearance of the effect in the field  $\rightarrow$  conduction electron (F  $\rightarrow$  CE) situation (cf. Ref. 96). This situation is also not strictly the same as in

the originally proposed AB experiment in another respect — the magnetic flux is produced by a large solenoid surrounding the influenced condensed matter, usually a loop or a cylinder — so that the  $\mathbf{B}$  field is not set to zero within the material. However, the preponderance of the  $\mathbf{B}$  field is always in the hole encompassed by that cylinder or ring, and the magnetic field causes only secondary effects in the material. In this situation, periodic oscillations in the conductance of the ring appear as a function of the applied magnetic field,  $\mathbf{B}$ . The periodicity of the oscillations is

$$\nabla \mathbf{B} = \frac{\hbar}{eA}, \quad (3.1.6)$$

where  $A$  is the area enclosed by the ring.

Under these conditions the AB effect is seen in normal metal,<sup>45–47,60–67</sup> bulk Mg,<sup>68</sup> semiconductors,<sup>69,70</sup> and on doubly connected geometries on GaAs/AlGaAs heterostructures.<sup>71</sup> The effect has been seen in structures such as cylindrical Mg films<sup>72,73</sup> and Li films,<sup>74</sup> wire arrays,<sup>75,76</sup> arrays of Ag loops,<sup>77</sup> small metal loops<sup>60,65</sup> and MBE-grown double quantum wells.<sup>66</sup>

Bandyopadhyay *et al.*<sup>78</sup> and Datta and Bandyopadhyay<sup>70</sup> have also discussed a novel concept for a transistor based on the electrostatic AB effect in MBE-grown quantum wells, where the current is modulated by quantum interference of electrons in two contiguous channels of a gate voltage. They predict that transistors based on this effect will have power-delay products that are orders of magnitude better than those of existing devices such as MODFETs and Josephson junctions. The transconductance will also be much higher than that of MODFETs. Unlike previous experimental treatments which assumed diffusive transport with negligible inelastic scattering, Datta and Bandyopadhyay<sup>70</sup> assume ballistic transport and perfect symmetry in the arms of the interferometer and in the voltage along the interferometer or two-channel structure.

Now, the AB ( $F \rightarrow CE$ ) effect is temperature-dependent as coherent transport is required. The effect has only been seen at very low temperatures. Measurements were made on parallel GaAs quantum wells at 4.2 K and below,<sup>66</sup> on 860 nm-i.d. Au loops at 0.003 K<sup>61</sup> and  $0.05 \text{ K} < T < 0.7 \text{ K}$ ,<sup>60,63</sup> on 75 nm-o.d. Sb loops at  $0.01 < T <$

1 K<sup>79</sup> and at 0.04 K,<sup>64</sup> and on Ag loop arrays at 4.2 K.<sup>77</sup> Measurements on 1.5–2.0-micron-diameter Mg cylinders of length 1 cm were made at 1.12 K.<sup>72</sup> The Thouless scaling parameter,  $V$ , or the sensitivity of energy levels to a change in the phase of the wave functions at the boundaries<sup>80,81</sup> implies that the necessary energy correlation range for small rings is accessible in the temperature range 0.0001–10 K.<sup>82</sup>

What is remarkable is that these experiments on the ( $F \rightarrow CE$ ) AB effect demonstrate that the effect *can* occur in *disordered* electrical conductors if the temperature is low enough. The effect in metals is a small magnetoresistance oscillation superimposed on the ohmic resistance in multiply connected conductors at low temperatures.<sup>72,74,82</sup> *This means that the conducting electrons must possess a high degree of phase coherence (internal correlation) over distances larger than the atomic spacing or the free path length.*<sup>f</sup> It was initially thought that the effects of finite temperature and the scattering from, and collision with, impurities, would cause incoherence and prevent the observation of the AB effect in bulk samples.<sup>83</sup> The metal loops used measure, for example, less than a micron in diameter and less than 0.1 microns in line thickness. Therefore, the electron is thought to be represented by a pair of waves — one traveling around the ring in the clockwise direction, and the other in the opposite direction, but following the time-reversed path of the first wave. Thus, although each wave has been scattered many times, *each wave collides with the same impurities*, i.e. acquires the same phase shifts, resulting in constructive interference at the origin. *The total path length of both waves is twice the circumference of the ring, meeting the requirement that the phase coherence of the electrons be larger than the circumference of the ring*, or, the transport through the metals arms considered as disordered systems is determined by the eigenvalues of a large random matrix.<sup>83</sup>

Thus, the conductance,  $G$ , of a one-dimensional ring in the presence of elastic scattering is<sup>84</sup>

$$G = \left( \frac{2e^2}{h} \right) \left( \frac{t}{1-t} \right), \quad (3.1.7)$$

<sup>f</sup>The author is indebted to an anonymous reviewer of this page for pointing out that the same coherence effect exists for electrons circulating in an antenna in the presence of ions. If this coherence effect did not exist, radiation would be impossible since emitted power would be proportional to  $I$  instead of  $I^2$ .

and an AB flux applied to the ring results in periodic oscillations of  $G$ , *provided that the phase coherence length of the ring is larger than the size of the system*. (The AB oscillations can be suppressed by magnetic fields, vanishing near resistance minima associated with plateaus in the Hall effect.<sup>85</sup>) Dupuis and Montambaux<sup>86</sup> have shown that in the case of the AB effect in metallic rings, the statistics of levels show a transition from the Gaussian orthogonal ensemble (GOE), in which the statistical ensemble shows time reversal, to the Gaussian unitary ensemble (GUE), in which time reversal symmetry is broken.

A related effect is the Altshuler–Aronov–Spivak (AAS) effect.<sup>87</sup> These authors considered an ultrathin normal metal cylindrical shell of moderate length but very small transverse dimensions at low temperature and how the magnetoresistance would depend on the intensity of magnetic flux axially threading the cylinder. They concluded that it would be an oscillating function of the total flux with a period of  $\hbar/2e$ , i.e. the same as the flux of the superconductive state. The analogous “flux quantum” of the AB effect is  $\hbar/e$ <sup>60,61</sup> and differs from the AAS situation, which involves coherent “backscattering.” The AAS effect has been observed in a 1000-Å-thick magnesium layer on a quartz fiber several millimeters long.<sup>72</sup> More recent treatments of the AAS effect<sup>82,88,89</sup> are based on the quantum-mechanical transmission ( $t$ ) coefficients of electrons and, unlike the original AAS treatment, find an  $\hbar/e$  periodic component as well as the  $\hbar/2e$  harmonic. Raising the temperature above a crossover,  $T_c$ , changes the flux periodicity of magnetic resistance oscillations from  $\hbar/e$  to  $\hbar/2e$ , where  $T_c$  is determined by the energy correlation range  $\hbar D/L^2$ , where  $D$  is the elastic diffusion constant,  $L$  is the length of the sample and the quantity  $\hbar D/L^2$  is the Thouless scaling parameter  $V$  for a metal.

The AAS effect arises because of a special set of trajectories — time-reversed pairs which form a closed loop — which have a fixed relative phase for any material impurity configuration. These trajectories do not average to zero and contribute to the reflection coefficients which oscillate with period  $\hbar/2e$ . The  $\hbar/e$  oscillations of the AB effect, on the other hand, arise from oscillations in the transmission coefficients and can at higher temperatures, average to zero. Below  $T_c$ , both contributions are of order  $e^2/\hbar$ .<sup>82</sup>

Xie and DasSarma<sup>90</sup> studied the AB and AAS effects in the transport regime of a strongly disordered system in which electron transport is via a hopping process — specifically via variable-range-hopping transport. Their numerical results indicate that only the  $\hbar/2e$  (AAS) flux-periodic oscillations survive at finite temperatures in the presence of any finite disorder.

The results of the metal loop experiments demonstrated that elastic scattering does not destroy the phase memory of the electron wave functions.<sup>61,91</sup> The flux periodicity in a condensed matter system due to the AB effect would not be surprising in a superconductor. However, the same periodicities in *finite* conductors is remarkable.<sup>90</sup> Numerical simulation of variable-range-hopping conduction<sup>90</sup> only finds AB oscillations ( $\phi_0 = \hbar/e$ ) in hopping conductance when a metal ring is small and at low temperature. At the large ring limit and higher temperature AAS ( $\phi_0 = \hbar/2e$ ) oscillations survive — a finding consistent with the experimental findings of Polyakov *et al.*<sup>92</sup> A suggested reason for the retention of long range phase coherence is that the phase memory is only destroyed exponentially as  $\exp[-L/L_i]$ , where  $L_i$  is a “typical inelastic scattering length” and the destruction depends on the energy changes in the hopping process dependent on long wavelength, low energy acoustic phonons. The search for an explanation for both AB and AAS effects has resulted in consideration of systems as neither precisely quantum-mechanical nor classical, but in between, i.e. “mesoscopic.”

“Mesoscopic” systems have been studied by Stone<sup>93</sup> in which the energy and spacing is only a few orders of magnitude smaller than  $kT$  at low temperatures. The prediction was made<sup>93</sup> that large AB oscillations should be seen in the transport coefficients of such systems. Such systems have a sample length which is much longer than the elastic mean free path, but shorter than the localization length. The magnetic field through a loop connected to leads changes the relative phase of the contribution from each arm of the loop by  $2\pi\Phi/\Phi_0$ , where  $\Phi_0 = \hbar c/e$  is the one electron flux quantum and  $\Phi$  is the flux through the hole in the loop — but only if the phase-dependent terms do not average to zero. In the mesoscopic range, if inelastic scattering is absent, these phase-dependent contributions do not self-average to zero.

Washburn *et al.*<sup>63</sup> and Stone and Imry<sup>82</sup> demonstrated experimentally that the amplitude of aperiodic and periodic conductance fluctuations decrease for the F  $\rightarrow$  CE AB effect with increasing temperature. There is a characteristic correlation energy:

$$E_C = \frac{\pi \hbar D}{2L^2}, \quad (3.1.8)$$

where  $D$  is the diffusion constant of the electrons, and  $L$  is the minimum length of the sample. If thermal energy  $k_B T > E_C$ , the conductance fluctuations decrease as  $(E_C/k_B T)^{1/2}$ . The conductance fluctuations also decrease when  $L_\phi$ , the phase coherence length, is shorter than the length,  $L$ , or the distance between voltage probes, the decrease being described by a factor  $\exp(-L/L_\phi)$ .<sup>79</sup> This gives a conductance fluctuation:

$$\Delta G n = \left( \frac{L_\phi}{L} \right)^{3/2}. \quad (3.1.9)$$

In condensed matter, therefore, the AB effect appears as the modulation of the electron wave functions by the  $A_\mu$  potential. The phase of the wave function can also be changed by the application of an electric field.<sup>64</sup> The electric field contributes to the fourth term in the four-vector product  $A_\mu(dx)^\mu$ , which contains the scalar potential  $\Phi$  associated with transverse electric fields and time. The phase shift in the wave function is

$$\Delta\phi = \int \frac{e\Phi dt}{\hbar}. \quad (3.1.10)$$

Experiments on Sb metal loop devices on silicon substrate have demonstrated that the voltage on capacitive probes can be used to tune the position (phase) of  $\hbar/e$  oscillations in the loop. Thus, there appear to be two ways to modulate the phase of electrons in condensed matter: application of the  $A_\mu$  potential by threading magnetic flux between two paths of electrons; and application of a scalar potential by means of a transverse electric field. AB fluctuations in metal loops are also not symmetric about  $H = 0$ : four-probe measurements yield resistances which depend on the lead configurations.<sup>62</sup>

More recently, it has been shown that, owing to transport via edge states and penetration of a strong magnetic field into the conducting

region, periodic magnetoconductance oscillations can occur in a singly connected geometry, for example as in a point contact or a “quantum dot” (a disk-shaped region in a two-dimensional electron gas).<sup>94,95</sup> As this effect is dependent on transport via edge states circulating along the boundary of a quantum dot and enclosing a well-defined amount of flux, the geometry is effectively doubly connected. The claim of singly-connectedness is thus more apparent than real. However, there is a difference between the AB effect in a ring and in a dot: in each period  $\nabla \mathbf{B}$ , the number of states below a given energy, stays constant in a ring, but increases by one in a quantum dot. In the case of a dot, the AB magnetoconductance oscillations are accompanied by an increase in the charge of the dot by one elementary charge per period. This can result in an increase in Coulomb repulsion which can block the AB magnetoconductance oscillations. This effect, occurring in quantum dots, has been called the Coulomb blockade of the AB effect.<sup>94,95</sup>

Finally, Boulware and Deser<sup>97</sup> explain the AB effect in terms of a vector potential coupling minimally to matter, i.e. a vector potential not considered as a gauge field. They provide an experimental bound on the range of such a potential as  $10^2$  km.

In summary, the AB and AAS effects, whether  $F \rightarrow FE$  or  $F \rightarrow CE$ , demonstrate that the phase of a *composite* particle’s wave function is a physical degree of freedom which is dependent on differences in  $A_\mu$  potential influences on the space–time position or path of a *first* particle’s wave function with respect to that of another, *second*, particle’s wave function. But the *connection*, or *mapping*, between spatiotemporally different fields or particles which originated at, or passed through, spatiotemporally separated points or paths with differential  $A_\mu$  potential influences, is only measurable by many-to-one mapping of those different fields or particles. By interpreting the phase of a wave function as a *local* variable instead of the norm of a vector, electromagnetism can be interpreted as a *local* gauge (phase) theory, if not exactly, then very close to the way which Weyl originally envisioned it to be.

Below, the interaction of the  $A_\mu$  field ( $x$ ), whether vector or potential, as an independent variable with dependent variable constructs will be referred to in an  $x \rightarrow y$  notation. For example, field  $\rightarrow$  free electron,

field  $\rightarrow$  conducting electron, field  $\rightarrow$  wave guide, field  $\rightarrow$  neutral particle and field  $\rightarrow$  rotating frame interactions will be referred to as  $(F \rightarrow FE)$ ,  $(F \rightarrow CE)$ ,  $(F \rightarrow WG)$ ,  $(F \rightarrow P)$  and  $(F \rightarrow RF)$  interactions. Although the  $\mathbf{A}_\mu$  field is a classical field, the AB and AAS effects are either  $F \rightarrow FE$  or  $F \rightarrow CE$  effects and might be considered “special” in that they involve quantum-mechanical particles, i.e. electrons. In the next section, however, we examine a phase rotation which can only be considered classical, as *both* independent and dependent variables are classical. Nonetheless, the same result — the  $\mathbf{A}_\mu$  potentials demonstrate physical effects — applies.

### 3.2. Topological phases: Berry, Aharonov–Anandan, Pancharatnam and Chiao–Wu phase rotation effects

When addressing the AB effect, Wu and Yang<sup>98</sup> argued that the wave function of a system will be multiplied by a nonintegrable (path-dependent) phase factor after its transport around a closed curve in the presence of an  $\mathbf{A}_\mu$  potential *in ordinary space*. In the case of the Berry–Aharonov–Anandan–Pancharatnam (BAAP) phase, another nonintegrable phase factor arises from the adiabatic transport of a system around a closed path in *parameter (momentum) space*, i.e. this topological phase is the AB effect in parameter space.<sup>99–108</sup> The  $WG \rightarrow F$  version of this effect has been experimentally verified<sup>109</sup> and the phase effect in general interpreted as being due to parallel transport in the presence of a gauge field.<sup>110</sup> The effect exists at both the classical and the quantum level (cf. Refs. 111 and 112).

There has been, however, an evolution of understanding concerning the origins of topological phase effects. Berry<sup>99</sup> originally proposed a geometrical (beside the usual dynamical) phase acquisition for a nondegenerate quantum state which varies adiabatically through a circuit in parameter space. Later, the constraint of an adiabatic approximation was removed<sup>103</sup> and also the constraint of degenerate states.<sup>28</sup> Then Aharonov and Anandan<sup>113</sup> showed that the effect can be defined for any cyclic evolution of a quantum system. Bhandari and Samuel<sup>114</sup> have also pointed out that Berry’s phase is closely connected with a phase discovered by Pancharatnam.<sup>115</sup> These authors

Downloaded from https://www.cambridge.org/core

also demonstrated that unitary time evolution is not essential for the appearance of a phase change in one beam of an interferometer. This is because the polarization state of light can be taken along a closed circuit on the Poincaré sphere, and the resulting polarization changes are not necessarily a function of a unitary time evolution. Thus, the current thinking is that the history of “windings” of a particle is “remembered,” or registered and indicated, by changes in phase either in a quantum-mechanical particle’s state or in a classical wave’s polarization. The topological phase effect appears to arise from the nontrivial topology of the complex projective Hilbert space — whether classical or quantum-mechanical<sup>116</sup> — and to be equivalent to a gauge potential in the parameter space of the system — again, whether classical or quantum-mechanical.

Jiao *et al.*<sup>117</sup> have also indicated at least three variations of topological phases: (i) the phase which arises from a cycling in the direction of a beam of light so that the tip of the spin vector of a photon in this beam traces out a closed curve on the sphere of spin directions — which is that originally studied by Chiao and Wu<sup>106</sup>; (ii) Pancharatnam’s phase arising from a cycling in the polarization states of the light while keeping the direction of the beam of light fixed, so that Stokes’ vector traces out a closed curve on the Poincaré sphere, i.e. the phase change is due to a polarization change; (iii) the phase change due to a cycle of changes in squeezed states of light.

Topological phase change effects, in the field  $\rightarrow$  photon version, have been observed in NMR interferometry experiments<sup>118,119</sup> and using ultracold neutrons<sup>120</sup>; in coherent states<sup>121,122</sup>; optical resonance<sup>123</sup>; and the degenerate parametric amplifier.<sup>124</sup> However, topological phase change effects are more commonly studied in a classical wave guide  $\rightarrow$  classical field (WG  $\rightarrow$  F) version, in which the parameter space is the momentum  $\mathbf{k}$  space.<sup>106,117,125</sup>

For example, the helicity or polarization state,  $\sigma$ , is<sup>106</sup>

$$\sigma = \mathbf{s} \cdot \mathbf{k}, \quad (3.2.1)$$

where  $\mathbf{s}$  is a spin or helicity operator and  $\mathbf{k}$  is the direction of propagation ( $k_x, k_y, k_z$ ). If  $\tau$  is the optical path length, then  $|\mathbf{k}(\tau)$ ,  $\sigma >$  is the spin or polarization state. Interpreted classically, the constraint of keeping  $\mathbf{k}$  parallel to the axis of a wave guide is due to the linear momentum

being in that direction. This means that a wave guide can act as a polarization rotator. Furthermore, as helicity (polarization),  $\sigma$ , is adiabatically conserved,  $\mathbf{s}$  is also constrained to remain parallel to the local axis of the wave guide. Therefore, the topology of a wave guide, e.g. a helix shape, will constrain  $\mathbf{k}$ , and also  $\mathbf{s}$ , to perform a trajectory  $C$  on the surface of a sphere in the parameter space  $(k_x, k_y, k_z)$  which prescribes the linear momentum. Thus, the topology of the constrained trajectory of radiation progressing between two *local* positions has a *global* effect indicated by a polarization (spin) change. If  $\gamma(C)$  is the topological phase, and  $\beta = \exp[i\gamma(C)]$  is a phase factor, the final polarization state after progression along a constrained trajectory, i.e. “momentum conditioning,” is

$$\sigma_2 = \beta \cdot \mathbf{s} \cdot \mathbf{k} \quad (3.2.2)$$

where the subscript indicates a second location on the trajectory.

As a monopole is theoretically required at  $\mathbf{k} = 0$ , owing to the radial symmetry of the parameter space and resulting singularity, a solid angle  $\Omega(C)$  can be defined on a parameter space sphere with respect to the origin  $\mathbf{k} = 0$ . Thus,  $\Omega(C)$  can be said to define the “excited states” of the monopole at  $\mathbf{k} = 0$ . Therefore

$$\sigma_2 - \sigma_1 = \beta \cdot \mathbf{s} \cdot \mathbf{k} - \sigma_1 = \sigma_1 \Omega(C) - \sigma_1 = \gamma(C). \quad (3.2.3)$$

The following question can be asked: What conservation law underlies the topological phase? A clue is provided by Kitano *et al.*,<sup>126</sup> who point out that the phase change can also be seen in discrete optical systems which contain no wave guides, for example in a configuration of (ideal or infinitely conducting) mirrors. Now, mirrors do not conserve helicity; they reverse it and the local tangent vector,  $\mathbf{t}$ , must be replaced by  $-\mathbf{t}$  on alternate segments of the light path. Mirror configurations of this type have been used in a laser gyro.<sup>127</sup> This suggests that changes of acceleration, whether along a wave guide or in a mirror reflection, under equivalence principle conditions are the compensatory changes which match changes in the topological phase, giving the conservation equation:

$$\gamma(C) + \oint \mathbf{A} \cdot d\mathbf{l} = 0. \quad (3.2.4)$$

That the phase effect change can occur in classical mechanical form is witnessed by changes in polarization rotation resulting from changes in the topological path of a light beam. Tomita and Chiao<sup>109</sup> demonstrated effective optical activity of a helically wound single mode optical fiber in confirmation of Berry's prediction. The angle of rotation of linearly polarized light in the fiber gives a direct measure of the topological phase *at the classical level*. (Hannay<sup>128</sup> has also discussed the classical limit of the topological phase in the case of a symmetric top.) The effect arises from the overall geometry of the path taken by the light and is thus a *global topological effect* independent of the material properties of the fiber. The optical rotation is *independent of geometry* and therefore may be said to quantify the “*topological charge*” of the system, i.e. the helicity of the photon, which is a relativistic quantum number.

Referring to Fig. 3.2.1, the fiber length is

$$s = [p^2 + (2\pi r)^2]^{1/2}, \quad (3.2.5)$$

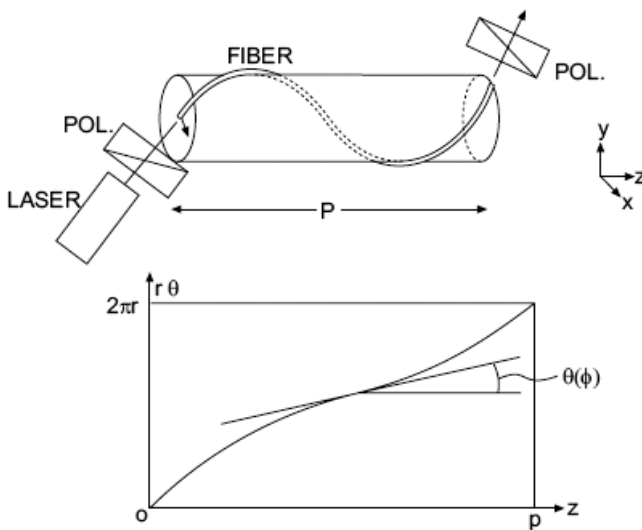


Fig. 3.2.1. (a) Experimental setup; (b) geometry used to calculate the solid angle in momentum space of a nonuniformly wound fiber on a cylinder. (After Ref. 109.)

and the solid angle in momentum space  $\Omega(C)$  spanned by the fiber's closed path  $C$ , a circle in the case considered, is

$$\Omega(C) = 2\pi(1 - \cos \theta). \quad (3.2.6)$$

The topological phase is

$$\gamma(C) = -2\pi\sigma \left(1 - \frac{P}{s}\right), \quad (3.2.7)$$

where  $\sigma = \pm 1$  is the helicity quantum number of the photon.

By wrapping a piece of paper with a computer-generated curve on a cylinder to which the fiber is fitted, and then unwrapping the paper, the local pitch angle, or the tangent to the curve followed by the fiber, can be estimated to be [Fig. 3.2.1(b)]

$$\theta(\phi) = \tan^{-1} \left( \frac{rd\phi}{dz} \right), \quad (3.2.8)$$

which is the angle between the local wave guide and the helix axes. In momentum space,  $\theta(\phi + \pi/2)$  traces out a closed curve  $C$ , the fiber path on the surface of a sphere. The solid angle subtended by  $C$  to the center of the sphere is

$$\Omega(C) = \int_0^{2\pi} [1 - \cos \theta(\phi)] d\phi. \quad (3.2.9)$$

The topological phase is then, more correctly,

$$\begin{aligned} \gamma(C) &= -\sigma \Omega(C) \quad \text{or} \\ \gamma(C) &= \nu \Omega(C) \end{aligned} \quad (3.2.10)$$

(where  $\nu = 1/2$  in the case of polarization charges, i.e. Pancharatnam's phase). There is thus a linear relation between the angle of rotation of linearly polarized light, and the solid angle  $\Omega(C)$  subtended by  $C$  at the origin of the momentum space of the photon.<sup>109</sup>

More recently, Chiao *et al.*<sup>129</sup> have demonstrated a topological phase shift in a Mach–Zehnder interferometer in which light travels along nonplanar paths in two arms. They interpret their results in terms of the Aharonov–Anandan phase and changes in projective Hilbert space, i.e. the sphere of spin directions of the photon, rather

than parameter (momentum) space. The hypothesis tested was that the evolution of the state of a system is cyclic, i.e. that it returns to its starting point adiabatically or not. Thus the  $C$  in Eq. (3.2.10) is to be interpreted as a closed circuit on the sphere of spin directions.

Chiao and Wu<sup>106</sup> consider topological phase rotation effects to be “topological features of the Maxwell theory which originate at the quantum level, but survive the correspondence principle limit ( $\hbar \rightarrow 0$ ) into the classical level.” However, this opinion is contested and the effect is viewed as classical by other authors (e.g. Refs. 102, 103, 130 and 131). For example, the evolution of the polarization vector can be viewed as being determined by a connection on the tangent bundle of the two-dimensional sphere.<sup>130,131</sup> The effect is then viewed as non-Abelian. The situation can then be described with a family of Hamiltonian operators,  $H_0 + \mathbf{k} \cdot \mathbf{V}$ , where  $H_0$  is rotationally invariant,  $\mathbf{V}$  is a vector operator and  $\mathbf{k}$  varies over the unit vectors in  $\mathbf{R}^3$ .<sup>131</sup>

The Chiao–Wu phase and the Pancharatnam phase are additive.<sup>117</sup> This is because the two topological phase effects arise in different parameter spaces: the former in  $\mathbf{k}$  space and the latter in polarization vector (Poincaré sphere) space. To see this, Maxwell’s equations can be recast into a six-component spinor form<sup>117</sup>

$$\nabla \times (\mathbf{E} \pm \mathbf{B}) = \frac{\pm \partial(\mathbf{E} \pm i\mathbf{B})}{\partial t}, \quad (3.2.11)$$

$$\Psi = \text{col}(\mathbf{E} + \mathbf{B}, \mathbf{E} - \mathbf{B}) = \text{col}((\Psi^+), (\Psi^-)), \quad (3.2.12)$$

where  $\text{col}$  denotes a column vector, to obtain a Schrödinger-like equation

$$\frac{i\partial\Psi}{\partial t} = H\Psi, \quad (3.2.13)$$

where the Hamiltonian,  $H$ , is given by

$$H = \begin{pmatrix} (\nabla \times) & 0 \\ 0 & (-\nabla \times) \end{pmatrix} \quad (3.2.14)$$

( $\nabla \times$  represents the curl). This spinor representation of Maxwell’s equations has a natural correspondence with natural optical activity in the frequency domain.<sup>132</sup>

The conventional dynamical phase becomes

$$\delta(H) = - \int_0^T (\Psi, H\Psi) dt \quad (3.2.15)$$

and the geometrical (topological) phase is, as before

$$\gamma(C) = \oint \mathbf{A} \cdot d\mathbf{l}, \quad (3.2.16)$$

but where the vector potential,  $\mathbf{A}$ , is explicitly defined as

$$A = -(\Psi, \nabla \Psi), \quad (3.2.17)$$

i.e. a connection defined on the state space.

By means of Stokes' theorem

$$\gamma(C) = \oint \mathbf{A} \cdot d\mathbf{l} = \int_S \nabla \times \mathbf{A} \cdot d\mathbf{S}, \quad (3.2.18)$$

$$\gamma(C) = \nu \Omega(C), \quad (3.2.19)$$

as before.

Figure 3.2.2 shows the different manifestations and representations of the topological phase effect. Figure (i) is a sphere of *spin directions* for representing the (Chiao–Wu) phase arising from the spin vector of a photon tracing out a closed curve on the sphere.<sup>117</sup> The topological phase is equal to the angle  $\Omega$ . Figure (ii) is a Poincaré sphere of polarization states, or *helicity*, of a photon for representing the (Pancharatnam) phase effect arising from cycling in the polarization states of the photon while keeping the direction of the beam fixed.<sup>114</sup> The topological phase is equal to the *negative* of one half the angle  $\Omega$ . Figure (iii) is a *generalized* Poincaré sphere for representing the angular momentum of light (with space-fixed axis),<sup>117</sup> or a *null flag or twistor* representation<sup>34,133</sup> for representing the Pancharatnam topological phase but with the phase equal to the *positive* value of one half the angle  $\Omega$ . The topological phase effects represented in (i) and (iii) are additive.

A more explicit relation between the topological phase effect and Maxwell's theory is obtained within the formulation of Maxwell's

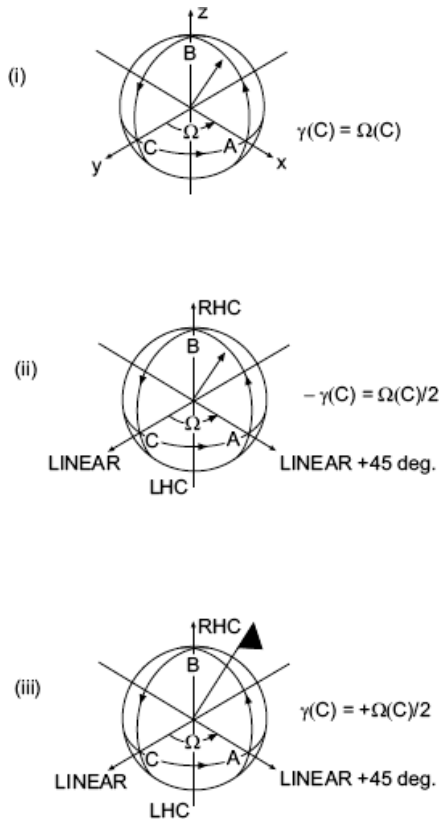


**Fig. 3.2.2.** (i) A sphere of *spin directions* for representing the (Chiao–Wu) phase arising from the spin vector of a photon tracing out a closed curve on the sphere. (After Ref. 117.) The topological phase is equal to the angle  $\Omega$ . (ii) A Poincaré sphere of polarization states, or *helicity*, of a photon for representing the (Pancharatnam) phase effect arising from cycling in the polarization states of the photon while keeping the direction of the beam fixed (After Ref. 114). The topological phase is equal to the *negative* of one half the angle  $\Omega$ . (iii) A *generalized* Poincaré sphere for representing the angular momentum of light (with space-fixed axis) (after Ref. 117), or a *null flag* or *twistor* representation (after Refs. 34 and 133) for representing the Pancharatnam topological phase but with the phase equal to the *positive* value of one half the angle  $\Omega$ . The topological phase effects represented in (i) and (iii) are additive.

theory by Biakynicki-Birula and Bialynicka-Birula.<sup>134</sup> Within this formulation, the intrinsic properties of an electromagnetic wave are its wave vector,  $\mathbf{k}$ , and its polarization,  $\mathbf{e}(\mathbf{k})$ . As Maxwell's theory can be formulated as a representation of Poincaré symmetry, all wave vectors

form a vector space. Thus, implicit in Maxwell's theory is the topology of the surface of a sphere (i.e. the submanifold,  $S^2$ ). The generators of the Poincaré group involve a covariant derivative in momentum space, whose curvature is given by a magnetic monopole field. However, the Maxwell equations can only determine the polarization tensor,  $\mathbf{e}(\mathbf{k})$ , *up to an arbitrary phase factor*, as Maxwell's theory corresponds to the structure group  $U(1)$ , i.e. Maxwell's theory *cannot* determine the phase of polarization in momentum space. This arbitrariness permits additional topological phase effects.

On the other hand, the topological phase is *precisely* obtained from a set of angles associated with a group element and there is just one such angle corresponding to a holonomy transformation of a vector bundle around a closed curve on a sphere.<sup>135</sup> The parameter space is the based manifold and each fiber is isomorphic to an  $N$ -dimensional Hilbert space. In particular, for the  $SU(2)$  case there is a single angle from the holonomy of the Riemannian connection on a sphere.<sup>135</sup> The observation that gauge structure appears in simple dynamical systems — both quantum-mechanical and classical — has been made.<sup>28</sup> For the special case of Fermi systems, the differential geometric background for the occurrence of  $SU(2)$  topological phases is the quaternionic projective space with a time evolution corresponding to the  $SU(2)$  Yang–Mills instanton.<sup>136</sup> *Locally*, the non-Abelian phase generated can be reduced to an Abelian form. However, it is not possible to define the connection defined on the bundle space except *globally*. This reflects the truly non-Abelian nature of the topological phase. The topological phase effect can be described in a *generalized* Bloch sphere model and an  $SU(2)$  Lie group formulation in the spin-coherent state.<sup>137</sup> Furthermore, while acknowledging that in general terms, and formally, Berry's phase is a geometrical object in projective Hilbert space (ray space), the nonadiabatic Berry's phase, physically, is related to the expectation value of spin (spin alignment), and Berry's phase quantization is related to spin alignment quantization.<sup>138</sup>

The topological phase effect even appears in quantum systems constrained by molecular geometry. For example, the topological phase effect appears in the molecular system  $\text{Na}_3$ .<sup>139</sup> Suppose that a system in an eigenstate  $C(r, t)$  responds to slowly varying changes in its

parameters  $R(t)$ , such that the system remains in the *same eigenstate apart from an acquired phase*. If the parameters,  $R(t)$ , completed a circuit in parameter space, then that acquired phase is not simply the familiar dynamical phase,  $[(i\hbar)^{-1}E(R(t))dt]$ , but an additional geometrical phase factor,  $\gamma_n(c)$ . The origins of this additional phase factor depend only on the geometry of the parameter space and the topology of the circuit traversed. Therefore, adiabatic excursions of molecular wave functions in the neighborhood of an electronic degeneracy result in a change of phase. That is, if the internuclear coordinates of a wave function traverse a circuit in which the state is degenerate with another, then the electronic wave function acquires an additional phase, i.e. it changes its sign. This change was predicted<sup>140–142</sup> and is a special case of the topological phase applying to a large class of molecular systems exhibiting conical intersections. Delacrétaz *et al.*<sup>139</sup> reported the evidence for half-odd quantization of free molecular pseudorotation and offered the first experimental confirmation of the sign change theorem and a direct measurement of the phase. The topological phase has also been observed in fast-rotating superfluid nuclei, i.e. oscillations of pair transfer matrix elements as a function of the angular velocity<sup>143</sup> and in neutron spin rotation.<sup>144</sup>

The appearance of the topological phase effect in both classical and quantum-mechanical systems thus gives credence to the view that the  $A_\mu$  potentials register physical effects at both the classical and the quantum-mechanical level. That such a role for these potentials exists at the quantum-mechanical level is not new. *It is new to consider the  $A_\mu$  potentials for such a role at the classical level.* One may ask how the schism in viewing the  $A_\mu$  potentials came about, i.e. why are they viewed as physical constructs in quantum mechanics, but as merely arbitrary mathematical conveniences in classical mechanics? *The answer is that whereas quantum theory is defined with respect to boundary conditions, in the formal presentation of Maxwell's theory boundary conditions are undefined.* Stokes' theorem demonstrates this.

### 3.3. Stokes' theorem re-examined

Stokes' theorem of potential theory applied to classical electromagnetism relates diverging potentials on line elements to

potentials on surface elements. Thus, Stokes' theorem describes a local-to-global field relationship.

If  $\mathbf{A}(x)$  is a vector field,  $S$  is an open, orientable surface,  $C$  is the closed curve bounding  $S$ ,  $d\mathbf{l}$  is a line element of  $C$ ,  $\mathbf{n}$  is the normal to  $S$  and  $C$  is traversed in a right-hand screw sense (positive direction) relative to  $\mathbf{n}$ , then the line integral of  $\mathbf{A}$  is equal to the surface integral over  $S$  of  $(\nabla \times \mathbf{A}) \cdot \mathbf{n}$ :

$$\oint_C \mathbf{A} \cdot d\mathbf{l} = \int_S (\nabla \times \mathbf{A}) \cdot \mathbf{n} da. \quad (3.3.1)$$

It is also necessary that  $S$  be the union of a finite number of smooth surface elements and that the first order partial derivation of the components of  $\mathbf{A}$  be continuous on  $S$ . Thus Stokes' theorem, as described, takes no account of: (i) space-time overlap in a region with fields derived from different sources; (ii) the exact form of the boundary conditions.

This neglect of the exact form of the boundary conditions in Stokes' theorem of classical mechanics can be contrasted with the situation in quantum mechanics. In quantum mechanics, the wave function satisfies a partial differential equation coupled to boundary conditions because the Schrödinger equation describes a minimum path solution to a trajectory between two points. The boundary condition in the doubly connected (overlap) region outside of the shielded volume in an AB experiment is the reason for the single-valuedness of the wave function, and also the reason for quantization. The situation is also different with spatial symmetries other than the usual, Abelian, spatial symmetry.

A non-Abelian Stokes theorem is<sup>145</sup>

$$h^{-1} \left( \frac{dh}{ds} \right) \approx ie \int_0^1 g^{-1} \mathbf{G}_{ij} g \left( \frac{\partial r^i}{\partial t} \right) \left( \frac{\partial r^j}{\partial s} \right) dt, \quad (3.3.2)$$

where  $h(s)$  is a path-dependent phase factor associated with a closed loop and defines a closed loop  $r(s, t)$ ,  $0 \leq t \leq 1$ ,  $s$  fixed, in the  $U(1)$  symmetry space,  $\mathbf{H}$  (equivalent to  $\mathbf{A}_\mu$ );  $\mathbf{G}$  is a gauge field tensor for the  $SU(2)$  non-Abelian group; and  $g$  is magnetic charge. Here, the boundary conditions, i.e. the path dependencies, are made explicit.

and we have a local field [with  $U(1)$  symmetry] to global field [with  $SU(2)$  symmetry] connection.

In classical electromagnetism, therefore, Stokes' theorem appears merely as a useful mathematical relation between a vector field and its curl. *In gauge theory, on the other hand, an amended Stokes' theorem would provide the value for the net comparative phase change in the internal direction of a particle traversing a closed path, i.e. a local-to-global connection.*

Lest it be thought that the  $A_\mu$  field which functions as the independent variable in the AB experiment is only a quantum effect with no relevance to classical behavior, the relation of the  $A_\mu$  potential to the properties of bulk condensed matter is examined in the following section. A more complete definition of Stokes' theorem is also given in Sec. 4 below [Eq. (4.10)].

Use of Stokes' theorem has a price: that of the (covert) adoption of a gauge for local-to-global connections. This is because Stokes' theorem applies directly to propagation issues, which are defined by local-to-global connections. Such connections are also required in propagation through matter. Thus, there is a requirement for Stokes' theorem in any realistic definition of macroscopic properties of matter, and in the next section we see that the physical effects of the  $A_\mu$  potentials exist not merely in fields traversing through various connecting topologies, but in radiation-matter interactions.

### 3.4. *Properties of bulk condensed matter — Ehrenberg and Siday's observation*

In the AB  $F \rightarrow FE$  situation, when the size of the solenoid is much larger than the de Broglie wavelength of the incident electrons, the scattering amplitude is essentially dominated by simple classical trajectories. But the classical manifestation of quantum influences is not peculiar to the AB effect. For example, macroscopic quantum tunneling is observable in Josephson tunnel junctions in which the phase difference of the junction can be regarded as a macroscopic degree of freedom, i.e. a classical variable.<sup>146,147</sup>

Even without known quantum influences or quantum-mechanical explanation, there is a classical justification for the  $A_\mu$  potential as a physical effect. For example, on the basis of optical arguments, the  $A_\mu$  potential must be chosen so as to satisfy Stokes' theorem, *thereby removing the arbitrariness with respect to the gauge*. Furthermore, an argument originating with Ehrenberg and Siday<sup>148</sup> shows that a gauge-invariant  $A_\mu$  potential is *presupposed* in any definition of the refractive index.

This argument is a derivation of the refractive index based on *Fermat's principle*: in any optical medium, a scalar quantity, e.g. the refractive index, finite everywhere in space, can be defined so that the line integral in the three-dimensional space taken between any fixed points must be an extremum which passes through these points. The optical path along a given line connecting a point 1 and a point 2 is

$$\int_1^2 m ds = \int_1^2 [mv + (A_\mu \cdot \mathbf{n})] ds, \quad (3.4.1)$$

where  $\mathbf{n}$  is the unit vector in the direction of the line,  $v$  is the velocity of the electron, and  $m$  is its mass. Defined in this way, an unambiguous definition of the refractive index indicates the *necessity* of a unique (gauge-invariant) definition of the  $A_\mu$  potential. Stated differently: an unambiguous definition of the refractive index implies defining the boundary conditions through which test radiation moves. These boundary conditions define a definite gauge and thereby definite  $A_\mu$  potentials.

This an example of physical  $A_\mu$  -dependent effects (the refractive properties of matter) seen when radiation propagates through matter — from one point to another. In the next section  $A_\mu$  effects are described when two fields are in close proximity. This is the Josephson effect, and again, the potential functions as a global-to-local operator.

### 3.5. The Josephson effect

Josephson<sup>149–152</sup> predicted that a d.c. voltage,  $V$ , across the partitioning barrier of a superconductor gives rise to an alternating current of