

# Time reversal symmetry and the geometric phase

A THESIS  
SUBMITTED IN PARTIAL FULFILMENT  
OF THE REQUIREMENTS FOR THE DEGREE  
OF  
PHD IN PHYSICS  
IN THE  
UNIVERSITY OF CANTERBURY  
by  
Mattias Torbjörn Johnsson



University of Canterbury  
1998

## Acknowledgements

I am very grateful to Professor Geoff Stedman for his patient supervision, and his always having time to discuss my problems and provide new ideas. I am also grateful to Professor Ian Aitchison who collaborated with me on a portion of the work in Chapter 6.

I thank becky and Grace for their friendship and support though my thesis. I am also indebted to Stuart for his help with the often recalcitrant computer system and his editing suggestions. My fellow student Siva is owed thanks for his willingness to have my sometimes peculiar ideas bounced off him and offering thoughtful comment. I would like to give special thanks to Shauna for her love, support and encouragement throughout almost the entire thesis episode. She always knew I could do it. Gratitude also goes to my parents. I am indebted to them for their love and kindness, and their sympathy when things became difficult.

Finally, I would like to offer up an anti-acknowledgement to the the evil minus sign that dogged me through over eight months of my PhD. I got you in the end.

# Abstract

This thesis examines the quantum-mechanical geometric phase with a view toward time reversal symmetry considerations.

The idea of time reversal in quantum mechanics is investigated, disagreements and inconsistencies in the literature are examined, and the action of the time reversal operator is extended to time-dependent Hamiltonians. With this background, and using a definition of time reversal symmetry based on the evolution operator, I demonstrate that the existence of a non-zero geometric phase can in all cases be attributed to a breakdown of time reversal symmetry in some form. This result holds for both adiabatic and nonadiabatic evolutions, and for arbitrary dimensional parameter spaces.

I explore the role of the geometric phase in a two-level Kramers system described by a parameter-dependent Hamiltonian such that the two levels can become degenerate for some value of the parameters, and discuss, from a mathematical point of view, the monopole geometric potential that results. I then extend this analysis by considering a pair of Kramers doublets, each doublet degenerate due to time reversal symmetry, where the parameters can be chosen so that each of the pair of doublets becomes degenerate with the other. I find the explicit functional forms for the two resulting nonabelian geometric gauge potentials and show that they can be identified exactly with the only two gauge-inequivalent  $SU(2)$  monopole potentials of Yang. Furthermore, following a conformal transformation these potentials can be mapped to those of the  $SU(2)$  instanton / anti-instanton pair.

Finally I examine the relevance of the geometric phase to the molecular physics of time-odd systems. Time-odd coupling in molecular physics is a much under-studied area, with many potentially interesting results. Specifically I study time-odd coupling in Jahn-Teller systems under the Born-Oppenheimer approximation, where the electronic position states are coupled to the lattice momentum rather than the usual time-even lattice position. As an example I solve the  $E \otimes (b_1 \oplus b_2 \oplus a_2)$  Jahn-Teller system exactly, showing that once again monopole-like geometric potentials arise, and comment on how this affects the angular momentum of the lattice subsystem.

# Contents

<b>1. Introduction</b> . . . . .	1
<b>2. Time reversal</b> . . . . .	9
2.1 What is time reversal symmetry? . . . . .	10
2.2 Time reversal symmetry in quantum mechanics . . . . .	16
2.3 The evolution operator . . . . .	20
2.4 Ring laser tests of time reversal symmetry violation . . . . .	23
<b>3. The geometric phase</b> . . . . .	27
3.1 Definition . . . . .	29
3.2 The geometric gauge potential . . . . .	30
3.3 The nonadiabatic geometric phase . . . . .	32
3.4 The nonabelian geometric phase . . . . .	34
3.5 Fibre bundles and the geometric phase . . . . .	40
<b>4. Time reversal constraints on the geometric phase</b> . . . . .	45
4.1 Basic results from time reversal symmetry . . . . .	46
4.2 Ihm's results . . . . .	49
4.3 Classifying phases using time reversal criteria . . . . .	53
4.4 Extending Ihm's results . . . . .	57
4.5 The time reversal signature of the geometric vector potential . . . . .	61
<b>5. The U(1) monopole and the geometric phase</b> . . . . .	65
5.1 The Dirac monopole . . . . .	66

5.2	The U(1) monopole and the geometric phase . . . . .	68
5.3	A simple example . . . . .	71
5.4	Monopole mathematics . . . . .	74
5.5	Time reversal and the monopole . . . . .	77
<b>6.</b>	<b>The nonabelian monopole . . . . .</b>	<b>79</b>
6.1	Nonabelian gauge theory . . . . .	80
6.2	The two-Kramers doublet Hamiltonian . . . . .	84
6.3	The geometric vector potential . . . . .	85
6.4	Yang's SU(2) monopole . . . . .	87
6.5	The SU(2) instanton . . . . .	93
6.6	Duality and existence of monopoles . . . . .	98
6.7	Applications . . . . .	100
<b>7.</b>	<b>Jahn-Teller systems and time-odd coupling . . . . .</b>	<b>103</b>
7.1	The adiabatic and Born-Oppenheimer approximations . . . . .	105
7.2	Jahn-Teller systems . . . . .	108
7.3	Time-odd coupling in Jahn-Teller systems . . . . .	114
7.4	Angular momenta in the $E \otimes (b_1 \oplus b_2 \oplus a_2)$ system . . . . .	121
<b>8.</b>	<b>Conclusion . . . . .</b>	<b>123</b>

# List of Figures

2.1	$\beta$ decay parity violation . . . . .	11
2.2	Time reversal asymmetry for a charged particle in a magnetic field . . . . .	14
2.3	“Reversed film” version of time reversal symmetry . . . . .	15
2.4	Structure of a planar ring laser . . . . .	24
3.1	The Aharonov-Bohm effect . . . . .	31
3.2	Projection from a Hilbert space to its ray space . . . . .	33
3.3	Example of fibre bundle theory on a Möbius strip . . . . .	43
4.1	Path extension causing a discontinuous phase change of $\pi$ . . . . .	55
4.2	Winding numbers in a non-simply connected space . . . . .	59
6.1	Path dependence of parallel transport . . . . .	82
7.1	Two E-type normal modes in a lattice with octahedral symmetry . . . . .	109
7.2	Adiabatic potential surfaces for the $E \otimes \epsilon$ Jahn-Teller system . . . . .	110
7.3	A rotational $A_2$ mode coupled to a doubly degenerate set of E-type electronic states . . . . .	115

# 1. Introduction

Quantum mechanics has now been with us for over seventy years, dating back to the early 1920s, and one might therefore be forgiven for thinking that, particularly in the non-relativistic formulation, it would have no more surprises to offer us. Luckily physics is an evolving discipline, constantly throwing up new insights, even in old, supposedly thoroughly explored fields. One such relatively recent surprising development was Berry's 1984 discovery of an extra phase that arises in solutions to the Schrödinger equation, a phase that had been consistently overlooked for decades [31].

While the phase associated with an isolated quantum mechanical wavefunction is, to an extent, arbitrary, since observables are related to the modulus of the wavefunction, this certainly does not mean phases are unimportant. Although the phase of an isolated system is unimportant, when wavefunctions are superimposed their phases interfere and this can give rise to observable consequences.

What Berry demonstrated was that phases can be seen as belonging to two classes. The first class is the easily understood dynamical phase  $\delta$ , which is merely the time integral of the instantaneous energy of the system:

$$\delta = \frac{1}{\hbar} \int E(t) dt. \quad (1.1)$$

Thus a high energy implies a short associated wavelength and hence a rapidly changing phase, while a low energy implies a long wavelength and a slowly changing phase. This aspect of the phase has been well known and understood since the inception of quantum mechanics. This phase is clearly dependent on the state of the system, and the speed at which it is being transported along a path.

The second type of phase, which has become known as Berry's, or the geometric, phase is much more subtle and interesting. What Berry initially proved was that if the Hamiltonian of a non-degenerate quantum system depends on external parameters, and if these parameters are adiabatically varied around a closed loop, then the system picks up an extra phase in addition to the dynamical one. This extra phase,  $\gamma$ , is fundamentally different in origin to the dynamical phase. It is non-integrable, i.e. depends on the path itself and not merely the endpoints, and purely geometric in origin — it depends only on the path taken in parameter space; the speed or energy with which it traverses this path is irrelevant [3]. This geometric phase is completely separate from the dynamical

phase, and can be experimentally isolated [195]. As will be shown in §3.1 this geometric phase is given by

$$\gamma_n(t) = i \int_{x(t_0)}^{x(t)} \langle n(x(t')) | \frac{\partial}{\partial x^\mu} | n(x(t')) \rangle dx^\mu(t'), \quad (1.2)$$

where the  $|n(t)\rangle$  are an adiabatic basis for the system at each point in the parameter space, and the parameters are represented by  $x_\mu$ .

The geometric phase arises from very simple quantum mechanical considerations, and it is worth considering why it was not noticed previously. The reason usually given when ignoring it is that the extra phase  $\gamma$  can be transformed to zero under a U(1) gauge transformation of the form  $|n(x)\rangle = e^{i\eta(x)} |n(x)\rangle$  and can be traced back to the 1928 paper by Fock [71]. This reasoning is correct provided the path taken is not closed. For closed, or cyclic, evolutions the proof fails as the transformation becomes multi-valued, and the extra phase cannot be gauged away [43].

One of the most interesting features of the geometric phase is that it can be recast naturally in a gauge theory context [31, 201]. That is, the phase can be considered as arising from the motion of a charged particle through a gauge field:

$$\gamma = \int_C A_\mu(x) dx^\mu \quad (1.3)$$

where the field  $A_\mu$  is defined as

$$A_\mu = \langle \psi(x) | \frac{\partial}{\partial x^\mu} | \psi(x) \rangle. \quad (1.4)$$

Notice that the phase has exactly the same form as that of the wavefunction of a charged particle moving in an electromagnetic field, the prototypical gauge field. The similarity extends further. As will be detailed in Chapter 3,  $A_\mu$  transforms exactly as a local gauge field under a phase transformation of the basis eigenkets  $|\psi(x)\rangle$ . This correspondence between the geometric  $A_\mu$  and more general gauge fields stems from their similarity in the mathematical language of fibre bundles [169], and will be explored in more detail in §3.5. For this reason I coin the term “geometric vector potential” and often use it interchangeably with  $A_\mu$  and the phrase “gauge field” throughout this thesis.

The gauge field approach becomes even more obvious in molecular physics [42, 126]. Molecular physics is normally studied, at least as a first approach, within the Born-Oppenheimer approximation. This divides the system of interest up into two subsystems: the slowly moving nuclei and the rapidly moving, high energy electrons. Because the electrons are so much lighter than the nuclei, the Born-Oppenheimer approximation assumes that they are slaved to the heavier nuclei and follow them adiabatically. Thus we have a situation tailor-made

for Berry's formulation of the geometric phase, except that the adiabatic parameters are now internal rather than external variables. In this formulation (see Chapter 7) the geometric vector potential makes an appearance in the Hamiltonian governing the nuclear motion. Furthermore, it appears minimally coupled to the nuclear momentum, just as one would expect in the presence of an electromagnetic field [42, 43].

Berry's original formulation of the geometric phase, while ground-breaking, was somewhat limited, restricted as it was to adiabatic evolutions and non-degenerate systems, and it was quickly generalized. The restriction to adiabaticity was removed by Aharonov and Anandan using an approach which highlighted the geometric nature of the phase [3], and the condition of non-degeneracy was removed by Wilczek and Zee who demonstrated that degeneracy led to the geometric vector potential becoming nonabelian [201], leading to all the richness of nonabelian gauge theories. The geometric phase can even be defined for evolutions that are not cyclic or even unitary [36, 128, 160].

Although Berry was the first to quantify the geometric phase and correctly identify its nature there were several earlier papers in the literature which offered hints as to its existence, although the full generality associated with it was not noticed [33]. The first group to explicitly mention that the vector potential need not vanish under an appropriate choice of phase appear to have been Herzberg and Longuet-Higgins in 1963. They showed that the Born-Oppenheimer electronic wavefunction underwent a sign change (that is, a phase shift of  $\pi$ ) if the nuclear coordinates followed a closed path around a conical intersection of two electronic potential-energy surfaces [87]. These things stood until Mead and Truhlar examined the behaviour of the electronic wavefunction in a more general way and showed that the multivaluedness of the wavefunction could be removed at the expense of introducing a vector potential [127]. Mead later coined the term "molecular Aharonov-Bohm effect" to describe the effect of the vector potential [124], a name which was remarkably prophetic given that only four years later the work of Berry [31] and Simon [169] would unify the real and molecular Aharonov-Bohm effects via the more general concept of the geometric phase. It is interesting to note that practically the only situation in which molecular physicists and chemists did not throw out the geometric vector potential was in the analysis of the diatom [133]. In fact, Van Vleck's 1929 Born-Oppenheimer analysis of the diatom treated precisely the effects of the the geometric vector potential, without being remarked on [191]. The geometric phase was almost discovered over fifty years early!

An aside on terminology: although there exists no iron-clad agreement it is common to refer to the adiabatic formulation as Berry's phase and the nonadiabatic version as the geometric phase. Throughout this thesis I have chosen to consistently refer to the phase as the geometric phase as I feel this is a more suitable overall description.

The purpose of this thesis is to explore the geometric phase and the associated geometric gauge fields from a perspective that has previously received only minimal attention. The perspective to which I refer is that of time-reversal symmetry in quantum mechanics. Time-reversal symmetry is one of the more neglected physical symmetries, even though it is probably one of the most interesting and exotic ones, capable of simplifying many problems and generating powerful selection rules [155, 175].

This thesis demonstrates that an intimate link exists between time-reversal symmetry and the geometric phase, or more precisely between time-reversal symmetry *violation* and the geometric phase. I precisely codify this relationship in Chapter 4 and then proceed to explore a number of specific examples, ranging from the the gauge theory of instantons to the Jahn-Teller effect in molecular physics, using time-reversal symmetry.

The format of this thesis is as follows: Chapter 2 introduces the concept of time-reversal symmetry and explores it in some detail. I will show that in the literature there exists some confusion as to the exact action of the time-reversal operator in quantum mechanics, with at least two approaches extant. I extend the action of these two operators, showing how they can be considered to act on time-dependent systems. This is quite novel. Most authors simply take the approach that time-dependent Hamiltonians are automatically not time-reversal symmetric. Considering the action of the time-reversal operator on time-dependent Hamiltonians and hence time-dependent evolution operators allows me to develop tools which I use in Chapter 4 when determining existence criteria for a non-trivial geometric phase. Finally I close the chapter with an analysis of the possibilities of using a planar ring laser, such as the one at the University of Canterbury, to test for violations of parity ( $P$ ) and time-reversal ( $T$ ) symmetry, coming to the conclusion that planar ring lasers in their standard geometry are only capable of detecting  $T$ -violating effects, and not  $P$ -violating effects as has been claimed [67]. Non-standard geometries involving circularly polarized light do allow the detection of  $P$ -violating and  $PT$ -violating effects, but then are incapable of detecting  $T$ -odd effects [177].

Chapter 3 consists mainly of review material, giving the derivation of the basic equations involved in the geometric phase. I also cover the extensions to Berry's basic results, which deal with nonadiabatic evolutions (sometimes called the Aharonov-Anandan phase) and degenerate systems which lead to nonabelian gauge potentials (also known as the Wilczek-Zee phase). In analysing these extensions I point out some common errors and inconsistencies in the formulation of the nonabelian phase that exist in the literature, and show that some of the confusion arises from different definitions of the gauge potential. I derive the gauge transformation rules for the two major definitions of the gauge potential, which I will use later in Chapter 6 when considering the  $SU(2)$  instanton. Because problems involving the nonabelian phase are often difficult due to the non-commutativity of the gauge potential in the path integral (or Wilson loop),

I derive an original result that states under what circumstances it is possible to decompose a nonabelian problem into a set of abelian ones. This result will be used in Chapter 4. Finally, I close the chapter with an analysis of the geometric phase from the mathematical perspective of fibre bundle theory. It is under this formulation that the truly geometric nature of the phase is made clear, as well as its generality and link to many other concepts such as parallel transport and, most interestingly, general gauge theories. This link between the geometric phase and gauge theories is quite beautiful, and will crop up again and again in this thesis, particularly in Chapters 5 and 6 where I consider  $U(1)$  monopoles and their extension, the  $SU(2)$  instanton.

Having thoroughly explored time-reversal symmetry in Chapter 2 and the geometric phase in Chapter 3, Chapter 4 proceeds to marry the two. The connection between time reversal and the geometric phase is a startlingly intimate one, although this has gone practically unnoticed in the literature. Aside from the odd well-known result such as geometric phases in time-even systems being quantized in multiples of  $\pi$  [103], the only work on the subject appears to have been done by Ihm [91, 92, 93]. In his boldest paper he attempted to attribute the existence of any non-trivial geometric phase to the breakdown of time-reversal symmetry [93]. His analysis contained the seeds of many good ideas but was fatally flawed in a number of ways which rendered his proof invalid. In Chapter 4 I examine and then correct his proof, coming to the same conclusion — non-zero geometric phases can always be traced to the violation of time-reversal symmetry. Further, Ihm's proof considered only adiabatic evolutions and three-dimensional parameter spaces, conditions which severely limit the scope of the theorem. To this end I extend the proof to include both nonadiabatic evolutions and arbitrary-dimensional parameter spaces. In the process of constructing the proof I arrive at a number of "selection rules" on form of the geometric phase matrix such as for a two-level Kramers system it must be  $SU(2)$ . I close the chapter with a discussion on the time-reversal character of the geometric phase and geometric vector potential. Should they be considered time-even or time-odd? I come to the conclusion that this depends on whether the parameters driving the evolution of the system are considered to be internally or externally governed. As a consequence Ihm's suggestion of using the geometric phase as a probe of  $PT$ -breaking properties of high-temperature superconductors [91] is shown to be unworkable.

Chapter 5 is concerned with  $U(1)$ , or Dirac, monopoles, which will arise again and again throughout this thesis. A large part of the chapter is review, explaining the topological properties of the monopole. After this background the monopole is linked to the geometric phase by demonstrating that a geometric vector potential functionally equivalent to that of Dirac's monopole is generated by any arbitrary two-level system that includes a point of degeneracy. The geometric phase generated by the two-level crossing is classified according to the scheme in Chapter 4. I then re-analyse the monopole from a more mathematical perspective, with the intention of demonstrating the mathematical techniques

within a simpler  $U(1)$  abelian context before applying them to the more involved nonabelian problems of Chapter 6. As a consequence of the mathematics I show that a simple topological reason for the quantization of spin into units of  $\frac{1}{2}\hbar$  exists. That is, rather than using algebraic commutation rules or group theory to prove the quantization of spin, it is possible to create a proof based on the geometric phase and monopole topology. While some connection between spin and geometric monopole topology has been noted [8], to my knowledge this is the first time it has been explicitly shown that it is possible to derive canonical spin quantization in this light. Chapter 5 concludes with a speculative look at other possible connections between time-reversal symmetry and monopoles in general, in both the geometric and electromagnetic form.

The monopole is closely related to time-reversal symmetry as can be seen by noting that the two-level crossing criteria are precisely fulfilled by a Kramers doublet, that is a pair of states that are degenerate due to time-reversal symmetry. Given the general importance of the two-level crossing, a natural question to ask is, what happens upon a *pair* of degenerate Kramers doublets crossing and becoming hyper-degenerate at a certain point? Does one obtain monopole-like behaviour? This theme and its topological consequences are the main thrust of Chapter 6. Because we must now deal with degeneracy the geometric phase formalism becomes the nonabelian version of Wilczek and Zee, and thus §6.1 is devoted to a review of the basics of nonabelian gauge theory. I then proceed to derive the most general Hamiltonian describing such a paired Kramers doublet system, and from it calculate the two associated geometric gauge potentials. These potentials are similar in form to the standard Dirac potentials but are  $SU(2)$  rather than  $U(1)$ . To determine whether my potentials do indeed represent a nonabelian monopole I turn to the work of Yang [205] who constructed the  $SU(2)$  monopole and came to the conclusion that only two forms of such a monopole exist, both in five dimensions. I show that my two potentials representing the higher and lower energy doublets exactly correspond to Yang's  $SU(2)$  monopole, even down to the way they behave under gauge transformations.

Another way of looking at my gauge fields is from the perspective of instantons. As early as 1988 Avron *et al.* noticed from topological considerations that a pair of Kramers doublets would give rise to a vector potential with the same properties as the  $SU(2)$  instanton, although they did not explicitly construct the potentials [25]. I demonstrate, using various coordinate systems and a conformal mapping, that my potentials are functionally equivalent to those of the instanton and anti-instanton. Furthermore, I show in §6.7 that the  $SU(2)$  monopole/instanton potential arises in many situations in molecular physics, connecting the beautiful mathematics of nonabelian gauge theory and particle physics with the more prosaic world of molecular and chemical physics.

I close my thesis in Chapter 7 with an examination of the effects that time-odd coupling and the geometric vector potential have within the arena of molecular physics using the adiabatic approximation. Time-odd coupling between the

nuclear (lattice) motion and the electronic states has been practically ignored through most of this century, usually being dismissed as negligible [76, 180]. The only exceptions appear to be Fletcher [69], Fletcher and Pooler [70] and, in the context of breakdown of sum rules in Ham reduction factors in Jahn-Teller systems, Payne and Stedman [144, 145, 146]. More recently it has been mentioned by Moore and Stedman [134] and Riley and Furlan [152]. It is therefore an area which is well overdue for exploration, particularly within the context of the geometric phase. The central theme of the chapter is Jahn-Teller systems as the geometric phase has been shown to have considerable importance in this situation. However, with the exceptions mentioned above, to my knowledge Jahn-Teller systems have only been considered with time-even coupling and, as the rest of this thesis has demonstrated, time-oddity gives rise to the most interesting geometric phase effects. What is particularly interesting is that in the classic time-even  $E \otimes e$  Jahn-Teller system so beloved of theorists everywhere the geometric phase is responsible for a shift in the nuclear angular momentum of  $\frac{1}{2}\hbar$  and thus changes the orbital angular momentum so it acquires a fermionic spectrum. This shift is linked to a geometric phase of  $\pi$ , which we expect from a time-even system. Given the link between spin and statistics and particle interchange phases [200], an intriguing question is what happens to the angular momentum under a time-odd coupling scheme where arbitrary phases are generated? Is it possible that some form of anyonic (fractional) statistics may be realized? In §7.3 and §7.4 I construct and solve the time-odd  $E \otimes (b_1 \oplus b_2 \oplus a_2)$  system to explore general time-odd coupling behaviour and answer these questions.

Finally, the conclusion to my thesis is found in Chapter 8 and consists of a summary of my major results along with suggestions for further work.



## 2. Time reversal

Time reversal has long been of interest to physicists, not least because time only seems to run one way, despite the fact that so far all explicitly known laws are time reversal symmetric. Although the examination of the various arrows of time is of great interest in both philosophy and physics, in this chapter I consider only time reversal in quantum mechanics.

The arrow of time occurs in quantum mechanics too. In our current understanding of the quantum world, the evolution of a system consists of two utterly different parts. There is the normal evolution of the system governed by the Schrödinger equation which is linear, deterministic and (globally) time reversal symmetric. Then there is the effect of an external measurement on the system which results (if we subscribe to the Copenhagen interpretation) in the reduction of the state vector, forcing the system to suddenly adopt one particular state [41, 46]. This part of the evolution is nonlinear, not deterministic and certainly not time reversal symmetric. This problem has yet to be resolved, although there are many attempts to do so (for example [41, 141]). Despite this difficulty it is still possible (and useful) to restrict our application of time reversal symmetry to the smoothly evolving, deterministic part and to construct a time reversal operator.

In this chapter I first consider the nature of time reversal. This is necessary as the literature is rife with confusing and sometimes conflicting approaches. Even the question “what is the form of the time reversal operator?” seems to have divided authors in to two major camps. In §2.3 I will look at the behaviour of the evolution operator under time reversal, another point of some disagreement, and develop an approach that will prove useful when I go on to apply time reversal considerations to the geometric phase.

In §2.4 I consider the relevance of a ring laser in the context of time reversal and parity symmetries, a question which grew from my initial interest in using a ring to search for global time reversal and parity symmetry violation. In doing this I reach an interesting conclusion regarding the ability of a ring to detect the breakdown of these symmetries.

## 2.1 What is time reversal symmetry?

A large part of the study of physics is the study of symmetry.

Throughout history symmetries in physical problems have been sought after due to the fact that they tend to give rise to simpler methods of solution. As physics has progressed symmetry has come to play a larger and larger role. One of the prime motivations for this was the formulation of Noether's theorem [63, 138] which exposed the general link between conservation laws and symmetries. With modern physics it has now reached the stage where physicists do more than *look* for symmetries in a specific problem in order to facilitate the solution; they *demand* that their theories have certain symmetries actually built into them. They must be couched in a language which makes these symmetries manifest, for example the use of vectors for Galilean invariance or covariant notation for Lorentz invariance.

With modern field theory this has been taken even further. Physicists now demand that global symmetries in the equations must be preserved if they are made local, that is, if they are made to depend on position. This demand in turn gives rise to new physics, predicted purely on the basis of symmetry. This is the essence of gauge theory upon which most modern theories depend.

The meaning and implementation of most symmetries are obvious, such as rotation, space translation and boosts. That is, it makes no difference if one performs an experiment in a laboratory frame that has been rotated by some angle to another, whether the experiment is performed here or ten kilometers away, or whether it is performed at rest or in constant motion. These basic symmetries correspond via Noether's theorem to the conserved quantities of angular momentum, linear momentum, and motion of center of mass. Similarly we have translation in time — it makes no difference whether one carries out an experiment now or next week — corresponding to the conservation of energy.

These are obvious major continuous symmetries that our universe appears to exhibit, and are contained in the Lorentz group. However there also exist two discrete symmetries — reflection in space and reversal in time. Reflection is easy to understand. Suppose we reflect the universe in a mirror, or more properly set up a coordinate system covering the region of interest and apply a parity transformation so that every point  $\mathbf{x}$  is replaced by its point at  $-\mathbf{x}$ :

$$\mathbf{x} \rightarrow \mathbf{x}' = -\mathbf{x}. \quad (2.1)$$

If the transformed universe one sees is sensible and appears to obey all the known laws of physics then the universe is left/right symmetric, that is, invariant under a parity transform.

This was thought to be the case until the 1950s, when a variety of hints began to suggest suggest otherwise, culminating in the classic experiment by Wu and

coworkers in 1957 [203]. By considering  $\beta$ -decay of spin-aligned cobalt-60 they demonstrated that the weak interaction maximally violates parity symmetry (Figure 2.1).

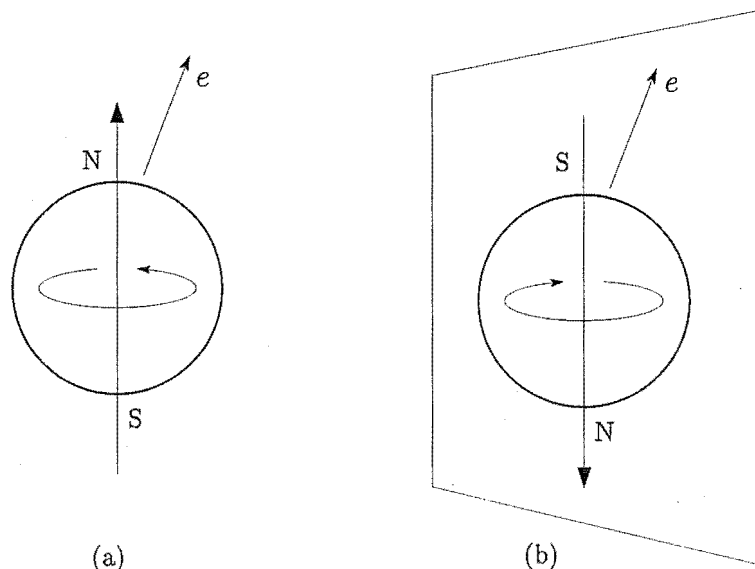


Fig. 2.1: (a) In  $\beta$  decay of  $^{60}\text{Co}$  most electrons are emitted in the direction of the nuclear spin. (b) In a mirror image world most electrons are emitted *opposite* to the nuclear spin.

Provided we do not deal with the weak nuclear force, however, the universe does indeed appear to be parity invariant. Thus all classical mechanical systems and quantum electrodynamics has parity as a valid symmetry.

We come now to the symmetry of time reversal. This is, as I will show, in many ways a much more subtle and misunderstood symmetry when applied to quantum mechanics, and has a variety of interpretations.

The first question is, what does one mean by “time reversal”? In all other symmetry operations such as reflection, rotation, translation and so on it is possible to actually carry out the operation. We can, for example, talk about what a particular experiment would look like if it were reflected in a mirror. Chiral objects exist in nature — we can generally create a mirror image experiment without undue difficulty. Unfortunately we do not know of any way to reverse the flow of time to see what results. It is, however, possible to do something that appears effectively the same using a film analogue. That is, one makes a movie of some particular happening, then plays the film backwards. This certainly creates the illusion of time running backwards, with all the attendant effects such as reversal of velocities and angular momenta and suchlike, and it is this I will use as the first model of time reversal. If the universe in the reversed film

does not break any physical laws and appears to be possible then we can say that the universe is indeed time reversal symmetric.

With this model the idea of time reversal symmetry seems ludicrous at first glance. In the real world we never see broken crockery reassembling itself on the floor and leaping up into the air to land on a bench, or a mixture of two gases in a box separating themselves out to lurk at either end of the container. However it is important to realize that this is not *impossible*, just very unlikely. It is possible, for example, for the molecules in the floor to conspire to vibrate in such a way as to send shock waves into the broken cup to knock the pieces together so exactly that it reassembles and is hurled in exactly the reverse trajectory to land on the bench; or for the gas molecules to collide in such a way that random thermal collisions eventually knock them all back to their respective ends of the container. This is known as microreversibility, and as the name suggests this arrow of time gets less and less obvious at smaller and smaller scales. As an example of how robust this particular arrow of time is, Blatt has considered a box containing gas of a mere 100 molecules [38]. If transitions from one side of the box to the other take place at one million per second the time required before we can expect to have all the molecules in one half of the box will be of the order of  $3 \times 10^{16}$  years.

The reason the above examples are unlikely is because of thermodynamics. There are an incredible number of possible final states and only a few initial states, or, to put it another way, one must specify the final conditions much more precisely than the initial ones in order to have the reversal occur. Thermodynamics is derived in the infinite particle limit, which never occurs in reality. Admittedly reality is a good approximation of this limit, which is why macroscopic systems appear to violate time reversal symmetry, but it is not exact.

So, to sum up: if the laws of physics do not *prohibit* what we see in the time-reversed movie, then the system we are observing is time reversal symmetric (TRS).

In classical mechanics, because of the effects of reversing velocities, the operation of *time reversal* can be considered equivalent to *motion reversal*. In classical mechanics systems are often TRS due to the fact that the Lagrangian describing the system is quadratic in the velocities, terms of first degree being absent, giving

$$L(\dot{x}, x) = L(-\dot{x}, x) \quad (2.2)$$

thus ensuring the equations of motion are invariant under motion reversal. In other words, if  $x(t)$  is a solution to Lagrange's equations,  $x(-t)$  is an equally valid solution.

At this point we must make an important distinction — the difference between local TRS and global TRS. If we assume that some set of laws exhibits TRS then that only means that an isolated system considered as a whole will exhibit the symmetry (global TRS). When we consider a subsystem of the total system it

may not behave in a TRS way, even though if we step back and view the complete system TRS will be restored. As an example consider a charged particle moving perpendicular to an external magnetic field (Figure 2.2). Its path will curve according to the Lorentz force law:  $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$ . If we reverse a film taken of this process without reversing the magnetic field we will find that the particle curves in the opposite direction to that predicted by the force law — the subsystem locally violates TRS. The reason for this, of course, is that now the assumption of a Lagrangian quadratic in velocity is incorrect. There is now a term of the form

$$L(\mathbf{A}) = q\mathbf{A} \cdot \mathbf{v}. \quad (2.3)$$

However, this Lagrangian only describes the subsystem of the charged particle. If we step back and look at the current distribution that causes the external  $\mathbf{A}$  field and reverse the motion of the charge carriers then the field will reverse, and the time reversed particle trajectory is now physically valid. Thus global TRS is preserved, at least in classical mechanics and electromagnetism.

It is worth noting, however, that while this is assumed by almost all physicists, there are objections that can be raised in pathological cases (generally within literature pertaining to the philosophy of science). For example Hutchinson gives a situation which consists of a particle with exactly specified initial conditions moving in a potential such that the particle comes to rest precisely at a local maximum in the potential [90]. At this point the particle will theoretically cease motion since it has no net force exerted on it provided it stays exactly at that point. If a film were made of the process of the particle slowing and stopping, the time-reversed version would have the particle suddenly beginning motion and retracing its path with no initial impulse to remove it from its stationary position. This would, at least technically, violate the laws of classical mechanics. At least two points must be born in mind, however. First, in reality it is impossible to specify the initial conditions of the particle to the infinite accuracy that would be required. Second, even were the particle at rest on the potential maximum only an infinitesimal displacement would be needed to restore the motion of the particle and thus appealing to random thermal vibrations or the uncertainty principle would defeat the assumption of a stationary particle.

It could be argued that time reversal symmetry is still compromised as the problem with appealing to small random variations to restart the motion of the particle is that they are just that, random. The particle could start moving down any direction on the potential “hill”, not necessarily the way it went “up”. Nonetheless should a film be made of the process (up to the point of zero particle velocity) and then reversed, the time-reversed film would not show anything that violated the laws of physics. The backward run would certainly be one possible alternative, provided some real physical mechanism is provided to remove the particle from its unstable potential maximum.

Thus the answer would appear to be: Yes, in reality classical mechanics is

time reversal symmetric. Theoretically, if all aspects of thermal vibrations, the quantum uncertainty principle and similar random factors are ignored and just pure classical mechanics is considered, in certain pathological cases some trajectories can be considered non-time reversal symmetric. If this reading is applied, however, the real time reversal symmetry of the system can easily be restored by using an open interval in the reversed-film analogy, i.e. by not including the end points of the trajectory.

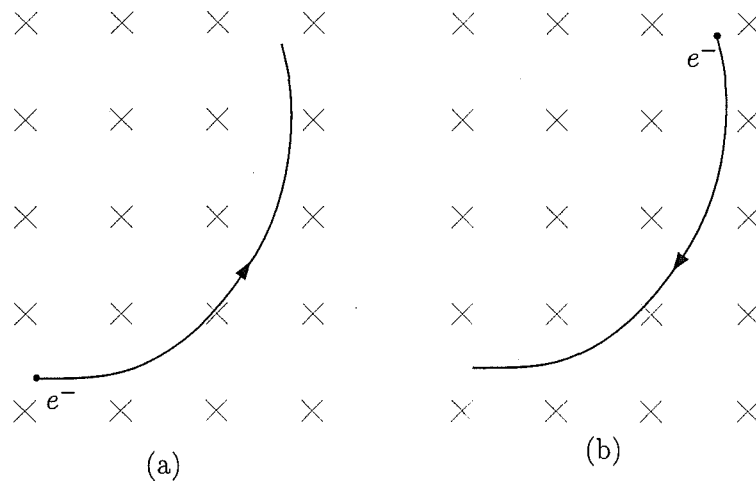


Fig. 2.2: (a) An electron bends to the left in a magnetic field as predicted. (b) In the time reversed trajectory the electron curves right, violating the Lorentz force law.

All this aside, the distinction between global and local TRS will prove to be important when I consider time reversal character of the geometric vector potential in §4.5, and the possibility of anomalous angular momentum quantization in Jahn-Teller systems in §7.4.

Does the universe have global time reversal symmetry? Somewhat indirect experiments suggest that it does not. It is a great deal harder to test for time reversal symmetry than parity. This is because while all particles are eigenstates of the parity operator, no particle is an eigenstate of the time reversal operator. The most direct test would be take some particular reaction, say  $n + p \rightarrow d + \gamma$  and run it in reverse:  $d + \gamma \rightarrow n + p$ . For the corresponding conditions on momentum, energy and spin the reaction rate should be the same in either direction (this is the principle of detailed balance, and it follows directly from time reversal invariance). This is easy to do for the strong and electromagnetic interactions and has been well tested. No evidence for time reversal violation has been seen. It is much harder to do this sort of experiment with the weak interaction; unless one uses a pure weak process such as neutrino-neutrino interaction effects from the strong and electromagnetic interactions wipe out any weak interaction signature. Since neutrinos are very difficult to deal with, most practical experi-

ments look for quantities that should be exactly zero if time reversal is a valid symmetry, such as the neutron electric dipole moment. For a recent review of current progress see for example [12, 172].

Lack of direct evidence of time reversal symmetry violation aside, there exists a more indirect experiment that suggests that this symmetry is, as in the case of parity, also globally violated. This evidence comes from the neutral  $K^0$  meson [75]. The Fitch and Cronin experiment demonstrated that the  $K^0 - \bar{K}^0$  system very slightly violated  $CP$  symmetry, that is, the combined symmetry of parity and charge conjugation [54]. However, in quantum field theory there exists the  $CPT$  theorem. Derived on very general assumptions it states that the combined symmetry of charge conjugation ( $C$ ), parity ( $P$ ) and time reversal ( $T$ ) *must* be an exact symmetry of nature [181, 197]. Consequently, since  $CP$  is known to be violated, there must exist a compensating violation in  $T$  in order to retain the symmetry of the product. Current and recent experimental searches for  $CP$  violation are summarised by Winstein and Wolfenstein [202].

The definition of TRS as “the motion reversed system does not violate any laws of physics” can be hard to apply. Let us try for something more workable. An intuitively appealing definition is as follows: Start with the system initially at time  $t = 0$  and let it run forward as determined by the initial conditions and external influences. At time  $t_1$  apply the operation of time reversal to it, then continue to let it evolve. If the system is time reversally symmetric we expect it to retrace its path, and at time  $2t_1$  we expect it to be back in the same configuration as at time  $t = 0$ , except that all momenta and similar time-odd quantities will be reversed (Figure 2.3). Classically that means that the reversed trajectory can be written

$$\mathbf{x}(t) = \mathbf{x}(2t_1 - t), \quad \mathbf{v}(t) = -\mathbf{v}(2t_1 - t). \quad (2.4)$$

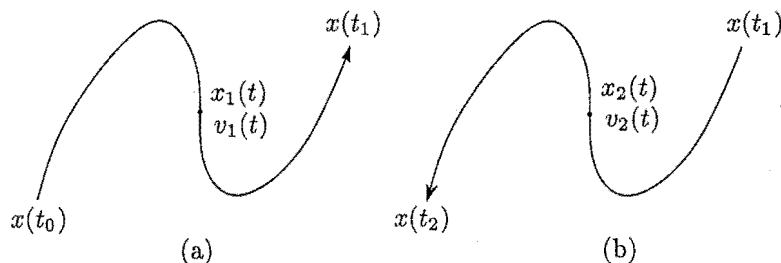


Fig. 2.3: In a time reversal symmetric system, applying time reversal at time  $t_1$  will result in the system exactly retracing its evolution so that  $x_2(t) = x(2t_1 - t)$  and  $v_2(t) = -v(2t_1 - t)$ .

If the external influences on the subsystem are time-dependent then things become a little more complicated. To my knowledge this situation has not been

considered in the literature, brushed under the rug with the implicit assumption that if time-dependent influences are involved then the entire notion of time reversal symmetry is useless. This is not the case. For example the Hamiltonian in the evolution operator can be time-dependent, and it is certainly a valid (and vital, in the case of the geometric phase) question whether a quantum state will retrace its path or not.

The main point is this: Although we are only time-reversing the subsystem under consideration, we must decide whether to reverse the time-dependence of the external field about  $t_1$  as well. For example, suppose we have a constant time-even external field  $X$  acting on a subsystem. Assume in this case Equation (2.4) holds and thus the subsystem is time-even. Now, if we let  $X(t)$  be time-dependent, and asymmetric about  $t_1$  then (2.4) will not hold and the subsystem would not appear TRS, whereas if  $X(t)$  is symmetric about  $t_1$  then the subsystem would be TRS.

One way around this is to assume that the time reversal operation will reverse the time-dependence of the external influences too, that is  $\bar{X}(t) = X(2t_1 - t)$  where  $\bar{X}(t)$  is the functional form of the external operator after the time reversal operation. This seems inconsistent with the idea of considering subsystems separately though, and thus I propose that a more logical approach is to leave external influences unaffected by the time reversal operation. Consequently, if we assume a system is TRS when ignoring external influences, then upon the introduction of time-dependent external influences it will only appear to remain TRS if the influences are time-even and symmetric in time about the time  $t_1$ , the point in time where the time reversal operation is carried out. This approach leads to some interesting results. For example if the external effects are time-*odd*, then the subsystem will appear to be TRS if the influence is antisymmetric about  $t_1$  and not TRS if the influence is symmetric about  $t_1$ .

## 2.2 Time reversal symmetry in quantum mechanics

Now that I have defined what time reversal *symmetry* for a system is, let us approach the trickier problem of what the time reversal operator  $T$  should do in quantum mechanics, and how it should be defined. We must be cautious in carrying over the notion of time reversal from classical to quantum mechanics as there are a number of subtleties.

My starting point will be kinematics. As in classical mechanics, we expect symmetries to come out in the kinematics. We require this in order to keep the equations of motion invariant. In quantum mechanics the kinematics are governed by commutation relations, and to have a kinematically admissible transformation, we require it to be consistent with the commutation relations [155]. So,

for example, if

$$x'_i = Tx_iT^{-1} \quad (2.5)$$

$$p'_i = Tp_iT^{-1} \quad (2.6)$$

$$L'_i = TL_iT^{-1} \quad (2.7)$$

then we must have

$$[x'_i, p'_j] = T[x_i, p_j]T^{-1} \quad (2.8)$$

$$= Tih\delta_{ij}T^{-1}. \quad (2.9)$$

In addition, to correspond with classical mechanics we require motion reversal, that is for quantities such as position, momentum and angular momentum we need

$$Tx_iT^{-1} = x_i \quad (2.10)$$

$$Tp_iT^{-1} = -p_i \quad (2.11)$$

$$TL_iT^{-1} = -L_i. \quad (2.12)$$

Putting equations (2.10) and (2.11) into (2.8) and (2.9) show that  $T$  must be an *antilinear* operator, i.e.

$$TiT^{-1} = -i. \quad (2.13)$$

That is,  $T$  includes in its action the complex conjugation operation  $K$ .

Up to this point all authors agree, although there are many different arguments for demonstrating antilinearity, and many prefer to work from the Schrödinger equation rather than the, to my mind, more logical commutation relations.

It is important to note that  $K$  is generally defined to act on complex numbers in operators and on the coefficients of basis kets in the expansion of an arbitrary state, but not on the basis kets themselves [82, 157]. Thus  $K$  is basis dependent.

To see this, consider two basis states  $|\phi_1\rangle$  and  $|\phi_2\rangle$ . Define a new basis

$$|\psi_{\pm}\rangle = \frac{1}{\sqrt{2}}(|\phi_1\rangle \pm i|\phi_2\rangle). \quad (2.14)$$

We require  $K|\psi_{\pm}\rangle = |\psi_{\pm}\rangle$  if  $K$  is to act only on coefficients of states. Thus  $K$  must be redefined with the basis.

Consequently, within this convention, for  $T$  to be a basis independent operator, and for its commutation product with other observables such as  $H$  to remain basis independent, it must include another unitary part  $W$  to compensate for the basis dependent action of  $K$ . This will ensure that  $T$  has basis independent matrix elements:  $\langle\alpha'|T|\beta'\rangle = \langle\alpha|T|\beta\rangle$ .

Not all authors take this approach. Goldberger and Watson [80], for example, allow  $K$  to operate on the basis kets as well, and state “all statements about time reversal are really representation dependent, since the operation of complex conjugation must be obtained explicitly.”

We now turn to the effect of  $T$  on expressions involving time explicitly. We expect the free-particle Schrödinger equation governed by the Hamiltonian  $H_0$  to be covariant under time reversal. Applying  $T$  and defining the time reversed quantities by an overbar we find

$$\begin{aligned} TH_0|\psi(x, t)\rangle &= Ti\hbar\frac{d}{dt}|\psi(x, t)\rangle \\ H_0|\bar{\psi}(x, t)\rangle &= -i\hbar\frac{d}{dt'}|\bar{\psi}(x, t)\rangle \\ &= i\hbar\frac{d}{dt}|\bar{\psi}\rangle \end{aligned} \quad (2.15)$$

showing that the Schrödinger equation is invariant as we require if we assume that part of the action of  $T$  is to send  $t \rightarrow -t$ , i.e. to anticommute with the operator  $\partial/\partial t$ . To put it another way, if the state  $|\psi(t)\rangle$  is a solution to the Schrödinger equation then so is the time-reversed state  $|\bar{\psi}(t)\rangle$ . This is, of course, what would seem reasonable to expect from a time reversal operator. Another point to note is that to get the covariance of the Schrödinger equation  $T$  must commute with the Hamiltonian, another desirable feature considering that we are dealing with the free particle case. This commutation is justified with greater rigour in §2.3.

This argument also holds in the Heisenberg picture. The equation of motion governing an operator  $O(t)$  is

$$\frac{dO(t)}{dt} = \frac{1}{i\hbar}[O(t), H]. \quad (2.16)$$

Applying  $T$  and assuming that  $T$  commutes with  $H$  we obtain

$$\frac{d\bar{O}(t)}{dt} = \frac{1}{i\hbar}[\bar{O}(t), H] \quad (2.17)$$

where I have taken  $t \rightarrow -t$  and  $i \rightarrow -i$ .

From the above arguments one can see that the time reversal operator  $T$  consists of four distinct parts:

$$T = WKU\tau \quad (2.18)$$

with  $U$  taking  $\mathbf{p} \rightarrow -\mathbf{p}$ ,  $\mathbf{x} \rightarrow \mathbf{x}$ ,  $\mathbf{B} \rightarrow -\mathbf{B}$  etc and  $\tau$  taking  $t \rightarrow -t$ . The operator  $K$  is antilinear and has the effect of complex conjugation, and the  $W$  operator counters the basis-dependent action of  $K$ . Barron [27], Capri [50], Fowler [28], Heine [86], and Kaempffer [99] take this approach. This form of the

time reversal operator is often called Wigner time reversal, and would appear to be the most reasonable form for the operator. However not all, or even most, authors use this definition. It is common to define a time reversal operator (I denote it  $T_{mr}$ , as it is often linked to the classical ideas of motion reversal) that anticommutes with  $i\hbar\partial/\partial t$ , essentially defining

$$T_{mr} = WKU. \quad (2.19)$$

However, with this approach the Schrödinger equation is not covariant. Also, any authors that discuss time reversal must use  $t \rightarrow -t$  at some point. To get around this, those who use  $T_{mr}$  are forced to use a Schrödinger equation that runs backward in time and explicitly insert  $-t$  into kets and wavefunctions. Böhm [40], Gottfried [82], Lee [114], Merzbacher [129], Messiah [130], Sachs [155] and Sakurai [157], take this approach.

To understand this better, consider the behaviour of a wavefunction describing a plane wave under the two time reversal operators. Using wavefunctions rather than states, and assuming a time reversally symmetric Hamiltonian, (2.15) becomes the statement that if the wavefunction  $\psi(x, t)$  is a solution of the Schrödinger equation then so is the time-reversed wavefunction

$$\bar{\psi}(x, t) = T\psi(x, t) = \psi^*(x, -t). \quad (2.20)$$

where the asterisk represents complex conjugation. Now, applying  $T_{mr}$  to the Schrödinger equation and letting  $T_{mr}\psi(x, t) = \tilde{\psi}(x, t)$ , we find that

$$H\tilde{\psi}(x, t) = -i\hbar\frac{d}{dt}\tilde{\psi}(x, t) \quad (2.21)$$

or, making the transformation  $t \rightarrow -t$  and comparing with the known solutions we obtain  $\tilde{\psi}(x, -t) = \psi^*(x, -t)$  which gives us the effect of operating on the wavefunction by  $T_{mr}$ :

$$\tilde{\psi}(x, t) = \psi^*(x, t). \quad (2.22)$$

Now consider a wavefunction described by a plane wave travelling to the right:

$$\psi(x, t) = e^{i(kx - \omega t)}. \quad (2.23)$$

Applying the two forms of time reversal we obtain

$$T\psi(x, t) = \bar{\psi}(x, t) = e^{-i(kx + \omega t)} \quad (2.24)$$

$$T_{mr}\psi(x, t) = \tilde{\psi}(x, t) = e^{-i(kx - \omega t)}. \quad (2.25)$$

Applying the momentum operator  $\mathbf{p} = -i\hbar\nabla$  to the two time-reversed wavefunctions we get

$$\mathbf{p}\bar{\psi} = \mathbf{p}\tilde{\psi} = -\hbar k \quad (2.26)$$

demonstrating that the momentum has reversed in sign in both cases. However,  $\bar{\psi}$  describes a wave travelling to the left, which is what is expected after time reversal, whereas  $\tilde{\psi}$  still describes a right-travelling wave. To get around this difficulty one can choose to interpret the wave from a point of view where time is flowing backwards. This, however, seems to be an unnecessary complication which can best be handled by using time-reversed rather than motion-reversed states.

The two operators  $T$  and  $T_{mr}$  both have something to do with time reversal, but in some situations with a time-dependent Hamiltonian  $T$  can prove to be a more useful operation, as shown in §4.2. The idea of different time reversal operators is not unheard of. For example Aharonov *et al.* consider the extended Aharonov-Bohm effect, where a cloud of charge rather than a pointlike charge is used. They find useful a “limited” time reversal operation that reverses momentum but leaves the electromagnetic vector potential unchanged [5].

### 2.3 The evolution operator

The evolution operator is defined as

$$U(t_1, t_0) = \mathcal{T} \exp \left[ -\frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] \quad (2.27)$$

where  $\mathcal{T}$  is the time-ordering operator, from here on assumed. It acts on a state via

$$U(t_1, t_0)\psi(t_0) = \psi(t_1). \quad (2.28)$$

Using the evolution operator it is possible to formulate a definition of TRS along the lines of the “movie” analogy laid out in §2.1. Running the film forward is no problem. In the Schrödinger representation the “frame” is given by  $|\psi(t)\rangle$ . “Running forward” is effectively given by the evolution operator  $U$ . It should be noted that this does assume that a system is always in some particular state of being, and is thus, to some extent, a classical interpretation [121]. This aside, to describe the film analogy mathematically, we take the time interval  $[t_0, t_2]$  and define  $t_1 = (2t_2 - t_0)/2$  and  $\Delta t = t_1 - t_0$ . Then appealing to the classical analogy (2.4) we can say: A system is time reversal symmetric if

$$U(t_1 + \alpha, t_1)T_{mr}\psi(t_1) = T_{mr}\psi(t_1 - \alpha) \quad (2.29)$$

for all  $\alpha \in [0, \Delta t]$ . It now becomes clear why I have labelled  $T_{mr}$  as such — it closely corresponds to the classical concept of a system retracing its path as in (2.4).

As shown in §2.2 however, the time reversal operator can include a component taking

$$t \rightarrow -t. \quad (2.30)$$

This particular transformation includes the assumption that the time reversal operator is applied at  $t = 0$ . In the film analogy we wish to apply it at time  $t = t_1$ , and reflect time about this point. Thus the correct transformation is  $\tau : t \rightarrow 2t_1 - t$ . This gives

$$U(t_1 + \alpha, t_1)T\psi(t_1) = T\psi(t_1 + \alpha), \quad (2.31)$$

because  $\tau : t_1 \rightarrow t_1$  and  $\tau : t_1 - \alpha \rightarrow (t_1 + \alpha)$ .

(2.29) implies that for a TRS Hamiltonian  $H$ , if a state  $\Psi = |\psi(0)\rangle$  evolves into  $\Phi = |\psi(t)\rangle$  after time  $t$ , then the motion-reversed state  $\bar{\Phi} = T_{mr}\Phi$  would evolve under  $H$  into  $T_{mr}\Psi = \bar{\Psi}$ . However using the other definition (2.31) the time reversed state  $T\Psi = \bar{\Psi}$  will evolve into  $T\Phi = \bar{\Phi}$ .

While the time reversal operator  $T_{mr}$  accords more closely with the classical notion of motion reversal it does not seem so reasonable in the quantum mechanical arena. Furthermore, the distinction between  $T$  and  $T_{mr}$  disappears in the time-independent context which is how time reversal is generally considered in any case. If we begin to consider how to define time reversal symmetry in a system with a time-dependent Hamiltonian then the difference becomes important. To see this, consider the natural question: How do the above definitions of time reversal symmetry compare with the more commonly used

$$THT^{-1} = H? \quad (2.32)$$

(2.32) is almost always used in a time independent context, and very little has been said of the form of time reversal symmetry should the Hamiltonian have some external time-dependent perturbation. If (2.31) is differentiated with respect to  $\alpha$  then the condition

$$TH(t)T^{-1} = H(t) \quad \forall t \in [t_0, t_2] \quad (2.33)$$

is obtained.

Differentiation of (2.29) yields

$$T_{mr}H(t_1 - \alpha)T_{mr}^{-1} = H(t_1 + \alpha) \quad (2.34)$$

for all  $\alpha \in [0, \Delta t]$ , which is not identical to (2.33) as might have been expected. Equations (2.33) and (2.34) are of course equivalent if the Hamiltonian is symmetric in time about point  $t_1$ , that is  $H(t_1 - t) = H(t_1 + t)$ . This is what is usual when considering time reversal; after all, when playing the film backwards one would expect to apply the same Hamiltonian in reverse order. Thus if the time reversal operator is applied at the time  $t_1$  then  $H(t_1 + \alpha) = H(t_1 - \alpha)$ . This is equivalent to global time reversal — one reverses all possible influences in the backwards run. This need not be the case, as was alluded to at the end of §2.1, particularly if one considers external influences not to be reversed by application of time reversal. A Hamiltonian can be TRS and not be symmetric about about

the point of time reversal if the operator  $T$  is used. Due to the definition (2.18) if the Hamiltonian has a time-odd time dependence, e.g. a term proportional to  $\sin(t - t_1)$ , coupled with a time-odd operator such as a magnetic field, then the system is time reversally symmetric regardless of the fact that the Hamiltonian is time odd.

An important question is what form the evolution operator, given by (2.27), takes under time reversal given the Hamiltonian is TRS.

This problem is considered by many quantum mechanics texts, but in general they only look at the case where the Hamiltonian is time translation invariant, in which case the problem is simple:

$$TU(t_1, t_0)T^{-1} = Te^{-\frac{i}{\hbar}(t_1-t_0)H}T^{-1} = e^{\frac{i}{\hbar}(t_1-(2t_1-t_0))H} = e^{-\frac{i}{\hbar}(t_1-t_0)H} = U(t_1, t_0). \quad (2.35)$$

The literature does not appear to cover the case where the Hamiltonian is time-dependent. If the Hamiltonian has explicit time dependence, so that it does not commute with itself at different times, then the problem is not trivial because the exponential form of the evolution operator becomes shorthand for a Dyson series [157]. The correct application of the time reversal operator to this series and the time ordering operator is not obvious. This can be avoided by adoption of (2.31) as the definition for a time reversal symmetric system. Assuming we have a TRS system

$$\begin{aligned} U(t_1 + \alpha, t_1)\psi(t_1) &= \psi(t_1 + \alpha) \\ TU(t_1 + \alpha, t_1)T^{-1}T\psi(t_1) &= T\psi(t_1 + \alpha) \\ TU(t_1 + \alpha, t_1)T^{-1} &= U(t_1 + \alpha, t_1) \end{aligned} \quad (2.36)$$

using (2.31). Now take  $\alpha = t_0 - t_1$  and  $\alpha = t_1 - t_0$  and insert into (2.36) to get

$$T \exp \left[ \frac{-i}{\hbar} \int_{t_1}^{t_0} H(t) dt \right] T^{-1} = \exp \left[ \frac{-i}{\hbar} \int_{t_1}^{t_0} H(t) dt \right] \quad (2.37)$$

and

$$T \exp \left[ \frac{-i}{\hbar} \int_{t_1}^{2t_1-t_0} H(t) dt \right] T^{-1} = \exp \left[ \frac{-i}{\hbar} \int_{t_1}^{2t_1-t_0} H(t) dt \right]. \quad (2.38)$$

Using these results we can obtain the time reversal transformation of the evolution operator over the entire region of time reversal symmetry,  $[t_0, t_2 = 2t_1 - t_0]$ :

$$\begin{aligned} TU(t_2, t_0)T^{-1} &= TU(t_2, t_1)U(t_1, t_0)T^{-1} \\ &= TU(2t_1 - t_0, t_1)T^{-1}TU(t_1, t_0)T^{-1} \\ &= U(2t_1 - t_0, t_1)U(t_1, t_0) \\ &= U(t_2, t_0). \end{aligned} \quad (2.39)$$

This argument shows that under  $T$  the evolution operator is invariant.

On the other hand, if  $T_{mr}$  is used as the time reversal operator, similar reasoning to the above yields

$$T_{mr} \exp \left[ \frac{-i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] T_{mr}^{-1} = \exp \left[ \frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right]. \quad (2.40)$$

Here there is a slight discrepancy between my result and that of Lee [114]. Lee is one of the few people to consider the time reverse of the evolution operator for a time-dependent Hamiltonian. Lee uses  $T_{mr}$  and obtains

$$T_{mr} U(t, t_0) T_{mr}^{-1} = U(-t, -t_0). \quad (2.41)$$

(2.41) would agree with (2.40) if the condition that  $H(t) = H(-t) \forall t$  is met. It is likely that Lee is assuming the Hamiltonian is TRS for all time, and is choosing the origin of time to be the point about which TRS holds (i.e.  $t_1 = 0$ ). If these conditions are met then the results (2.41) and (2.40) are equivalent.

## 2.4 Ring laser tests of time reversal symmetry violation

A ring laser is an extremely high precision instrument, capable of very sensitive interferometry measurements. It consists of two counter-propagating laser beams constrained to remain in a plane (Figure 2.4). The Canterbury ring laser, for example, has a sensitivity to rotation of  $\Omega \sim 3 \times 10^{-9} \tau^{-1/2}$  where  $\tau$  is the observation time and  $\Omega$  is the rotation rate [176]. Thus a one hour observation could detect a rotation of  $\sim 10^{-10}$  rad/s.

What a ring laser essentially measures is the difference in optical path length between the two counter-propagating beams. The optical path length is in turn given by the refractive index of the medium through which the beam passes. If there exists a nonreciprocal part to the refractive index so that each of the two beams experiences a different index then a beat frequency between the two beams will result.

For example should the ring laser as a whole have some form of net rotation (the Sagnac effect) then the effective beam path one way around the ring will be shorter than the other. This will give rise to a beat frequency [176]

$$f = \frac{4\mathbf{A} \cdot \boldsymbol{\Omega}}{\lambda P} \quad (2.42)$$

where  $\mathbf{A}$  is the oriented area enclosed by the ring,  $\boldsymbol{\Omega}$  is the angular velocity vector of the rotation,  $\lambda$  is the beam wavelength and  $P$  is the perimeter length

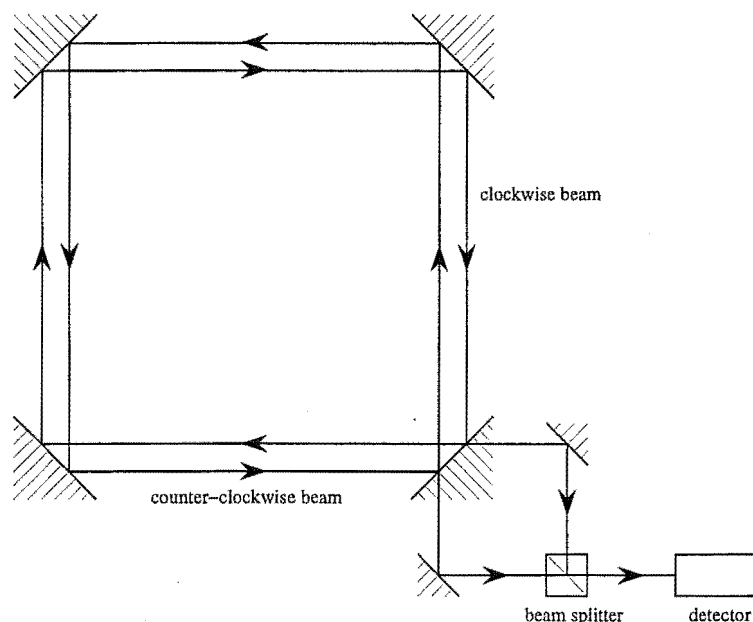


Fig. 2.4: Overhead view of a typical ring laser

of the ring. As the ring is necessarily mounted on a rotating platform (the earth) it is possible to measure the earth's rotation rate with great precision.

Although rotation effects are of great interest and importance, a ring laser has many other possibilities. The existence of nonreciprocal effects immediately suggest one of the more fundamental: symmetry breaking. Due to the high sensitivity of a ring an obvious question is whether it is possible to make tests on the nature of the fundamental time reversal and parity symmetries violated in the weak interaction.

Although both parity and time reversal symmetry violations can give rise to non-reciprocal effects, Stedman and I have shown that the ring laser in its standard planar geometry can fairly generally be said to be sensitive only to time-odd effects [177].

To see this note that the optical path length is  $P_{\pm} = \int n_{\pm} dx$  where  $n_{\pm}$  is the refractive index of the clockwise (CW) and counterclockwise (CCW) beams respectively. In order to obtain a non-reciprocal effect there must exist a term  $\delta n$  in the refractive index of the medium satisfying

$$\delta n(\mathbf{e}_+, \mathbf{k}, \mathbf{e}_-) = -\delta n(\mathbf{e}_-, -\mathbf{k}, \mathbf{e}_+) \quad (2.43)$$

where we denote the polarization of the CW beam by  $\mathbf{e}_+$  and the polarization of the CCW beam by  $\mathbf{e}_-$ , and we have allowed the refractive index to be general, depending on the wavevector of the beam, its polarization and the polarization of the counter-propagating beam. (2.43) demands that opposing beams experience a different refractive index at the same point in the medium. By Lloyd's theorem

each term contributing to the refractive index must be either symmetric or antisymmetric; we ignore the symmetric contributions.

First assume that both the CW and CCW beams have the same ellipticity:

$$\mathbf{e}_+ = \mathbf{e}_*. \quad (2.44)$$

Thus either the beams are linearly polarized or have the same handedness.

Now, time reversal violation is equivalent to the existence of a term of the form

$$\delta n(\mathbf{e}_+, \mathbf{k}, \mathbf{e}_-) = -\delta n(\mathbf{e}_*, -\mathbf{k}, \mathbf{e}_*). \quad (2.45)$$

which follows from considering the effect of the time reversal operator acting on a wave defined by  $(\mathbf{e}, \mathbf{k})$ . We have

$$T\mathbf{e} = \mathbf{e}^* \quad T\mathbf{k} = -\mathbf{k}. \quad (2.46)$$

Thus equation (2.45) states that a time reversed beam will experience a different refractive index to the original beam. The possible existence of a term such as (2.43) can be seen by considering the refractive index within an extended golden rule approach, i.e. application of  $T$  to a product of matrix elements must give just an overall sign [175].

It is now simple to see that if (2.44) holds then (2.45) reduces to (2.43), the condition required to observe an effect. Consequently  $T$  violation is observable in the  $\mathbf{e}_+ = \mathbf{e}_*$  geometry.

Consider now parity violation in this geometry. A  $P$  violating material has a refractive index that depends on handedness:

$$\delta n(\mathbf{e}_+, \mathbf{k}) = \delta n(\mathbf{e}_*, -\mathbf{k}) \neq \delta n(\mathbf{e}_+, -\mathbf{k}). \quad (2.47)$$

With the condition (2.44) this gives  $\delta n(\mathbf{e}_+, \mathbf{k}) = \delta n(\mathbf{e}_-, -\mathbf{k})$  which contradicts (2.43). Thus we cannot detect parity violation alone in this geometry providing beam coupling is ignored.

One other possible geometry (probably the main alternative) is one in which

$$\mathbf{e}_+ = \mathbf{e}_- \neq \mathbf{e}_*. \quad (2.48)$$

In this geometry the counter-propagating beams have different handedness. Thus a right circularly polarized CW beam and a left circularly polarized CCW beam are distinct from the case with a left circularly polarized CW beam and a right circularly polarized CCW beam.

Analogous to (2.45) is the fully general condition that if  $P$  is violated then there exists a term such that

$$\delta n(\mathbf{e}_+, \mathbf{k}, \mathbf{e}_-) = -\delta n(\mathbf{e}_+, -\mathbf{k}, \mathbf{e}_-). \quad (2.49)$$

With the geometry given by (2.48), equation (2.49) ensures that the condition (2.43) is met, allowing experimental detection of parity violation in this geometry.

Finally, combining (2.45) and (2.49) we see that  $PT$  symmetry requires

$$\delta n(\mathbf{e}_+, \mathbf{k}, \mathbf{e}_-) = \delta n(\mathbf{e}_+^*, \mathbf{k}, \mathbf{e}_-^*). \quad (2.50)$$

Thus a RCP CW beam and a LCP CCW beam are degenerate, as are a LCP CW beam and a RCP CCW beam. Thus in the (2.48) geometry  $PT$  violation would break this degeneracy and yield an effect via (2.43) but  $T$  violation would not.

To sum up: The standard geometry for a planar ring laser is given by (2.44) and is thus only capable of detecting  $T$ -violating effects. An alternate geometry (2.48) is capable of detecting  $P$ - and  $PT$ -violating effects but not  $T$ -violating effects.

This conclusion had partially been reached a number of times previously. For example Alekseev *et al.* [11] were one of the first groups to consider the possibility of detecting parity violation due to the weak interaction in atomic transitions using a pair of separately polarized lasers; Zhang *et al.* [208] wrote an account of using a ring laser to measure a (time-odd) external magnetic field; and Kapitunlik *et al.* [100] explicitly stated that a ring laser was sensitive to  $T$ -violating effects and used one to measure Faraday and Kerr effects with great precision.

The fact that in a conventional geometry a ring laser is insensitive to  $P$ -violation demonstrates that some attempts to detect atomic parity violation from the electroweak interaction have been doomed to failure (see for example Elliott and Small [67]): such  $P$ -violating but  $T$ -even couplings as  $\boldsymbol{\sigma}_e \cdot \mathbf{p}_e$  are invisible in such geometries. On the other hand experiments to look for a time-odd nuclear interaction with electronic transitions using ring lasers are quite valid. Kozlov and Porsev [108], for example, look for nonreciprocity in a term of the form  $\mathbf{k} \cdot \mathbf{E}$  using an interferometer that could easily be adapted to use a ring laser. Ring lasers have also been used to study other time-odd phenomena. For example they have been used to test for  $T$  violation in high- $T_c$  superconductors [173] and the study of magneto-optical phenomena such as the Faraday and Kerr effects [100].

### 3. The geometric phase

In quantum mechanics the state of a system is represented by a ket in a Hilbert space. Kets that lie on the same ray in Hilbert space represent the same physical state even though they differ by a phase. This does not mean that the phase is unimportant, however. Although a single phase is unimportant in isolation, we almost always deal with interacting systems where phases can interfere with each other and produce physical consequences. It is therefore of interest to consider the various phases that can arise in physics.

The geometric phase could have been discovered anytime during or after the formulation of quantum mechanics in the 1920s. It was partially anticipated by Herzberg and Longuet-Higgins who demonstrated that in the molecular Born-Oppenheimer approximation electronic wavefunctions became double-valued when the nuclear coordinates described a closed path about a conical intersection [87] and also by Mead and Truhlar who showed that this double-valuedness could be removed at the cost of introducing a vector potential into the nuclear Hamiltonian [127]. Its true geometric and consequently wide-ranging properties were however unknown until the seminal work of Berry [31]. The phase is consequently often known in the adiabatic context as the Berry phase, as distinct from the nonadiabatic case which is sometimes called the Anandan-Zee phase. In this thesis I drop the distinction and consistently refer to the phase as the geometric phase.

Berry demonstrated that a quantum-mechanical state evolving adiabatically in time under a slowly varying parameter-dependent Hamiltonian could acquire a geometric phase factor if the parameters were to return to their initial values after having traversed a closed path. This phase factor is independent of the well-known dynamical phase factor and is nonlocal, or nonintegrable — that is it depends on the geometry of the path as a whole and not merely the endpoints. Following Berry's initial formulation the phase was rapidly generalized in a number of ways, and was demonstrated to have effects in a startlingly large range of applications.

Aside from §3.4 this chapter consists mainly of review material, albeit sometimes with an unusual perspective or emphasis. I consider the forms which the geometric phases can take, and the various formalisms that have been used. In §3.1 I give its simplest derivation, which corresponds to the situation that Berry originally considered — an adiabatically evolving parametric Hamiltonian. In §3.2 I demonstrate one of the most interesting properties of the geometric phase, the fact that in an adiabatic context it can be seen to arise from a gauge potential.

The restriction to adiabatic evolutions, although yielding remarkable gauge structures, can be limiting. The geometric phase was shown to have meaning in nonadiabatic context by Aharonov and Anandan [3], a result which highlighted the geometric nature of the phase. I briefly review their approach in §3.3.

The gauge approach of §3.2 can be taken further. For example although Berry initially considered only non-degenerate states, if degeneracy is introduced it can be shown that the gauge potential becomes nonabelian and acts precisely as a nonabelian gauge field [201], even transforming correctly under gauge transformations. I demonstrate this in §3.4, as well as developing some results I will require later in my consideration of the  $SU(2)$  instanton geometric gauge potential in Chapter 6. I also point out some common errors and inconsistencies about the nonabelian phases and gauge potentials in the literature. To end the section I prove a result of my own which shows in which situations it is possible to reduce the nonabelian case to a set of abelian cases.

Finally in §3.5 I give the mathematical background for the geometric phase in terms of fibre bundle theory. This treatment demonstrates the beautiful topological and geometric nature of the phase, as well as providing links between the formalisms of Berry and Aharonov and Anandan, and links between the phase, parallel transport and gauge theories in general. This section becomes somewhat mathematical, and it is easy to lose track of the physical implications of the formalism that is being constructed. If this is the case, an excellent non-mathematical article covering the relationship between gauge theories and geometry exists [30].

There are many extensions beyond the fundamental approaches I consider in this chapter, highlighting the rich generality of the geometric phase. For example there exists an alternate way of looking at the phase using “geometric quantum angles”, which considers the eigenspace of an observable under general evolution [19] and can be considered locally measurable unlike the usual non-local formulation of the phase [17]. Salmistraro has linked the two formalisms [158].

The geometric phase has also been formulated relativistically in a number of ways. Kuratsuji and Iida expressed it within a path integral approach [110, 111], and it has also been shown an analogue exists for the Klein-Gordon equation rather than the more usual Schrödinger equation [18]. Its role has also been considered in classical field theories [13, 74] and general relativity ([16] and references therein).

There are also interesting analogues that exist in classical systems. Hannay, for example, has explored the phase in classical systems in the context of action angle variables for adiabatically varied Hamiltonians [85], and Cina has considered the case of a magnet in a slowly varying external field [55]. Shapere and Wilczek have considered the phase as it arises in smoothly deformable bodies and fluid mechanics [166, 168]. Similarly Littlejohn and Reinsch have studied

the nonabelian geometric gauge fields that arise in Hamiltonians describing the internal degrees of freedom for systems in classical mechanics [119].

### 3.1 Definition

To understand the geometric phase we must first define a cyclic state. Suppose the Hamiltonian for a system has a periodic time dependence such that

$$H(t + \tau) = H(t). \quad (3.1)$$

An initial eigenvector  $|\psi(0)\rangle$  is a cyclic initial state of  $H$  if

$$|\psi(\tau)\rangle = e^{i\chi}|\psi(0)\rangle, \quad (3.2)$$

that is, if the eigenvector  $|\psi\rangle$  returns to itself up to a phase.

If the Hamiltonian is time-independent then the existence of cyclic initial states is trivial: the solution to the Hamiltonian will consist of stationary states of the form

$$|\psi(t)\rangle = e^{-iEt/\hbar}|\phi\rangle \quad (3.3)$$

where  $|\phi\rangle$  is the solution to the time-independent Schrödinger equation. Thus the state is always periodic – it is always equal to itself up to a phase.

One way to ensure the existence of cyclic states in the time-dependent case is as follows. Consider a Hamiltonian depending on a set of time-dependent external parameters  $x^\mu(t)$ . At any time  $t$  the  $n$ th energy eigenstate  $|n(x(t))\rangle$  of the Hamiltonian satisfies the equation

$$H(x(t))|n(x(t))\rangle = E_n(x(t))|n(x(t))\rangle \quad (3.4)$$

where  $x(t)$  is a vector in the parameter space. From this one can construct a set of basis vectors for  $H$  that cover the entire parameter space. In what follows it is assumed that this set of basis vectors  $|n(x)\rangle$  is single-valued. Now suppose we begin with an eigenstate  $|\psi(x(0))\rangle$  of the Hamiltonian at time  $t = 0$ , and allow it to evolve via the time-dependent Schrödinger equation

$$H(x(t))|\psi(x(t))\rangle = i\hbar\frac{\partial}{\partial t}|\psi(x(t))\rangle. \quad (3.5)$$

We assume that parameters are varied adiabatically so that the system is always in an eigenstate of the Hamiltonian. It should be noted that the assumption that sufficiently slow change will ensure that the state is preserved in an eigenstate

of the Hamiltonian, while valid, is not trivial [26]. Consequently we expect the solution to (3.5) to be of the form

$$|\psi(x(t))\rangle = \exp\left[-\frac{i}{\hbar}\int_0^t E_n(x(t'))dt'\right] \exp[i\gamma_n(t)]|\psi(x(0))\rangle. \quad (3.6)$$

Inserting (3.6) into the Schrödinger equation (3.5) we find that

$$\frac{d}{dt}\gamma_n(t) = i\langle n(x(t))|\partial_\mu n(x(t))\rangle \frac{d}{dt}x^\mu(t) \quad (3.7)$$

and consequently

$$\gamma_n(t) = i\int_{x(t_0)}^{x(t)} \langle n(x(t'))|\partial_\mu n(x(t'))\rangle dx^\mu(t') \quad (3.8)$$

where path of integration is the path of adiabatic changes of  $x$  in parameter space.

Thus adiabatic evolution generates two separate phases: The dynamical phase  $\delta$  which is simply related to the instantaneous energy of the state, and the geometric phase  $\gamma$ .  $\gamma$  is fundamentally different in origin to the dynamical phase. It is non-integrable and purely geometric in origin — it depends only on the path taken in parameter space; the speed or energy with which it traverses this path is irrelevant. It can also be experimentally measured independently of the dynamical phase [195].

As can be seen, the geometric phase arises from very simple quantum mechanics. The question must be asked: Why were its properties not investigated until the seminal work of Berry [31] recognized its fundamental geometric nature? Probably the most important reason can be traced to Fock in 1928 [71]. Fock demonstrated that the extra phase  $\gamma$  can be transformed to unity under a phase gauge transformation of the form  $|n(x)\rangle \rightarrow e^{i\eta(x)}|n(x)\rangle$ . Fock was correct up to a point. However he did not consider *cyclic* evolution, and his proof fails for this situation. In this case the extra phase cannot be gauged away [31].

The existence of this phase is experimentally well established over a wide range of physics. The first direct experimental observation was by Tomita and Chiao who observed a phase shift in a beam of polarized light propagating through an optical fibre [187]. Since then it has been observed in a wide range of systems ranging from the classical (precession of Foucault's pendulum [85]), to the optical [36], to the purely quantum (shifts in the energy spectrum of the Na<sub>3</sub> [59]).

## 3.2 The geometric gauge potential

Part of the interest and, perhaps more importantly, usefulness of the geometric phase lies in the way in which it can be recast in a gauge field theory, provided

one remains in the adiabatic approximation. To see this we drop the explicit time-dependence in the parameters and rewrite (3.8) as

$$\gamma = \int_{x_0}^{x_1} A_\mu dx^\mu \quad (3.9)$$

where

$$A_\mu = \langle n(x) | i\partial_\mu | n(x) \rangle. \quad (3.10)$$

The vector field  $A_\mu$  in the equations above bears a striking resemblance to the electromagnetic vector potential, specifically with regard to altering the wavefunctions by phases. That is, in electromagnetic quantum theory a vector potential modifies the phase of the wavefunction of a charged particle [157]:

$$|\psi\rangle \rightarrow \exp \left[ i \frac{q}{\hbar} \int A_\mu^{\text{em}} dx^\mu \right] |\psi\rangle. \quad (3.11)$$

It is this behaviour that gives rise to the celebrated Aharonov-Bohm effect (Figure 3.1), and was the first demonstration that it is gauge potentials rather than their associated field strengths that are the truly fundamental entities [4]. One of the most interesting things about the Aharonov-Bohm effect is that it is non-local — the wavefunction of the electron can be affected even if the field is zero along its entire path. This idea carries over to the geometric phase. It is also a non-local effect, and is non-integrable, that is requiring knowledge of the entire path rather than just the endpoints. In fact, it is actually possible to interpret the Aharonov-Bohm effect in terms of the geometric phase [14, 31]. I will say more about the non-local and hence geometric nature of the phase in §3.3.

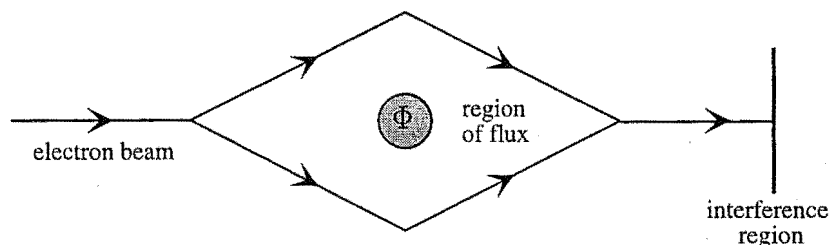


Fig. 3.1: The wavefunction of charged particles acquire a phase shift if their path encloses a region of magnetic flux.

The analogy between the geometric and electromagnetic vector potentials has considerable depth. For example we can construct a geometric flux analogous to the magnetic flux in order to calculate (3.9) via a surface integral rather than a path integral, provided we are dealing with a three dimensional parameter space

$\mathbf{R}$  and the path in parameter space is closed. Noting that  $\nabla \times (f \nabla g) = (\nabla f \times \nabla g)$  we have

$$\begin{aligned} \gamma_n &= i \oint \langle n(\mathbf{R}) | \nabla n(\mathbf{R}) \rangle \cdot d\mathbf{R} \\ &= - \iint \mathbf{B}_n^{\text{geo}} \cdot d\mathbf{S} \end{aligned} \quad (3.12)$$

where

$$\begin{aligned} \mathbf{B}_n^{\text{geo}} &= \nabla \times \langle n(\mathbf{R}) | \nabla | n(\mathbf{R}) \rangle \\ &= \Im \langle \nabla n(\mathbf{R}) | \times | \nabla n(\mathbf{R}) \rangle \\ &= \Im \sum_{n \neq m} \langle \nabla n(\mathbf{R}) | m(\mathbf{R}) \rangle \times \langle m(\mathbf{R}) | \nabla n(\mathbf{R}) \rangle \\ &= \Im \sum_{n \neq m} \frac{\langle n(\mathbf{R}) | \nabla H | m(\mathbf{R}) \rangle \times \langle m(\mathbf{R}) | \nabla H | n(\mathbf{R}) \rangle}{(E_m - E_n)^2} \end{aligned} \quad (3.13)$$

and  $H$  is the Hamiltonian. The advantage of using (3.13) over (3.12) is that it is not necessary to choose a locally single-valued basis set [31]. Any basis set satisfying (3.4) will suffice.

If the parameter space is greater than three-dimensional it is still possible to construct a geometric flux analogue using the generalized Stokes' theorem and the theory of differential forms [51].

To strengthen the analogy between the geometric potential and the electromagnetic potential we examine how the geometric potential changes under a rephasing of the basis vectors. We allow the rephasing to be local (that is, dependent on position) so that

$$|n(x)\rangle \rightarrow e^{i\eta(x)} |n(x)\rangle. \quad (3.14)$$

Under this transformation we find that

$$A_\mu^{\text{geo}} \rightarrow A'_\mu{}^{\text{geo}} = A_\mu^{\text{geo}} + i\partial_\mu \eta. \quad (3.15)$$

Consequently the geometric potential behaves just as if it were a U(1) gauge field, once again exactly analogous to the electromagnetic vector potential.

### 3.3 The nonadiabatic geometric phase

So far I have considered only adiabatic evolution. Geometric phases can be defined for nonadiabatic evolutions too, and in many ways considering such an evolution makes the true geometric nature of the phase more apparent. This was first shown by Aharonov and Anandan [3]. They demonstrated that for a given

cyclic evolution, not necessarily adiabatic, there was associated a particular phase. This phase is geometric in the sense that it is identical for all the curves in the Hilbert space  $\mathcal{H}$  which project to the same closed curve in the projective Hilbert space  $\mathcal{P}$ , that is the space of rays of  $\mathcal{H}$ . This is shown pictorially in Figure 3.2. Although points  $p$  and  $q$  are distinct in  $\mathcal{H}$ , they lie on the same ray and are related by a  $U(1)$  phase transformation. In §3.5 I will show that this transformation corresponds to the group operation, in this case an element of  $U(1)$ , on a fibre bundle. This phase is also completely independent of the Hamiltonian which evolves the system through the curve in  $\mathcal{H}$ . To show this I follow Aharonov and Anandan's original argument.

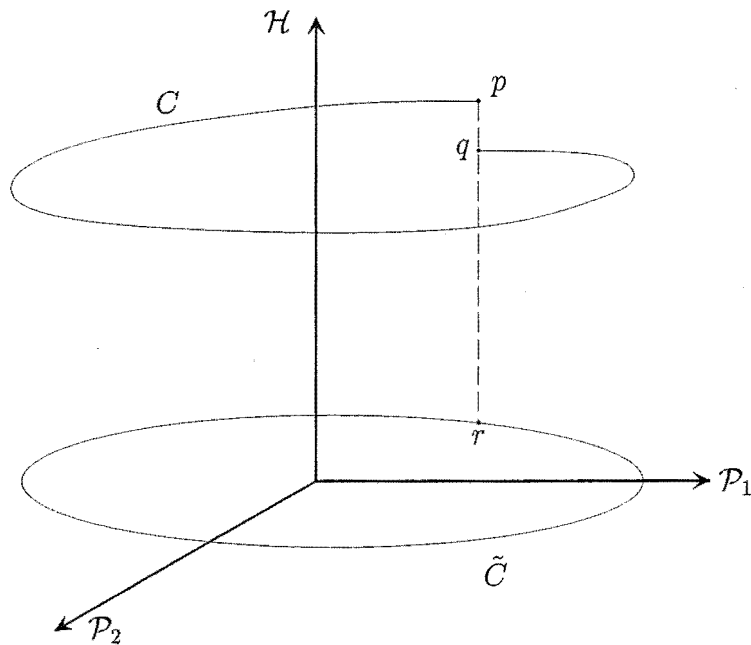


Fig. 3.2: An open loop in a Hilbert space  $\mathcal{H}$  can project to a closed loop in its ray space  $\mathcal{P}$ .

Suppose that the normalized state  $|\psi(t)\rangle \in \mathcal{H}$  evolves according to the Schrödinger equation

$$H(t)|\psi(t)\rangle = i\hbar \frac{d}{dt}|\psi(t)\rangle \quad (3.16)$$

such that  $|\psi(\tau)\rangle = e^{i\phi}|\psi(0)\rangle$ , with  $\phi$  real. Now define  $|\tilde{\psi}(t)\rangle = e^{-if(t)}|\psi(t)\rangle$  such that  $f(\tau) - f(0) = \phi$ . Then  $|\tilde{\psi}(\tau)\rangle = |\tilde{\psi}(0)\rangle$  and from Eq. (3.16)

$$-\frac{df}{dt} = \frac{1}{\hbar} \langle \psi(t) | H | \psi(t) \rangle - \langle \tilde{\psi}(t) | i \frac{d}{dt} | \tilde{\psi}(t) \rangle. \quad (3.17)$$

Now, if the dynamical part of the phase is removed by defining

$$\gamma = \phi + \frac{1}{\hbar} \int_0^\tau \langle \psi(t) | H | \psi(t) \rangle dt \quad (3.18)$$

it follows that

$$\gamma = \int_0^\tau \langle \tilde{\psi} | i \frac{d}{dt} | \tilde{\psi} \rangle dt. \quad (3.19)$$

If  $C$  is the path taken in Hilbert space, and  $\tilde{C}$  is the path taken in projective Hilbert space, then the same  $|\tilde{\psi}(t)\rangle$  can be chosen for every curve  $C$  for which  $\Pi(C) = \tilde{C}$ , by appropriate choice of  $f(t)$ . Hence  $\gamma$  is independent of  $\psi$  and  $H$  for a given closed curve  $\tilde{C}$ . In fact, for a given  $\tilde{C}$ ,  $H(t)$  can be chosen such that the second term in Eq. (3.18) is zero, which may be regarded as an alternative definition of  $\gamma$ . Also, from Eq. (3.19),  $\gamma$  is independent of the parameter  $t$  of  $\tilde{C}$  and is uniquely defined up to  $2\pi n$ . Hence  $e^{i\gamma}$  is a geometric property of the unparameterized image of  $\tilde{C}$  only.

This argument was soon extended to the case where the state being transported is degenerate giving rise to the nonabelian nonadiabatic geometric phase [14]. It is possible to generalize still further. Samuel and Bhandari showed that the geometric phase was still meaningful in evolutions that were not cyclic or even unitary [160].

The relation between the adiabatic and nonadiabatic formalisms is very strong, but they are not identical. The former is much closer to an experimental viewpoint where the experimenter has direct control over the evolution, and the latter, while more general and avoiding the assumption of adiabaticity, is not directly controllable by the experimenter. The relation between the two formalisms has been discussed by Bohm and Mostafazadeh using the classifying theorem for principal fibre bundles [44].

For most of the rest of this thesis I will concentrate on the adiabatic version of the formalism. This allows an easier gauge field interpretation of the phase, and this is more useful when I consider the gauge fields generated by time reversal symmetry, and by time-odd molecular systems.

### 3.4 The nonabelian geometric phase

In the previous sections it was assumed that the state being adiabatically transported was not degenerate. The problem of transporting a degenerate state was first considered by Wilczek and Zee who showed that this will lead to the geometric gauge potential becoming nonabelian [201].

Suppose the solution to a parameter-dependent Hamiltonian yields a degenerate space of  $n$  states. Letting  $H(x(t))$  represent the Hamiltonian depending on the set of parameters  $x^\mu$  as in the previous section, we once again assume that the parameters describe a closed adiabatic path such that  $x^\mu(0) = x^\mu(\tau)$ . Given that the  $n$  degenerate levels must be mapped back on themselves after the evolution, the natural question is whether this mapping is a trivial transformation. The

answer is that it is not. This is seen by letting a set of single-valued states be given at each point  $x^\mu$  by

$$\bar{H}(x(t))|\tilde{\psi}_i(t)\rangle = E|\tilde{\psi}_i(t)\rangle. \quad (3.20)$$

where the tilde denotes a single-valued basis and will continue to do so in the rest of this thesis. Assuming we have initial conditions  $|\psi_i(0)\rangle = |\tilde{\psi}_i(0)\rangle$  we allow the initial space to evolve via the time-dependent Schrödinger equation

$$H(x(t))|\psi_i(t)\rangle = i\hbar \frac{d|\psi_i(t)\rangle}{dt} \quad (3.21)$$

and write the solution as

$$|\psi_i(t)\rangle = U_{ij}|\tilde{\psi}_j(t)\rangle. \quad (3.22)$$

Substituting (3.22) into (3.21) and premultiplying by  $\langle\tilde{\psi}_k|$  and  $U^{-1}$  we obtain

$$\begin{aligned} i\hbar(U^{-1})_{li}\dot{U}_{ij}\langle\tilde{\psi}_k|\tilde{\psi}_j\rangle + i\hbar(U^{-1})_{li}U_{ij}\langle\tilde{\psi}_k|\dot{\tilde{\psi}}_j\rangle &= E(U^{-1})_{li}U_{ij}\langle\tilde{\psi}_k|\tilde{\psi}_j\rangle \\ (U^{-1}\dot{U})_{lj}\delta_{kj} + \delta_{lj}\langle\tilde{\psi}_k|\dot{\tilde{\psi}}_j\rangle &= \frac{-iE}{\hbar}\delta_{lj}\delta_{kj} \\ (U^{-1}\dot{U})_{lk} &= \frac{-iE}{\hbar}\delta_{lk} - \langle\tilde{\psi}_k|\frac{d}{dt}|\tilde{\psi}_l\rangle. \end{aligned} \quad (3.23)$$

This is solved in terms of path ordered integrals by

$$U_{ij} = P \exp\left[-\frac{iE}{\hbar}\delta_{ij} + iA_{ij}\right] \quad (3.24)$$

where

$$A_{ij} = \langle\tilde{\psi}_j|i\frac{d}{dt}|\tilde{\psi}_i\rangle. \quad (3.25)$$

This is completely analogous to (3.9) and (3.10) for the abelian case.

I have written out the above derivation in detail as my result differs from the original result of Wilczek and Zee [201]. Their definition of  $U_{ij}$  differs by a sign. That is, choosing the energy  $E = 0$  they have

$$U_{ij}^{WZ} = P \exp\left[-i\langle\tilde{\psi}_j|i\frac{d}{dt}|\tilde{\psi}_i\rangle\right]. \quad (3.26)$$

This can be seen to be incorrect by simplifying the degenerate space to a single level and thus obtaining the abelian case. Using the result of Wilczek and Zee the phase factor becomes

$$U^{WZ} = \exp\left[-i\langle\tilde{\psi}|i\frac{d}{dt}|\tilde{\psi}\rangle\right] \quad (3.27)$$

which is in direct contradiction to equations (3.6) and (3.9). This error is certainly of experimental importance, resulting in a different phase matrix whenever the phases are not multiples of  $\pi$ .

Incorrect definitions of the phase matrix  $U_{ij}$  are common throughout the literature. The problem is compounded due to a lack of agreement as to how the gauge potential matrix  $A_{ij}$  is defined, and how the phase matrix  $U_{ij}$  is defined in terms of  $A_{ij}$ . For example it is common to define the phase matrix via right multiplication, or, employing the summation convention

$$|\psi_i(t)\rangle = U_{ji}|\tilde{\psi}_j(t)\rangle. \quad (3.28)$$

The advantage with this approach is that it enables a logical indexing of the vector potential, i.e.  $A_{ij} = \langle \tilde{\psi}_i | i \partial_t | \tilde{\psi}_j \rangle$  which is to be contrasted with (3.25). This scheme loses the obvious phase-factor-as-operator interpretation, however. It appears to be the standard choice when considering the geometric vector potential arising in the Born-Oppenheimer approximation in molecular physics. In addition, different definitions of the vector potential  $A_{ij}$  exist. For example, mathematical physicists often define  $A_{ij} = \langle \tilde{\psi}_i | \partial_t | \tilde{\psi}_j \rangle$  which omits the factor of  $i$ . This is due to the fact that when the vector potential is considered from a topological point of view it can be seen to be a connection on a fibre bundle. Extra constants such as  $i$  and  $\hbar$  arise in the assigning of physical reality in the form of quantum mechanics to pure topological results, and when analysing a problem topologically it is simpler to define  $A_{ij}$  to be exactly the topological object corresponding to a connection. This is quite valid provided the definition of the phase factor is then changed also.

This cloud of different definitions has led to confusion and outright error even after the different definitions are allowed for. In addition to Wilczek and Zee errors of this nature can be found in Apsel *et. al.* [22], Choi *et. al.* [53] and Ihm [93], for example.

It is natural to ask what happens to the nonabelian gauge potential if the adiabatic basis (3.20) is reshuffled by a unitary transformation. Since any basis is as good as any other, we expect such a transformation not to alter any observable physics. If we define

$$A_{ij\mu} = \langle \psi_j | i \partial_\mu | \psi_i \rangle \quad \text{and} \quad |\psi'_i\rangle = \Lambda_{ik} |\psi_k\rangle \quad (3.29)$$

then one finds

$$\begin{aligned} A_{ij\mu} \rightarrow A'_{ij\mu} &= \langle \psi'_j | i \partial_\mu | \psi'_i \rangle \\ &= \Lambda_{ik} A_{kl\mu} \Lambda_{lj}^\dagger + i \partial_\mu \Lambda_{ik} \Lambda_{kj}^\dagger \\ &= \Lambda A_\mu \Lambda^\dagger + i \partial_\mu \Lambda \Lambda^\dagger \end{aligned} \quad (3.30)$$

in matrix form. Thus the geometric potential transforms exactly as is expected of a nonabelian gauge field under a gauge transformation — a change of basis is equivalent to a gauge transformation, and both are physically unobservable. It should be noted that if one chooses the right multiplication definition of the phase matrix (equation (3.28)) and the associated definition of the geometric

potential  $A_{ij\mu} = \langle \tilde{\psi}_i | i\partial_\mu | \tilde{\psi}_j \rangle$  then for consistency one is forced to have a right-multiplication definition of the basis change also:  $|\psi'_i\rangle = |\psi_k\rangle \Lambda_{ki}$ . If this is the case then the effect of the basis change on  $A_\mu$  becomes

$$A_\mu \rightarrow A'_\mu = \Lambda^\dagger A_\mu \Lambda + i\Lambda^\dagger \partial_\mu \Lambda. \quad (3.31)$$

This is not a normal form for a gauge transformation, but if we invert the transformation and write  $A_\mu$  as a function of  $A'_\mu$  it can be seen to be a valid transformation by considering  $\Lambda$  to be the inverse element of the gauge group, that is  $\Lambda = \Lambda'^\dagger$ . These two transformation rules will be useful when I consider the general Hamiltonian which generates an SU(2) instanton potential in Chapter 6.

The abelian geometric phase can only change the phase of a quantum state and thus is often difficult to measure unless it can be recombined with a reference state to produce interference. The nonabelian phase on the other hand can change the expectation value of a physical observable [161], and consequently one would expect experimental tests to be straightforward. Several experimental schemes to detect the nonabelian geometric phase have been put forward. Segert has suggested considering normally forbidden transitions of optically pumped Pb<sup>208</sup> with accidental degeneracies created via external collinear electric and magnetic fields [162]. Mead has suggested examination of an atom with an odd number of electrons in an external electric field so that the angular momentum states are split by the Stark Hamiltonian [125]. He showed how the nonabelian phase could mix states resulting in a sudden reversal of the component of angular momentum parallel to the electric field as it was slowly rotated. Choi and Ji have considered a very similar situation although in more detail, considering excited hydrogenic atoms to second order perturbation effects [53]. Zee has shown how Tycko's measurement of the abelian phase in nuclear quadrupole resonance [189] can, with sufficient cunning, be extended to measure the effect of a nonabelian phase in the same system [206].

Despite the plethora of experimental schemes that exist, to my knowledge there has only been one experiment performed to actually measure the nonabelian phase. Zwaniger *et al.* [212] considered the nuclear quadrupole resonance spectrum of <sup>35</sup>Cl in a single crystal of sodium chlorate. Their experiment was very similar to that of Tycko, except that as Zee suggested they rotated the crystal around two independent axes, the anti-commutative nature of the rotations giving rise to a nonabelian phase. The results they obtained were consistent with a nonabelian phase.

This lack of explicit experimental verification has not stopped theorists exploring the consequences, however. It is impossible to even begin to cite a representative set of references, but by way of example it has been considered in such diverse fields such as molecular chemistry [42, 126], collision and scattering theory [101, 213], electron and nuclear spin-rotation interaction for paramagnetic and diamagnetic molecules [163, 164], vortex forces in Fermi superfluids [190], and elementary particle physics [113].

The nonadiabatic nonabelian geometric phase is a little more complex, but similar in many essentials. I give a brief overview of the approach of Anandan [14] before extending his work to consider when the full nonabelian formalism is really necessary. To follow Anandan we consider an  $n$ -dimensional subspace  $V_n(t)$  of the Hilbert space  $\mathcal{H}(t)$ .  $V_n(t)$  is defined to be the subspace of vectors obtained by applying the Schrödinger equation to the vectors in  $V_n(0)$  over the time interval  $[0, t]$ . Now, if  $V_n(0) = V_n(\tau)$  for some  $\tau$  we call the subspace cyclic. We choose an orthonormal basis set  $|\psi_i(t)\rangle$  of  $V_n(t)$  such that  $|\tilde{\psi}_i(\tau)\rangle = |\tilde{\psi}_i(0)\rangle$  for all  $i$ . Now choose for  $V_n(0)$  a basis set  $|\psi_i(0)\rangle$ . Evolve these states via the Schrödinger equation and denote the resulting basis for  $V_n(t)$  by  $|\psi_i(t)\rangle$ . Then

$$|\psi_i(t)\rangle = \sum_{j=1}^n U_{ji}(t) |\tilde{\psi}_j(t)\rangle. \quad (3.32)$$

$U$  is a unitary matrix, and is given by [14] as

$$U_{ij}(t) = P \exp \left[ \int_{x(0)}^{x(t)} i(A_{ij} - K_{ij}) dt \right] \quad (3.33)$$

where

$$A_{ij} = i \langle \tilde{\psi}_i | \frac{d}{dt} | \tilde{\psi}_j \rangle \quad (3.34)$$

$$K_{ij} = \frac{1}{\hbar} \langle \tilde{\psi}_i | H | \tilde{\psi}_j \rangle. \quad (3.35)$$

This is very similar to the adiabatic case given by (3.24). The matrix  $K$  corresponds to the instantaneous energy part of the phase and the matrix  $A$  to the geometric gauge potential in the abelian case.

With this background we turn to consider a natural question: in what circumstances is it necessary to utilize the full nonabelian formalism, and when will the abelian approach suffice? The answer is that the problem can be reduced to the abelian case precisely when a basis of cyclic initial states can be found for the subspace  $V_n(t)$ . I now prove this result.

A cyclic initial state is a state which periodically returns to the same state vector in Hilbert space, up to a phase. That is,  $|n(t)\rangle$  is cyclic if  $|n(\tau)\rangle = e^{i\phi} |n(0)\rangle$ . One can ensure the existence of cyclic states if the evolution is adiabatic and the Hamiltonian is parameter dependent. If the path in parameter space is closed then the eigenstates of the instantaneous Hamiltonian will be cyclic. The picture is not so clear cut in the nonadiabatic case.

First, note that if  $U_{ij}(\tau)$  is diagonal then the  $|\psi_i\rangle$  in (3.32) must necessarily be cyclic:

$$|\psi_j(\tau)\rangle = U_{ij}(\tau) |\tilde{\psi}_i(\tau)\rangle = U_{ij}(\tau) |\psi_i(0)\rangle = e^{i\phi_j} |\psi_j(0)\rangle. \quad (3.36)$$

The converse is also true. If the  $|\psi_i\rangle$  are cyclic the  $|\tilde{\psi}_i\rangle$  basis can be chosen proportional to the  $|\psi_i\rangle$ , i.e. choose

$$|\psi_i(t)\rangle = e^{i\phi_i(t)}|\tilde{\psi}_i(t)\rangle. \quad (3.37)$$

Then

$$e^{i\phi_j(t)}|\tilde{\psi}_j(t)\rangle = |\psi_j(t)\rangle = U_{ij}(t)|\tilde{\psi}_i(t)\rangle \quad (3.38)$$

showing  $U(t)$  must be diagonal due to the orthogonality of the  $|\tilde{\psi}_i\rangle$ . If  $U(t)$  is diagonal we see that there is no state mixing during the evolution; each state simply acquires a path-dependent phase factor on the return of the subspace  $V_n$  to itself. Consequently an abelian approach for each individual level similar to [3] can be used.

With the identification of the abelian case with cyclic initial states the next question is: when does a basis of cyclic initial states exist? Again, the answer is simple — they exist if the evolution operator is diagonalizable, that is, has a complete set of eigenvectors. This immediately gives the existence of cyclic states: if the evolution operator (here unconventionally denoted  $\mathcal{E}$  to avoid confusion with  $U_{ij}$ ) generating an evolution over an interval  $[0, \tau]$  has eigenvectors  $|\phi_i(0)\rangle$  then

$$\mathcal{E}(\tau, 0)|\phi_i(0)\rangle = e^{i\phi_i}|\phi_i(0)\rangle \quad (3.39)$$

demonstrating that the  $|\phi_i\rangle$  are cyclic. Note that the associated eigenvalue has modulus one since  $\mathcal{E}$  is unitary.

One way of assuring a complete set of eigenvectors for  $\mathcal{E}$  is for it to act within a finite-dimensional space — any finite dimensional unitary matrix has a complete set of orthogonal eigenvectors [98]. Consequently, if an infinite dimensional state space is reducible into a countable infinity of finite dimensional spaces where states remain within their finite subspace during their evolution, it is possible to consider each subspace as a separate evolutionary entity within which its evolution operator can be diagonalized.

An infinite dimensional space is reducible in this way if there exists an operator  $Q$ , with an associated finite dimensional eigenspace  $\mathcal{Q}_n(t)$ , that commutes with the Hamiltonian  $H(t)$  for all time. As  $Q$  and  $H$  commute it is possible to find a set of basis states that are eigenvectors of both  $Q$  and  $H$ . Note that if  $H$  does not commute with itself at different times then the basis must be time dependent. Let  $|\phi_i(t)\rangle$  be such a basis for  $\mathcal{Q}_n(t)$ , and let  $|\Phi(t)\rangle = \sum_i a_i(t)|\phi_i(t)\rangle$  be some arbitrary vector in  $\mathcal{Q}_n(t)$ . Then

$$H(t)|\Phi(t)\rangle = H(t) \sum_i a_i(t)|\phi_i(t)\rangle = \sum_i E_i(t)a_i(t)|\phi_i(t)\rangle \quad (3.40)$$

demonstrating that vectors in  $\mathcal{Q}_n$  are invariant under  $H$ , and hence stay in  $\mathcal{Q}_n$  during their entire evolution. It is now possible to choose a basis for  $\mathcal{Q}_n$  in which the evolution operator, now considered to be acting only within  $\mathcal{Q}_n$ , is diagonal.

Note that even if the space  $\mathcal{Q}_n$  is not finite dimensional it is still reducible in this way provided that the initial state is a combination of at most a finite number of states of  $\mathcal{Q}_n$ .

This approach will prove useful in §4.1 where I take the commuting operator  $Q$  to be the time reversal operator  $T$  to derive an example of a time reversal selection rule.

Finally, note that even though it may be *possible* to find a commuting operator  $Q$  and construct a cyclic basis in which  $U$  can be made diagonal, it may not be *convenient*. For example, although considering Kramer's doublets, Tycko [189] chooses a basis within each degenerate level so that  $S_z$  (spin component) is diagonalized, and Mead [126] uses the basis  $T|\psi_1\rangle = |\psi_2\rangle$ ,  $T|\psi_2\rangle = -|\psi_1\rangle$ , neither of which lead to cyclic states.

### 3.5 Fibre bundles and the geometric phase

It is often the case that new branches of physics suddenly make use of previously abstract areas of mathematics. This is what Wigner described as the “unreasonable effectiveness of mathematics in the physical sciences” and includes, for example, the differential geometry created by Riemann subsequently proving to be the perfect description of space *a la* general relativity. (Of course, rarely, this situation can be reversed. Take, for example, modern superstring theory. It is a wonderful theory — but one in which the required mathematical tools are probably years in the future.) Another example is the concept of a fibre bundle — an object created by mathematicians long before it was realized that it precisely described gauge theories in physics. More specifically, the gauge fields so familiar to physicists are exactly the connections on fibre bundles [170].

Given the topological nature of the geometric phase, along with its similarity to a gauge field, it is only natural to ask whether it too can be described in terms of fibre bundle theory. It turns out that the answer is in the affirmative, and it is in this formalism that the full beauty and generality of the geometric phase is realized.

First I will consider the closely related link between parallel transport and the geometric phase. If a vector is parallel transported around a closed curve on some surface (that is, require that the vector moves without a change in its magnitude and without rotation about the instantaneous normal to the surface), then it is possible that it will return to its starting point rotated by some amount. This is known as a holonomy transformation. This means that even though the vector is transported without any rotation locally, the global properties of the surface can be such that it can end up rotated. This rotation is a consequence of intrinsic curvature of the manifold. The angle of rotation is governed purely by the curvature of the surface and the area enclosed by the path. For example

on a sphere the angle through which a vector is rotated upon being parallel transported around a closed loop is simply given by the area enclosed by path, or equivalently the solid angle subtended by the path.

This concept has a natural analogue with kets in a Hilbert space. Consider the path  $C$  traced out by an evolving ket through projective Hilbert space. Parameterize  $C$  by  $s$ , and let  $|\psi(s)\rangle$  be the vector corresponding to  $C(s)$ . Parallel transport of the vector  $|\psi(s)\rangle$  is accomplished by assuming that the magnitude of  $|\psi(s)\rangle$  is constant, and that  $|\psi(s)\rangle$  and  $|\psi(s+ds)\rangle$  have the same phase (that is,  $\langle\psi(s)|\psi(s+ds)\rangle$  is real and positive, analogous to the idea of no rotation). These conditions imply [15]

$$\langle\psi(s)|\frac{d}{ds}|\psi(s)\rangle = 0. \quad (3.41)$$

A similar condition was obtained by Simon [169]. Equation (3.41) is just a rule for calculating the phase/angle of a ket/vector at some point along a path, that is, a rule as to how the phase changes with respect to the path. Such a rule is called a connection, and the rule I have just described is called the parallel connection. The parallel connection is in many ways the most natural connection to use as it singles out the changes in phase that are purely geometric in nature. Transporting a ket around a closed path using this rule will result in a phase change when it returns to its starting point and is compared with the untransported ket, and this phase is exactly the same as the one derived in §3.3. There are other connections that can be used, for example the Schrödinger connection (change in phase rule is given by the Schrödinger equation) or the single-valued connection [126].

This idea of holonomy, parallel transport and connections is the language of fibre bundle theory, which I will now explain in more detail.

A fibre bundle is a topological space that is locally, but not globally, a product of two other spaces. To see this let  $E$  and  $X$  be differentiable manifolds and  $\Pi : E \rightarrow X$  be a projection. Then  $(E, X, \Pi)$  is called a bundle, with  $E$  called the total space and  $X$  the base space. We now restrict our attention to those bundles for which  $\Pi^{-1}(x)$  is homeomorphic to some manifold  $F$  for all  $x$ .  $F$  is called the typical fibre.

$E$  is obviously related to  $X \times F$  in some way, and it is the pinning down of this relationship that leads to the definition of the fibre bundle. Requiring that  $E = X \times F$  is overly restrictive so we instead require  $E$  to only locally be a product. A useful and workable definition can then be shown to be as follows [58]: A fibre bundle  $(E, \Pi, F, G, X)$  is defined by the following requirements:

- A topological space  $E$  (the total space).
- A topological space  $X$  (the base space), and a projection  $\Pi : E \rightarrow X$  of  $E$  onto  $X$ .

- A topological space  $F$  called the fibre.
- A Lie group  $G$  of homeomorphisms of the fibre  $F$ .
- A set of open coordinate patches  $U_i$  covering  $X$ , associating with each  $U_i$  a homeomorphism

$$\phi_i : \Pi^{-1}(U_i) \rightarrow U_i \times F \quad (3.42)$$

where  $\phi_i^{-1}$  satisfies  $\Pi\phi_i^{-1}(x, f) = x$  with  $x \in U_i$ ,  $f \in F$ .

The group  $G$  arises in considering the transition from one set of local coordinates given by  $\phi_\alpha, U_\alpha$ , say, to another set given by  $\phi_\beta, U_\beta$ . Suppose  $U_\alpha$  and  $U_\beta$  overlap. Then  $\phi_\alpha \circ \phi_\beta^{-1}$  is a continuous invertible map of the form:

$$\phi_\alpha \circ \phi_\beta^{-1} : (U_\alpha \cap U_\beta) \times F \rightarrow (U_\alpha \cap U_\beta) \times F. \quad (3.43)$$

We let  $x \in U_\alpha \cap U_\beta$  and  $f \in F$ , then fix  $x$  and only allow  $f$  to vary. Now the map  $\phi_\alpha \circ \phi_\beta^{-1}$  for fixed  $x$  is just a map from  $F$  to  $F$ . We denote this map by  $g_{\alpha\beta}$ , called a transition function, and it is the homeomorphism on the fibre  $F$ . The set of all these homeomorphisms for all choices of  $U_\alpha, \phi_\alpha$  form the group,  $G$ .  $G$  is called the structure group of the fibre bundle.

This definition is one only a mathematician could love, although it can be shown to correspond to more intuitive concepts (Nash and Sen give good examples [135]). These ideas are shown in Figure 3.3, where I demonstrate the above definition using a Möbius strip along with its corresponding base space, the circle  $S^1$ . Note that the group  $G$  in this case is given by the two elements  $\{e, g\}$  where  $e$  is the identity and  $g^2 = e$ .

The purpose and use of this definition will hopefully become clearer when we consider its application to the geometric phase.

One particular type of fibre bundle can be constructed by taking  $E = X \times G$  so that the typical fibre  $F$  and the group  $G$  are identical. In such a case the fibre bundle is called the principal fibre bundle.

Consider a path  $C_X$  in the base space of a principal fibre bundle. Since  $\Pi$  is a projection there are many paths  $C_E$  (called lifts) in the total space  $E$  that will map to  $C_X$  under  $\Pi$ . To proceed further we need a method of picking out a particular lift, and to do this we need to be able to compare points on different fibres. This is accomplished with the definition of a connection. It is convenient (but not necessary) to define connection in terms of the tangent vectors to paths in  $E$  [58]. Let  $T_u(E)$  be the tangent space at  $u \in E$ . A subspace of  $T_u(E)$  is those vectors which are only tangent to the fibre passing through  $u$ . This subspace is called the vertical subspace at  $u$  and is denoted  $V_u(E)$ . Together with this vertical subspace  $V_u(E)$  there is a horizontal subspace  $H_u(E)$  defined such that

$$T_u(E) = V_u(E) \oplus H_u(E) \quad (3.44)$$

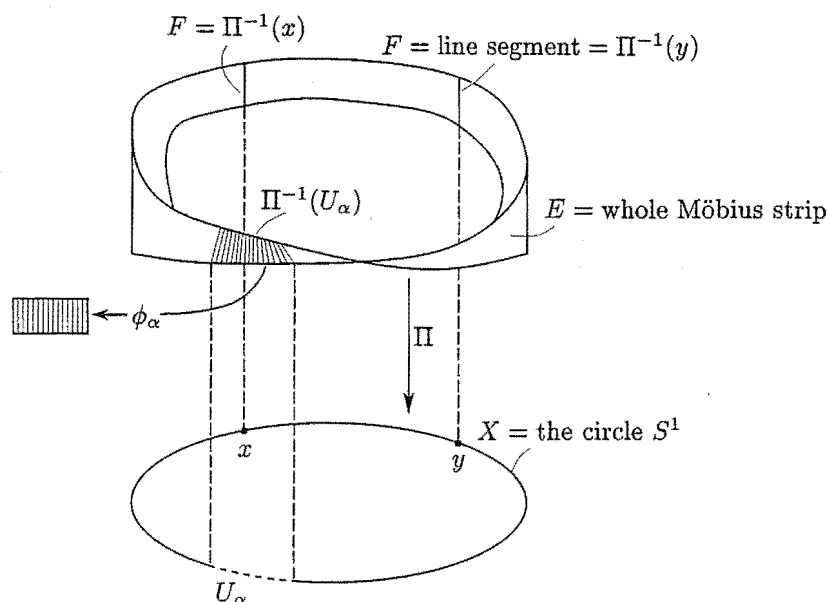


Fig. 3.3: Example of fibre bundles defined on a Möbius strip

Equation (3.44) does not uniquely define  $H_u(E)$ . For example, if  $T_u(E) = \mathbb{R}^2$  and  $V_u(E)$  is the vector  $(0, 1)$  then the only restriction on the basis vector of  $H_u(E)$  is that it cannot be collinear with  $(0, 1)$ . In this case an obvious choice of connection is the orthogonal subspace, i.e. we take  $H_u(E)$  to be  $(1, 0)$ . This is called the natural connection and can be defined whenever the tangent space has an inner product defined on it. The lift defined in this way is called the horizontal lift and can be shown to be identical to the idea of parallel transport [135]. Consequently this connection is also called the parallel connection. This explains why the Christoffel symbols in general relativity are considered as connections.

The last concept we need to define before we can apply the language of fibre bundles to the geometric phase is holonomy. Take a closed curve  $C_X$  in the base space of the fibre bundle. Obviously the horizontal lift is not necessarily closed — only its map  $\Pi : C_X$  need be. The initial and final points  $u_i$  and  $u_f$  need only be in the same fibre  $F$ . Consequently the two points are related by an element of the structure group  $G$  so that

$$u_f = u_i g \quad g \in G. \quad (3.45)$$

The group element  $g$  is called the holonomy, and will obviously depend on the choice of connection.

We are now in a position to apply this mathematical machinery to the physical world and explain the geometric phase. To do this we take the total space to be the space of normalized states  $\mathcal{N}$  and the base space to be the projective Hilbert space  $\mathcal{P}$  as in §3.3. Thus the projection mapping  $\Pi$  mapping the total

space to the base space is given by

$$\Pi = |\phi\rangle\langle\phi|. \quad (3.46)$$

If we take the structure group to be  $U(1)$  corresponding to the phase lost upon projection then we have defined a principal fibre bundle. The action of the group generates the fibres. Each fibre points in the same direction differing only by a phase. We will call this direction the vertical direction. Since we have an inner product available it is easy to define the horizontal space as the vector orthogonal to the vertical direction. Thus if  $|\phi(t)\rangle$  is a curve in  $\mathcal{N}$  then its tangent vector in  $H$  is  $|\dot{\phi}(t)\rangle$ . This tangent vector needs to be decomposed into horizontal and vertical components as in (3.44). We have  $|\phi(t)\rangle$  pointing in the vertical direction. Thus if we define a connection

$$\langle\phi(t)|\frac{d}{dt}|\phi(t)\rangle = 0 \quad (3.47)$$

we ensure that the path traced out by  $|\phi(t)\rangle$  has only a horizontal subspace component. Thus (3.47) defines a horizontal lift and is the natural connection. The holonomy is the  $U(1)$  group element connecting the final and initial points of the path and is thus given by

$$e^{i\gamma} = \langle\phi(0)|\phi(\tau)\rangle \quad (3.48)$$

where  $|\phi(t)\rangle$  satisfies (3.47) at all points.

Noting that  $H|\phi(t)\rangle = i\hbar|\dot{\phi}(t)\rangle$  and comparing (3.47) with (3.18) and (3.48) we see that the holonomy of the natural connection is the geometric phase. (Note that the  $|\phi(t)\rangle$  are equivalent to the multi-valued vectors  $|\psi(t)\rangle$  in §3.3 and not to the single-valued vectors  $|\tilde{\psi}(t)\rangle$ ).

This geometric approach can be carried over to any gauge theory. The connection on any principal bundle is given by the gauge potential one-form  $A_\mu$ , and the associated field strength two-form  $F_{\mu\nu}$  is the curvature of the bundle [135, 170]. To extend the above analysis to the nonabelian case it is only necessary to change the structure group  $G$  from  $U(1)$  to whatever nonabelian group is desired.

## 4. Time reversal constraints on the geometric phase

As this chapter will demonstrate, time reversal symmetry and the geometric phase are inextricably bound together. Given the strong nature of the connection it is perhaps somewhat surprising that very little of the established literature considers it in any depth.

Some standard results have been known for a long time. One of these is that, at least in the adiabatic case, the geometric phase for a time reversal symmetric system must be a multiple of  $\pi$  [103, 126]. Another is that the nonabelian phase factor associated with the two-fold Kramers degeneracy must be an  $SU(2)$  matrix [107, 125]. Other than these two results very little has been done to link time reversal and the geometric phase.

The sole exception appears to be work done by Ihm [91, 93, 94]. I will show that although Ihm's work contains several errors it is nonetheless ground-breaking and, to my knowledge, the only place the link between time reversal symmetry and the geometric phase has been considered with any degree of depth. The startling conclusion he reached was that a non-zero geometric phase (or non-unit matrix  $A_{ij}$  in the nonabelian case) is always due to a breakdown of time reversal symmetry of some type. That is, unless there is some form of time reversal symmetry breaking the geometric phase is always zero.

This chapter will explore thoroughly this link between time reversal symmetry and the geometric phase, reviewing, correcting and extending what has already been done in the literature.

In §4.1 I review the basics of the  $\pi$  quantization rule, and extend it to the nonadiabatic case. I then apply it to nonabelian systems using the method of breaking a nonabelian problem into a series of abelian problems as discussed in §3.4. In §4.2 I go over the work of Ihm and correct numerous errors in his basic results he later uses as tools to prove a series of cases, and show how some of his results can be saved by use of the time reversal operator that includes a component that takes  $t \rightarrow -t$  rather than the more standard operator.

In §4.3 I go over the proof that Ihm uses to demonstrate that a non-trivial geometric phase always requires the breakdown of time reversal symmetry in

some way. Ihm divides all possible situations into five cases and goes through them one by one. I take the same approach, commenting on and, if necessary, correcting each of his proofs. My conclusion is that Ihm is essentially correct in his statement that broken TRS is required for a geometric phase, provided one accepts a reasonable definition of time reversal symmetry violation.

Ihm developed his theorem in a very specific regime. He worked purely in a three-dimensional parameter space, and considers only adiabatic evolutions. These constraints are very limiting (regardless of Ihm's statement that "Generalization to higher dimensions is of little practical use"!) and in §4.4 I discuss to what degree these restrictions can be relaxed. I extend his conclusions to a parameter space of arbitrary dimension, and consider in which cases the requirement of adiabaticity can be dropped.

Finally in §4.5 I discuss whether the geometric phase, or more correctly the geometric vector potential should be considered as time even or time odd.

An aside on notation: In this chapter and the ones that follow I primarily use the time reversal operator referred to as  $T_{mr}$  in Chapter 2. Hence for notational simplicity I will refer to the operator  $T_{mr}$  as  $T$ , and the operator which includes the operation of  $t \rightarrow -t$  that I called  $T$  I will now refer to as  $T_i$ , standing for time inversion.

## 4.1 Basic results from time reversal symmetry

To begin I demonstrate one of the most fundamental results about the geometric phase for a time reversal symmetric Hamiltonian — in the abelian case the phase is constrained to be a multiple of  $\pi$ . The proof requires that  $T^2 = +1$ , which will be the case if the system under consideration has integer-spin (see, for example, [157]).

First it is necessary to prove that in the case of time reversal symmetry and  $T^2 = +1$  it is possible to choose a basis that is entirely real. This can be done a number of ways (see, for example, [77]) but the following method emphasizes how the phases are involved.

If  $|\phi(\mathbf{R})\rangle$  is an eigenket of the parameter-dependent Hamiltonian  $H(\mathbf{R})$  which commutes with the time reversal operator  $T$ , then it is possible to ensure that the basis kets  $|\phi(\mathbf{R})\rangle$  are also eigenkets of  $T$ :

$$T|\phi(\mathbf{R})\rangle = e^{i\alpha(\mathbf{R})}|\phi(\mathbf{R})\rangle. \quad (4.1)$$

We can now rephase these basis kets by

$$|\psi(\mathbf{R})\rangle = e^{i\alpha(\mathbf{R})/2}|\phi(\mathbf{R})\rangle \quad (4.2)$$

so that

$$T|\psi(\mathbf{R})\rangle = |\psi(\mathbf{R})\rangle. \quad (4.3)$$

To calculate geometric phases a single-valued basis is required. We obtain this basis by rephasing the kets  $|\psi\rangle$ :

$$|\tilde{\psi}(t)\rangle = e^{-in(t)}|\psi(t)\rangle \quad (4.4)$$

where the  $|\tilde{\psi}\rangle$  are single-valued and  $\mathbf{R}$  is parameterized by  $t$ . We take  $|\tilde{\psi}(0)\rangle = |\psi(0)\rangle$  and assume the initial vector is cyclic with period  $\tau$ . Since the  $|\psi(t)\rangle$  are not single-valued we can only say that  $|\psi(0)\rangle$  must equal  $|\psi(\tau)\rangle$  up to a phase,  $\phi$ . Thus we have

$$|\psi(0)\rangle = e^{i\phi}|\psi(\tau)\rangle. \quad (4.5)$$

Applying the time reversal operator and taking note of (4.3) we obtain

$$|\psi(0)\rangle = e^{-i\phi}|\psi(\tau)\rangle. \quad (4.6)$$

Consequently we see that  $|\psi(0)\rangle = \pm|\psi(\tau)\rangle$ , and thus, because  $|\tilde{\psi}(0)\rangle = |\tilde{\psi}(\tau)\rangle$ , we obtain the result that

$$\eta(\tau) = n\pi. \quad (4.7)$$

We are now in a position to calculate the geometric phase. We find

$$\begin{aligned} \gamma &= \int_0^\tau \langle \tilde{\psi}(t) | i \frac{d}{dt} | \tilde{\psi}(t) \rangle dt \\ &= \int_0^\tau e^{in} \langle \psi(t) | \left( \frac{d\eta}{dt} |\psi(t)\rangle + e^{-in} i \frac{d}{dt} |\psi(t)\rangle \right) dt \\ &= \int_0^\tau \left( \langle \psi(t) | i \frac{d}{dt} |\psi(t)\rangle + \langle \psi(t) | \psi(t) \rangle \frac{d\eta}{dt} \right) dt. \end{aligned} \quad (4.8)$$

As the  $|\psi(t)\rangle$  are normalized we have

$$0 = \frac{d}{dt} \langle \psi | \psi \rangle = \langle \dot{\psi} | \psi \rangle + \langle \psi | \dot{\psi} \rangle = \langle \psi | \dot{\psi} \rangle^* + \langle \psi | \dot{\psi} \rangle = 2\text{Re} \langle \psi | \dot{\psi} \rangle, \quad (4.9)$$

and thus

$$\begin{aligned} \gamma &= \int_0^\tau \frac{d\eta}{dt} dt \\ &= \eta(\tau) - \eta(0) \\ &= n\pi. \end{aligned} \quad (4.10)$$

Thus we see that in the case of time reversal symmetry and integer spin the only possible geometric phases are multiples of  $\pi$ . a restriction comparable with the adiabatic results of other workers, e.g. [93, 106, 126].

It is possible to extend this rule to the nonabelian case using my result at the end of §3.4. Recall that if it is possible to find an operator with an associated finite dimensional space that commutes with the Hamiltonian for all time then it is possible to create a basis such that the phase matrix  $U_{ij}$  is diagonal. It is often difficult to find an operator that will commute with the Hamiltonian for all time since symmetries are usually highly position dependent. For example in molecular systems where the basis kets are parameterized by nuclear positions  $\mathbf{R}$  the only symmetries that exist for all  $\mathbf{R}$  are diatomic molecules with  $C_{\infty v}$  or  $D_{\infty h}$  symmetry and triatomic systems with  $C_s$  symmetry [126]. The time reversal operator, however, is independent of spatial symmetry and thus is often a symmetry (and hence a commuting operator) for all time. Let us therefore assume a time reversal symmetric system so that

$$TH(t)T^{-1} = H(t). \quad (4.11)$$

Now, by definition there exists an operator that will commute with  $H(t)$  at all times — the time reversal operator,  $T$ . Thus it is always possible to construct a basis in which  $U_{ij}(t)$  is diagonal, provided the eigenspaces of  $T$  under consideration are considered finite. In the time reversal symmetric system under consideration there are two possible sub-cases. Either the system has overall integer-spin so  $T^2 = 1$ , or overall half-integral spin so that  $T^2 = -1$ .

As shown at the beginning of this section, in the integer-spin case it is always possible to create an orthonormal basis  $|\psi_i\rangle$  such that

$$T|\psi_i\rangle = |\psi_i\rangle. \quad (4.12)$$

Now, assume we have an integer-spin system, and are using the basis (4.12). Labelling single-valued basis vectors by a tilde we have

$$|\psi_j(\tau)\rangle = U_{ij}(\tau)|\tilde{\psi}_i(\tau)\rangle = U_{ij}(\tau)|\psi_i(0)\rangle. \quad (4.13)$$

But due to the choice of basis we also have

$$|\psi_j(\tau)\rangle = TU_{ij}(\tau)T^{-1}T|\psi_i(0)\rangle = TU_{ij}(\tau)T^{-1}|\psi_i(0)\rangle. \quad (4.14)$$

Comparing (4.13) and (4.14) we have

$$TU_{ij}T^{-1} = U_{ij}. \quad (4.15)$$

As the matrix  $U$  is nothing other than the coefficients relating one basis to another via a similarity transformation, time reversal should merely complex conjugate its entries. Consequently (4.15) shows  $U^* = U$  and thus  $U$  must be real. Since we have assumed the system is time reversal symmetric  $U$  can be

made diagonal. Finally, because  $U$  is a real diagonal unitary matrix we have the result that all its entries must be  $\pm 1$ , since a real unitary matrix can only have eigenvalues 1 or  $-1$ . Hence an integer-spin time reversal symmetric system can only acquire phases of 0 or  $\pi$ .

One should note that I have shown that the phase as a whole is a multiple of  $\pi$ , and the phase includes the dynamical part also. However in this abelian case the matrix  $K_{ij}$  in equation (3.33) is a scalar and hence commutes with  $A_{ij}$  and can be considered separately. By similar arguments to those in (4.8) and (4.9) and noting that in the basis (4.12) the Hamiltonian is given by  $i\hbar\partial_t$  it is not difficult to show that the dynamical phase is also a multiple of  $\pi$ , and thus the geometric phase must also be a multiple of  $\pi$ . Alternately one can argue that now that the nonabelian problem has been reduced to a set of abelian ones, identical reasoning to the abelian case previously can be followed to obtain an identical conclusion.

This result can also be compared with the erroneous statement of Simon who claimed that in the case of a real basis no geometric phase was possible [169]. His proof rests on the fact that his curvature (geometric field strength) two-form is zero in this case. However this is due to the fact that, as shown by Ceulemans, in some cases the curvature form does not exist if there are degeneracies present [51].

This result showing the restriction of the geometric phase of multiples of  $\pi$  need not hold for systems in which  $T^2 = -1$ . In this case the best it is possible to do is to create a basis such that

$$T(a|\phi_{2i}\rangle + b|\phi_{2i+1}\rangle) = -b^*|\phi_{2i}\rangle + a^*|\phi_{2i+1}\rangle, \quad (4.16)$$

i.e. vectors transform within a two dimensional subspace under  $T$ , corresponding to the Kramers degeneracy. Although this means that  $U_{ij}$  can be diagonalized within this subspace and the abelian approach used, the basis states  $|\phi_{2i}\rangle$  and  $|\phi_{2i+1}\rangle$  cannot themselves be cyclic states. A combination of them is required, and thus a similar proof to the integer-spin case cannot be followed through. It is, however, still possible to use time reversal arguments to constrain the form of the phase matrix, and I will do this in §4.4.

## 4.2 Ihm's results

As §4.1 has shown there do appear to be strong links between time reversal symmetry and the geometric phase. One of the few people to consider this link in detail is Ihm ([94], and in fuller detail [93]). Ihm's approach is based on an integral version of the condition for time reversal symmetry similar to that which I developed in Chapter 2. There are, however, several fatal errors in his approach. Nonetheless, some of his conclusions can be rescued by application the time reversal operator  $T_i$  rather than the more usual  $T$ .

Ihm attempts to classify all types of geometric phases by purely time reversal symmetry considerations and reaches the conclusion that the existence of any geometric phase is purely due to broken time reversal symmetry of some kind.

The approach is as follows. We choose to use the time reversal operator  $T$ . From (2.40) we know how the evolution operator transforms under  $T$  allowing us to obtain

$$T^{-1} \exp \left[ -\frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] T \exp \left[ -\frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] |\psi(t_0)\rangle = |\psi(t_0)\rangle \quad (4.17)$$

where I have explicitly written out the form of the evolution operator. Ihm's integral version of the time reversal symmetry [93] can be shown to correspond to (4.17) under a suitable integral transformation.

In the adiabatic case we have seen how it is possible to consider the geometric phase as arising from a gauge potential, and thus for a nondegenerate state we have

$$|\psi(\tau)\rangle = \exp \left[ -\frac{i}{\hbar} \int_0^\tau E(t) dt \right] \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] |\psi(0)\rangle. \quad (4.18)$$

Bringing in the evolution operator explicitly and noting that the cyclic initial vectors are eigenstates of the evolution operator after time  $\tau$ , we see that (4.18) becomes

$$\exp \left[ -\frac{i}{\hbar} \int_0^\tau H(t) dt \right] = \exp \left[ -\frac{i}{\hbar} \int_0^\tau E(t) dt \right] \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right]. \quad (4.19)$$

It is at this point that Ihm starts to introduce errors. He states that in the case of a degenerate state (4.19) is generalized to the nonabelian case by letting the vector potential become matrix-valued:

$$A_{ij} = \langle \psi_j | i \nabla | \psi_i \rangle. \quad (4.20)$$

This step is not trivial. Because the phase matrix  $U_{ij}$  associated with the definition (4.20) acts on states via

$$|\psi_i(\tau)\rangle = U_{ij} |\tilde{\psi}_j(\tau)\rangle = U_{ij} |\psi_j(0)\rangle \quad (4.21)$$

it is easy to confuse it with the matrix representation of an operator, which Ihm appears to have done. It is important to realize that it acts on sets of states rather than the column-vector representation of a state. Consequently to claim that because of the appearance of (4.21)  $U_{ij}(\tau, 0)$  can replace the evolution operator  $\mathcal{E}(\tau, 0)$  for cyclic initial vectors of period  $\tau$  is incorrect. The question thus becomes: does there exist a privileged basis in which Ihm's statement is true? I now show that such a basis exists. Consider a cyclic initial state  $|\Psi(0)\rangle = a_i |\psi_i(0)\rangle$ . Under the action of the evolution operator we have

$$\begin{aligned} \mathcal{E}(\tau, 0) |\Psi(0)\rangle &= |\Psi(\tau)\rangle \\ &= a_i \mathcal{E}(\tau, 0) |\psi_i(0)\rangle. \end{aligned} \quad (4.22)$$

In order for the phase matrix  $U$  to act like the matrix representation of the evolution operator we must have (using (4.21))

$$U a_i |\psi_i(0)\rangle = a_i |\psi_i(\tau)\rangle = a_i U_{ij} |\tilde{\psi}_j(\tau)\rangle = a_i U_{ij} |\psi_j(0)\rangle. \quad (4.23)$$

A representation that satisfies this is given by letting

$$\begin{aligned} |\psi_1(0)\rangle &= (1, 0, 0, \dots, 0)^T \\ |\psi_2(0)\rangle &= (0, 1, 0, \dots, 0)^T \\ &\vdots \\ |\psi_n(0)\rangle &= (0, 0, 0, \dots, 1)^T. \end{aligned} \quad (4.24)$$

With this representation (4.23) becomes

$$\begin{aligned} U \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} &= a_i \begin{pmatrix} U_{i1} \\ U_{i2} \\ \vdots \\ U_{in} \end{pmatrix} \\ \begin{pmatrix} a_1 U_{11} + a_2 U_{12} + \dots + a_n U_{1n} \\ a_1 U_{21} + a_2 U_{22} + \dots + a_n U_{2n} \\ \vdots \\ a_1 U_{n1} + a_2 U_{n2} + \dots + a_n U_{nn} \end{pmatrix} &= \begin{pmatrix} a_1 U_{11} + a_2 U_{21} + \dots + a_n U_{n1} \\ a_1 U_{12} + a_2 U_{22} + \dots + a_n U_{n2} \\ \vdots \\ a_1 U_{1n} + a_2 U_{2n} + \dots + a_n U_{nn} \end{pmatrix} \end{aligned} \quad (4.25)$$

Hence in this representation we see that the evolution operator  $\mathcal{E}(\tau, 0)$  has the representation  $U(\tau)^T$ . If we had chosen the alternate definition of the vector potential and phase matrix operation

$$A_{ij} = \langle \psi_i | i \nabla | \psi_j \rangle \quad |\psi_i\rangle = U_{ij} |\tilde{\psi}_j(\tau)\rangle \quad (4.27)$$

then the same argument shows that evolution operator has the representation  $U(\tau)$ .

Having been forced into a specific representation we can proceed. Ihm uses (4.27) as his definition of  $A_{ij}$  so I shall adopt this convention also, allowing us to let (4.19) be generalized to the nonabelian case by letting  $\mathbf{A}$  become a vector of matrices. Noting that in the adiabatic case (3.24) shows that the dynamical part of the phase is diagonal and hence commutes with  $A_{ij}^{\mu}$ , we insert (4.19) into (4.17) to obtain

$$T^{-1} \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] T \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] = 1. \quad (4.28)$$

As this point Ihm makes a second error and instead of (4.28) obtains

$$T^{-1} \exp \left[ -i \oint \mathbf{A} \cdot d\mathbf{R} \right] T \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] = 1, \quad (4.29)$$

that is the sign of one exponent is reversed. This error turns out to be vital for Ihm's later conclusions.

We now consider an infinitesimal rectangular path around a point. As the size of the rectangle tends to zero,  $\oint A_\mu dx^\mu$  may or may not converge to zero depending on whether the integral is singular or not. Assuming that it does converge to zero then we can make the approximation

$$\exp[i \oint A_\mu dx^\mu] \approx 1 + i \oint A_\mu dx^\mu. \quad (4.30)$$

For an infinitesimal closed path Stokes' theorem still holds in the nonabelian case [125] provided the parameter space is three dimensional, and hence

$$\exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] \approx i \mathbf{B} \cdot d\mathbf{S} \quad (4.31)$$

where  $\mathbf{B}$  is the geometric magnetic field analogue to the geometric vector potential, and can also be written in component form as

$$B_i = \epsilon_{ijk} (i\partial_j + A_j)(i\partial_k + A_k). \quad (4.32)$$

This may be compared with Ihm's (incorrect) version of  $B_i$  which includes an extra factor of  $-i$ .

Thus, replacing  $\oint \mathbf{A} \cdot d\mathbf{R}$  by  $\epsilon B_i$  where  $\epsilon$  is the infinitesimal area and  $B_i$  is the component of  $\mathbf{B}$  normal to the area of integration, I now follow Ihm. We expand (4.29) as

$$T^{-1}(I - i\epsilon B_i)T(I + i\epsilon B_i) = I. \quad (4.33)$$

Comparing terms of first order in  $\epsilon$  Ihm obtains

$$TBT^{-1} = -B. \quad (4.34)$$

Noting that the operator  $T$  is composed of a unitary operator plus the operation of complex conjugation, and that from (4.32)  $B_i$  is hermitian, one can take the trace of both sides of (4.34) to get

$$\text{tr}(B_i^*) = \text{tr}(B_i) = \text{tr}(-B_i) \quad (4.35)$$

demonstrating that

$$\text{tr}(B_i) = 0. \quad (4.36)$$

This implies that  $\text{tr} \oint \mathbf{A} \cdot d\mathbf{R} = 0$  and hence the phase factor  $\exp[i \oint \mathbf{A} \cdot d\mathbf{R}]$  is special unitary,  $SU(n)$ .

This is Ihm's conclusion. However it must be remembered that this result is based on (4.29) which I have shown to be incorrect. Using the correct equation, (4.28), and following the same line of reasoning as above yields

$$TBT^{-1} = B. \quad (4.37)$$

Consequently taking traces as above merely gives  $\text{tr}(B_i) = \text{tr}(B_i)$  and no new information.

The result that nonabelian phase factor is  $SU(n)$  can be rescued, however. Looking back to (2.39) and (2.40) to see how the evolution operator transforms under the two forms of the time reversal operator we see that we can write (4.17) as

$$T_i^{-1} \exp \left[ \frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] T_i \exp \left[ -\frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] |\psi(t_0)\rangle = |\psi(t_0)\rangle. \quad (4.38)$$

Following through (4.18) and (4.19) accordingly with this new definition we find that (4.28) becomes

$$T_i^{-1} \exp \left[ -i \oint \mathbf{A} \cdot d\mathbf{R} \right] T_i \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] = 1. \quad (4.39)$$

Consequently equations (4.33) through (4.36) follow as does the conclusion regarding the special unitary nature of the vector potential  $\mathbf{A}$ .

### 4.3 Classifying phases using time reversal criteria

Using the results in the previous section, Ihm tries to demonstrate that the existence of a non-zero geometric phase can be ascribed purely to broken time reversal symmetry by dividing all possible situations into five separate cases and examining each individually. I consider each of his cases in detail and comment and correct where necessary.

#### *Case 1: Nondegenerate, nonsingular, TRS*

In this case we assume we have a time reversal symmetric system, and  $\mathbf{A}$  is not singular, i.e. the flux enclosed with the path shrinks to zero as the area enclosed also tends to zero. Because we are dealing with a nondegenerate level we have the abelian case. Hence  $\mathbf{B}$  is one-dimensional and thus from (4.36) we see that  $\mathbf{B} = 0$  and the phase must necessarily be zero.

*Case 2: Degenerate from other than TRS, nonsingular, TRS*

Ihm assumes double degeneracy, that is a system with two degenerate levels. I shall assume the general case with an  $n$ -fold degenerate level. Since I have assumed that the degeneracy arose from other than time reversal considerations we cannot have Kramers degeneracy. Consequently we can choose a basis  $|\psi_i\rangle$  of the form (4.12). Thus

$$\begin{aligned}
 \mathbf{A}_{ij} &= \langle \psi_i | i \nabla | \psi_j \rangle \\
 &= (T \langle \psi_i | i \nabla | \psi_j \rangle T^{-1})^* \\
 &= \langle T \psi_i | T i \nabla T^{-1} | T \psi_j \rangle^* \\
 &= -\langle \psi_i | i \nabla | \psi_j \rangle^* \\
 &= -\mathbf{A}_{ji}
 \end{aligned} \tag{4.40}$$

where I have used the hermiticity of  $\mathbf{A}_{ij}$ . Thus the vector potential  $\mathbf{A}_{ij}$  must be purely imaginary and the phase matrix  $U_{ij}$  must be real and unitary. This case is by far the most confused of Ihm's cases. Allowing for typographical errors and seen in the best possible light his argument appears to be as follows: Write the diagonal elements of  $U_{ij}$  in polar form, i.e. as  $U_{ii} = r e^{i\gamma_{ii}}$ . He then defines the geometric phase as the  $\gamma_{ii}$  terms. Because the  $U_{ii}$  are real, the  $\gamma_{ii}$  must be either 0 or  $\pi$ . Further, "... it is much more natural to say that the amplitude of  $r e^{i\gamma_{ii}}$  goes continuously to a negative number  $-r$  rather than to say that  $|\psi_i\rangle$  suddenly picks up a phase  $\pi$  as the amplitude goes through the value 0." The conclusion he draws is that in this case one can say that there is no geometric phase.

Laying aside the extremely odd definition of the geometric phase in this case, assigning a phase change to the amplitude of a polar quantity would not appear to be reasonable. It would seem that here Ihm is nibbling at the edges of the work of Mead [126] who uses the methods of Pancharatnam [143] to construct a single-valued basis and compare phases with a multi-valued basis. This is accomplished by comparing basis kets to a reference ket using a projection operator. At certain points on the path the projection operator may annihilate the reference ket which can be shown to result in a contribution to the geometric phase of exactly  $\pi$ .

Nonetheless, Ihm is correct in his conclusion. As I demonstrated in §4.1, with the assumptions for Case 2 it is possible to choose a basis such that the phase matrix is diagonalized and the phase changes are always integer multiples of  $\pi$ . Since a path enclosing zero area must necessarily have a phase of zero by Stokes' theorem and we can continuously deform such a path to obtain the desired path, we see that these jumps must be due to singularities in the geometric vector potential at various points and that if there are no singularities then the phase must indeed be zero (Figure 4.1). Although one must be careful using this argument in the nonabelian case because then there exists no analogue to flux or simple equivalent to Stokes' theorem [89], it must be remembered that because the Hamiltonian commutes with the time reversal operator we have reduced the

nonabelian problem to a series of abelian problems for which a flux analogue exists.

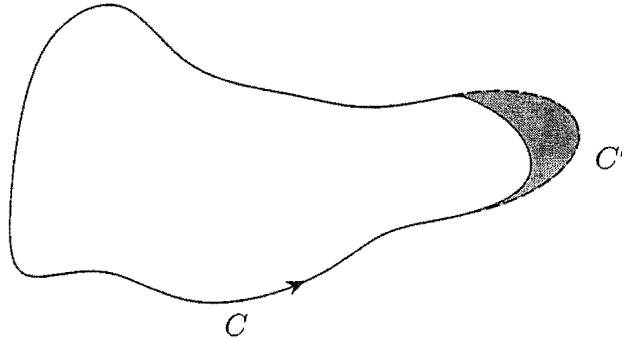


Fig. 4.1: The infinitesimal change from path  $C$  to  $C'$  gives rise to an extra geometric phase of  $\pi$ . Thus the shaded area must enclose a flux singularity by Stokes' theorem.

### Case 3: Singular, TRS case

I have treated this case at the end of Case 2. I include it here for completeness and to demonstrate the usefulness of the integral definition of time reversal symmetry, and also to demonstrate that the existence of a phase of  $\pi$  due to singularity can be viewed as due to broken time reversal symmetry even though the Hamiltonian is apparently TRS.

If  $n = 1$  (Case 1) then the phase factor is simply a complex number and hence utilizing the time reversal operator  $T_i$  and (4.39) we obtain

$$T_i^{-1} \exp \left[ -i \oint \mathbf{A} \cdot d\mathbf{R} \right] T_i \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] = \left[ \exp \left( i \oint \mathbf{A} \cdot d\mathbf{R} \right) \right]^2 = 1 \quad (4.41)$$

and consequently

$$\oint \mathbf{A} \cdot d\mathbf{R} = n\pi. \quad (4.42)$$

It should be noted here that Ihm's version of the proof is again incorrect due to his use of the operator  $T$  and his erroneous equation (4.29).

For the nonabelian case the argument has already been made in Case 2, and once again any phase generated must be a multiple of  $\pi$ . It is instructive to note that by the same argument one can see that if there is a singularity somewhere

then this will almost certainly mean the existence of a non-trivial phase, except in such non-physical cases as  $\oint \mathbf{A} \cdot d\mathbf{R}$  oscillating as the area  $\epsilon \rightarrow 0$ .

It is not immediately obvious, however, how this phase of  $\pi$  can be ascribed to some form of broken time reversal symmetry. The answer, at least in a three-dimensional parameter space, can be understood in terms of flux tubes (for a good discussion of their properties see, for example, [5, 9]). Suppose there exists a singularity such that if the path encloses it the total phase jumps by  $\pi$ . This can be modelled by the existence of an infinitely long line of magnetic field with negligible cross-section, such that the total flux is  $\Phi$ , i.e. a line singularity. If we replace the idea of magnetic flux with the geometric analogue arising from the curl of the geometric vector potential we obtain the desired system. Now such a situation would certainly not appear to be time reversal symmetric, including as it does a magnetic field which reverses under time reversal. However the only way to determine the existence of this flux tube is to globally measure the phase change arising from a circuit around it. If  $\Phi$  has a strength of  $\pi$  then a circuit around the flux tube will differ from a circuit around the time reversed system by a phase of  $2\pi$  and the two cases will be physically indistinguishable. Thus we have a situation where even though the Hamiltonian is not formally time reversal symmetric it is not possible to physically measure this and the integral definition of time reversal symmetry (4.38) still holds globally.

#### *Case 4: Spontaneous TRS breaking*

In this case we assume degeneracy is due to time reversal symmetry, that is  $|\psi\rangle$  and  $T|\psi\rangle$  are degenerate and linearly independent. This is an example of a spontaneously broken symmetry — although the Hamiltonian describing the system is time reversal symmetric the actual state of the system has two equally valid solutions within this symmetry. Thus the system picks out a ground state that does not fully reflect the total symmetry and the symmetry of the system is said to be spontaneously broken.

Ihm does not actually demonstrate that this case *must* give rise to a non-trivial geometric phase, but rather gives examples to show that phases can arise in this situation. I refer the reader back to §4.1 for my demonstration that the phase matrix in this case must be  $SU(2)$ .

This case of Kramers degeneracy has been well studied in the literature. It is, in fact, what the bulk of proposed experimental schemes to measure the nonabelian geometric phase consider. Mead has considered this case as it relates to a Stark Hamiltonian governed by a rotating electric field [125], and Zee has considered nuclear quadrupole resonance [206]. Avron *et al.* gave an incisive theoretical analysis utilizing a quaternionic formalism of systems governed by the Hamiltonian  $\sum_{ij} Q_{ij} J_i J_j$  [25]. In this case  $Q$  is a second rank tensor and the  $J$  are spin operators so that their square is time even. All these examples are based on spin half systems. Ihm considers things from another perspective, and constructs

a time even system that is doubly degenerate due to spontaneously broken time reversal symmetry yet is bosonic rather than fermionic [93]. His analysis of this system demonstrates that non-trivial phases can always be generated and is based on the quantum mechanical rotated rotator considered by Berry [32].

I consider Kramers degeneracy in much greater detail in Chapter 6, as well as its logical extensions.

#### *Case 5: Non-TRS system*

In this final case we consider the remaining situation: The Hamiltonian is not time reversal symmetric. First, we make the assumption that

$$\oint \mathbf{A} \cdot d\mathbf{R} = 0 \quad (4.43)$$

for all possible paths in the parameter space under consideration. Now, utilizing the time reversal operator  $T_i$  and appealing to (4.19) and (4.43) we obtain

$$\begin{aligned} & T_i^{-1} \exp \left[ \frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] T_i \exp \left[ -\frac{i}{\hbar} \int_{t_0}^{t_1} H(t) dt \right] \\ &= T_i^{-1} \exp \left[ -i \oint \mathbf{A} \cdot d\mathbf{R} \right] T_i \exp \left[ i \oint \mathbf{A} \cdot d\mathbf{R} \right] \\ &= I. \end{aligned} \quad (4.44)$$

Looking back to the definition of time reversal symmetry (2.39) we see that (4.44) demonstrates that the system is actually time reversal symmetric, a contradiction. Consequently the assumption in (4.43) is false, and there must exist paths which generate a non-trivial phase.

Having considered these five cases Ihm then draws the conclusion that to generate a non-zero geometric phase requires a violation of time reversal symmetry in some form. After his arguments are corrected and errors patched up it is seen that this conclusion is actually true. Perhaps the argument that the spontaneously broken symmetry case is TRS-violating is stretching the definition of time reversal violation a little, but that is a semantic point.

It should also be noted that Ihm's cases were only considered for Hamiltonians depending on a three-dimensional parameter space and in the adiabatic limit, which is quite restricting. I will now discuss to what extent the above conclusions can be generalized.

## 4.4 Extending Ihm's results

Firstly consider the restriction to three parameters. This has the advantage of remaining in familiar vector calculus territory, and also allows the construction

of meaningful magnetic flux analogues via Stokes' theorem.

The vector calculus of three dimensions has its natural extension in the theory of differential forms, where an arbitrary number of dimensions are allowed for. This extension admits a generalized Stokes' theorem [135]:

$$\int_X d\omega = \int_{\partial X} \omega \quad (4.45)$$

where  $X$  is an oriented  $n$ -dimensional manifold,  $\partial X$  its boundary,  $\omega$  is an  $(n-1)$ -form and  $d$  is the operation of exterior differentiation. Thus all cases dependent on the three dimensional nature of the parameter space can still be followed through in an  $n$ -dimensional parameter space by replacing  $\mathbf{A} \cdot d\mathbf{R}$  by the  $n$ -dimensional one-form  $A_\mu dx^\mu$ . Ceulemans and Szopa have a useful article discussing some of the basics of using differential forms to calculate geometric phases in a quantum mechanical formalism [51]. As an example of general higher-dimensional parameter spaces we consider Mead [126] who constructs the field tensor corresponding to (3.13) and finds

$$F_{\mu\nu} = \frac{1}{i} \sum_k \frac{\langle 0 | \partial_\mu H | k \rangle \langle k | \partial_\nu H | 0 \rangle - \langle 0 | \partial_\nu H | k \rangle \langle k | \partial_\mu H | 0 \rangle}{(E_0 - E_k)^2}. \quad (4.46)$$

This relates back to Case 3. We see that if we choose the basis kets in (4.46) as given in (4.12) then the matrix elements of  $H$  and hence  $\partial_\mu H$  will be real [123]. Consequently the flux given by (4.46) must be zero unless there exists electronic degeneracy. Thus we see that it is possible to expand on the conclusions in Case 3: A singularity in the geometric potential is required in order to give rise to a phase of  $\pi$  in the time reversal symmetric case; furthermore a singularity is equivalent to the existence of a point of degeneracy.

I now examine the argument that the singularity can be considered as being due to a time reversal violating term in the Hamiltonian, and adapt it to an  $n$ -dimensional parameter space. The previous argument was based on the fact that a one-dimensional flux tube can give rise to a phase factor of  $\pi$ . From (4.46) we see that the shape of the singularity is due to the the shape of the subspace of the parameter space that is defined by points at which degeneracy occurs between at least two of the basis states. Thus the question we must ask is, given a time reversal symmetric system what can we say about the form of the degeneracy space? To answer this I use a result from Mead's beautiful proof of the noncrossing rule [123]. Mead demonstrated that in a time-even system that allowed a basis of the form (4.12), for  $i$  levels to simultaneously cross (become degenerate) there must exist a set number of conditions on the Hamiltonian. The number of conditions is

$$c(i) = \frac{1}{2}(i+2)(i-1). \quad (4.47)$$

In this case we only need two levels to cross which yields two constraints on the Hamiltonian and hence on the parameter space. Thus in an  $n$ -dimensional

parameter space the degeneracy manifold will have dimension  $n - 2$ . Hence if  $n = 3$  the subspace will be one-dimensional, corresponding to a flux tube.

We note that if a subspace has two dimensions less than the embedding space it is possible to unambiguously define whether a closed loop has encircled the subspace or not. For example if the embedding space is two dimensional and the subspace is a zero-dimensional point, loops can be defined, as they can in the case of a line in the three dimensional space. This can be extended to spaces of arbitrary dimensionality inductively. A sketch proof is as follows: The idea of a closed loop encircling topological obstacles has an exact mathematical definition within the theory of homotopy groups [135, 89]. The fundamental homotopy group, denoted  $\pi_1$ , is defined as having elements that consist of classes of closed loop within the space that cannot be continuously deformed into each other. These closed loops can be shown to obey the axioms of a group. If the homotopy group is not trivial, i.e. loops cannot all be deformed into each other, the space is said to be non-simply connected. Thus, for example, the space  $X$  consisting of the punctured plane  $S^1 = \mathbb{R}^2 - \{0\}$  (labelled  $S^1$  since the circle is the deformation retract of the punctured plane and hence has the same homotopy group) is non-simply connected and has the fundamental homotopy group  $\pi_1(X) = \mathbb{Z}$ , the integers. This number is also known as the winding number — how many times the loops wraps around the puncture. Furthermore, this number is a topological invariant — a loop with winding number one can be continuously deformed into any other loop with winding number one, but not a loop with winding number two, and so on. This is shown in Figure 4.2. Figures 4.2 a) – c) have winding numbers zero, one and two respectively. The space in the Aharonov-Bohm experiment is an example of such a non-simply connected space, and this is the reason for much of its interesting behaviour.

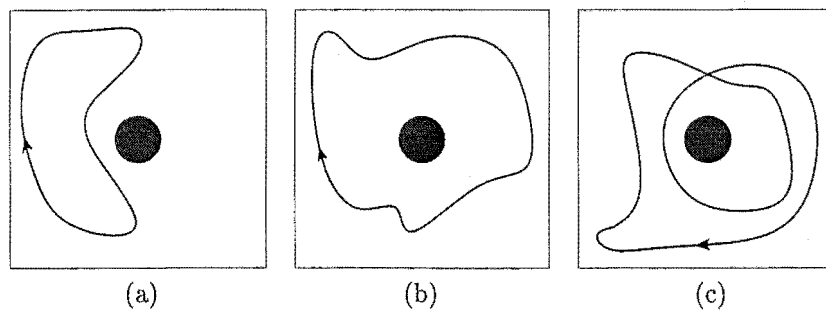


Fig. 4.2: Inequivalent classes of loop in a non-simply connected space

Now, there exists a theorem in homotopy theory which states that the fundamental group of the product of two topological spaces  $X$  and  $Y$  is isomorphic to the direct product of their fundamental groups:

$$\pi_1(X \times Y) \simeq \pi_1(X) \oplus \pi_1(Y). \quad (4.48)$$

To move from the punctured plane to the case of a three dimensional space with a line singularity we need only take the product of the space  $S^1$  with the real numbers:  $S^1 \times \mathbb{R}$ . The fundamental homotopy group of the real line is trivial and only has one element, the identity. Hence

$$\pi_1(S^1 \times \mathbb{R}) \simeq \pi_1(S^1) \oplus \pi_1(\mathbb{R}) \quad (4.49)$$

$$\simeq \mathbb{Z} \oplus I \quad (4.50)$$

$$\simeq \mathbb{Z}. \quad (4.51)$$

The case of a plane singularity in a four dimensional space is just  $S^1 \times \mathbb{R} \times \mathbb{R}$ , and consequently its fundamental homotopy group is also  $\mathbb{Z}$ . Identical results in higher dimensions follow by induction, completing the proof.

Thus in higher dimensions one can construct generalized flux tubes that behave as singularities and impart a phase of  $\pi$  when the path in parameter space encloses them. Thus the conclusion of Case 3 that a geometric phase of  $\pi$  can still be ascribed to a formally time-odd Hamiltonian, even in the apparently time-even case, is still valid for an arbitrary dimensional parameter space.

The extension to nonadiabatic evolutions is not so clear cut. Looking at (3.35) we see that in the nonadiabatic case the dynamical (i.e. the part governed by the Hamiltonian) phase is not necessarily diagonal as it was in the abelian case. Consequently we cannot expect this dynamical part to cancel when we insert (4.19) into (4.17). Instead of (4.28) we get

$$T_i^{-1} \exp \left[ -i \int (A_{ij} - K_{ij}) dt \right] T_i \exp \left[ i \oint (A_{ij} - K_{ij}) dt \right] = 1 \quad (4.52)$$

where  $K_{ij} = 1/\hbar \langle \tilde{\psi}_i | H | \tilde{\psi}_j \rangle$  as in (3.35). Thus unless  $K_{ij}$  commutes with  $A_{ij}$  for all  $0 \leq t \leq \tau$  one cannot use the same methods to prove Cases 1, 3, and 5, and although we have seen it is possible to choose a basis such that  $A_{ij} - K_{ij}$  is diagonal, neither  $A_{ij}$  or  $K_{ij}$  need be separately diagonal.

Cases 1, 2 and 3 have already been demonstrated in the nonadiabatic regime using my alternative method of proof in §4.1, along with my comments in Case 2. However Case 4, spontaneous symmetry breaking, has not been extended to nonadiabatic evolutions. I now consider it.

Using a degenerate basis pair of the form (4.16) and labelling them  $|\psi_1(t)\rangle$  and  $|\psi_2(t)\rangle$  we have

$$|\psi_1(\tau)\rangle = U_{11}|\tilde{\psi}_1(0)\rangle + U_{12}|\tilde{\psi}_2(0)\rangle \quad (4.53)$$

$$|\psi_2(\tau)\rangle = U_{21}|\tilde{\psi}_1(0)\rangle + U_{22}|\tilde{\psi}_2(0)\rangle \quad (4.54)$$

where  $|\tilde{\psi}_i(t)\rangle$  is a single-valued basis and  $|\tilde{\psi}_i(0)\rangle = |\psi_i(0)\rangle$ . Applying the time reversal operator to (4.53) and noting that

$$T|\psi_1\rangle = |\psi_2\rangle \quad T|\psi_2\rangle = -|\psi_1\rangle \quad (4.55)$$

we obtain

$$|\psi_2(\tau)\rangle = TU_{11}T^{-1}|\psi_2(0)\rangle - TU_{12}T^{-1}|\psi_1(0)\rangle. \quad (4.56)$$

Comparing (4.53) and (4.56) and exploiting the orthogonality of the  $|\psi_i\rangle$  we see that

$$U_{11} = U_{22}^* \quad (4.57)$$

$$U_{12} = -U_{21}^*. \quad (4.58)$$

Hence  $U$  has a form

$$U = \begin{pmatrix} a + ib & -c + id \\ c + id & a - ib \end{pmatrix} \quad (4.59)$$

where  $a, b, c$  and  $d$  are real parameters. Now, from its definition in §3.4 we know that the phase matrix  $U_{ij}$  is unitary, that is  $U^\dagger U = 1$ . Applying this to the matrix (4.59) we get

$$\begin{pmatrix} a^2 + b^2 + c^2 + d^2 & 0 \\ 0 & a^2 + b^2 + c^2 + d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (4.60)$$

yielding the constraint  $a^2 + b^2 + c^2 + d^2 = 1$ . Finally we calculate the determinant of  $U$ :

$$\begin{aligned} \det(U) &= (a + ib)(a - ib) - (c + id)(-c + id) \\ &= a^2 + b^2 + c^2 + d^2 \\ &= 1. \end{aligned} \quad (4.61)$$

Consequently the phase matrix in this basis must be special unitary, or in this case  $SU(2)$ . A similar result has been derived by Mead in the adiabatic case [126]. This demonstrates that even in the nonadiabatic case spontaneous breaking of time reversal symmetry is sufficient to give rise to non-trivial geometric phases.

We are now only left with Case 5. Unfortunately this proof is based entirely on (4.39) which as we have seen does not generalize to nonadiabatic evolutions, and I have not been able to construct a similar proof for this case. There are, however, well studied examples of time-odd, nonadiabatic systems which give rise to arbitrary geometric phases. An example of such a system is a neutron in a rapidly varying external magnetic field [15]. In any case, the principal purpose of this section was to demonstrate the result that a non-zero geometric phase is purely due to the violation of time reversal symmetry in some form, and this has been achieved.

## 4.5 The time reversal signature of the geometric vector potential

In Chapter 3 we saw that the geometric vector potential was analogous to the electromagnetic potential in many ways. One question that arises is whether

this similarity extends to its behaviour under time reversal.

It is well known that the electromagnetic vector potential is time-odd, as can be seen by requiring Maxwell's equations to reflect the time reversal symmetry of electromagnetism, or by considering it to be generated by the motion of electrical charges whose velocity vectors change sign under time reversal. The geometric vector potential, on the other hand, cannot be seen to arise from the motion of objects and neither can we appeal to a set of equations governing a classical analogue as we can for electromagnetism. Thus its character under time reversal is not immediately obvious. In searching for an answer we will see that one must be careful in distinguishing between the cases of local and global time reversal symmetry.

The question acquires more importance when considering the role of the geometric phase in molecular physics. In Chapter 7 we will see that such molecular systems are normally solved under the Born-Oppenheimer approximation, where the system is separated into separate regimes: the slowly moving nuclei and the much faster and lighter electrons which follow the nuclei adiabatically. Each regime is solved separately, and it turns out that the geometric vector potential makes an appearance in slow nuclear equations of motion. The effective Hamiltonian describing the nuclear motion is

$$H(x) = -\frac{\hbar^2}{2M}(\partial_\mu - i\mathbf{A}(x))^2 + V(x) \quad (4.62)$$

where  $\mathbf{A}(x)$  is the geometric vector potential. It is coupled to the nuclear momentum exactly as an external electromagnetic field would be coupled. The importance of the time reversal signature of the geometric vector potential is now clear. Can the usual time reversal selection rules governing molecular behaviour and, for example, optical transitions, apply to the Born-Oppenheimer approximation for molecular calculations?

Ihm has characterized the time reversal signature of the geometric vector potential as time-even [91]. I believe this is incorrect and that its signature is actually time-*odd*. Ihm's argument for time-evenness is simple. The vector potential is given as

$$\mathbf{A}(x) = \langle \psi(x) | i \frac{\partial}{\partial x_\mu} | \psi(x) \rangle \quad (4.63)$$

and as such must be a real quantity, being the expectation value of a hermitian operator. Because it is real and time reversal involves complex conjugation Ihm claims  $T\mathbf{A}T^{-1} = \mathbf{A}$ . This reasoning is incorrect. If a real external electromagnetic field is imposed on a system the minimal coupling prescription ensures that it is coupled to the momentum as in (4.63). One does not, however, then claim that because  $\mathbf{A}_{\text{em}}$  is real it must be time-even. To determine the real behaviour of the geometric vector potential under time reversal we must consider what happens to it when we run the entire system backwards, as per the "movie" analogy developed in Chapter 2.

First, because (excluding the weak interaction) all processes in physics are *globally* time reversal symmetric, if we allow a closed system to run forward accumulating a geometric phase, then time reverse the system and and evolve it back to the starting point, the phase must necessarily undo itself. This experiment has actually been done using an optical geometric phase and a phase conjugate mirror [188], and, as expected, the geometric phase was seen to be globally time-even. Suppose on the forward run the system accumulates a total geometric phase  $\gamma$ . Applying time reversal will effectively reverse the sign of the phase (since the phase appears in the phase factor as  $e^{i\gamma}$  and the  $i$  will change sign) to  $-\gamma$ . During the time-reversed backward run the trajectory of system will retrace its path through parameter space. Given that the geometric phase is given by the line integral

$$\gamma = \int_{x_1}^{x_2} A_\mu dx^\mu \quad (4.64)$$

it is obvious that reversing the path will result in a total phase of  $\gamma$  wiping out the phase of  $-\gamma$  only if the sign of  $A_\mu$  is also changed. This is exactly analogous to the way the sign of the electromagnetic vector potential must reverse in order for the phase of a charged particle to undo under time reversal.

If one wishes to apply Ihm's argument of time-reversing the matrix element defining  $\mathbf{A}$  one must bear in mind that upon time reversing the system the electronic states reverse:

$$|\psi(x)\rangle \longrightarrow T|\psi(x)\rangle = |\bar{\psi}(x)\rangle. \quad (4.65)$$

Thus we have

$$\begin{aligned} \bar{\mathbf{A}} &= \langle \bar{\psi} | i\partial_\mu | \bar{\psi} \rangle \\ &= \langle T\psi | i\partial_\mu | T\psi \rangle \\ &= -\langle T\psi | T i\partial_\mu \psi \rangle \\ &= -\langle \psi | i\partial_\mu | \psi \rangle^* \\ &= -\langle \psi | i\partial_\mu | \psi \rangle \end{aligned} \quad (4.66)$$

since  $i\partial_\mu$  is a hermitian operator, demonstrating that  $\mathbf{A}$  is indeed time-odd.

It is important to realize, however, that  $\mathbf{A}$  is sometimes to be treated as an external parameter, and thus may not need to be affected by the time reversal operator. An example is that of a sample in an external magnetic field. When considering the Zeeman interaction one expects time-odd effects even though as a whole the system is time-even. For example such effects can arise from an  $\mathbf{A} \cdot \mathbf{p}$  interaction, which is treated as time odd because the vector potential is external, even though  $\mathbf{A}$  and  $\mathbf{p}$  are separately each time-odd and one might expect their product to be time-even. Only the sample is of interest, and due to the external field its time reversal symmetry is locally broken.

The situation is similar for the geometric vector potential. If the geometric phase arises from the system being driven around the parameter space by an external influence, say an imposed field or physical rotation of the system under study, then the associated geometric vector potential must be considered time-even. If, on the other hand, the potential arises from internal interactions, such as when it is generated in the Born-Oppenheimer approximation, then it most certainly should be treated as time-odd.

Using his assumption that the geometric vector potential was time-even, even within the Born-Oppenheimer approximation, Ihm suggested using the geometric phase as a probe for examining the properties of high-temperature anyonic superconductors [91]. Because within a crystal lattice vibrational modes (phonons) couple to the geometric potential as  $\mathbf{A} \cdot \mathbf{p}$  he believed that it would be possible to transmit a circularly polarized phonon mode through a crystal, reflect it and have the acoustic activity factor cancel itself on the return leg but have the geometric phase factor double itself due to the time-oddness of the interaction  $\mathbf{A} \cdot \mathbf{p}$  factor, similar to magnetic circular dichroism with circularly polarized light passing through a fluid [175]. Unfortunately, as we have seen in this section, the geometric phase factor is time-odd in the situation he was considering, and thus his planned experiment could not succeed.

## 5. The U(1) monopole and the geometric phase

The magnetic monopole is a very interesting object, whose properties were first examined by Dirac in the interests of symmetrizing Maxwell's equations with respect to electric and magnetic effects [60]. It also appears in the study of geometric phases [31]. The purpose of this chapter is to examine this U(1) monopole in some detail in order to understand the links between the magnetic monopole and the one arising from geometric gauge fields. It also sets the scene for my work on the SU(2) monopole and instanton in Chapter 6, using the U(1) abelian case to demonstrate the approaches and mathematical techniques I will use later in the nonabelian case.

This chapter begins by examining the U(1) Dirac magnetic monopole in some detail, studying some of its topological properties and the implications for electric and magnetic charge quantization [109].

In §5.2 I go on to show how a very similar object appears in the study of geometric phases. More specifically I consider a general two-level crossing and show that this will generate a geometric vector potential that is functionally identical to Dirac's monopole. To make this idea more concrete, in §5.3 I consider a specific case of the two-level crossing, that of a particle with non-zero spin interacting with an external, adiabatically varying magnetic field. Because the geometric potential of a monopole can give rise to arbitrary phases, I classify the example of the spin-magnetic field system and the general two-level crossing according to the scheme developed in Chapter 4.

In §5.4 I examine the Dirac monopole from a more mathematical perspective, treating it as being described by a non-trivial principal U(1) bundle over the sphere  $S^2$ . This approach elegantly demonstrates the quantization rule for monopole magnetic charges. It also allows an analogous treatment of the geometric monopole since, topologically speaking, they are identical. Consequently the geometric monopole strength is also quantized. Perhaps most interesting of all, drawing on the result from §5.3, I demonstrate that from this quantization follows the fact that spin is quantized in units of  $\frac{1}{2}\hbar$ .

I close the chapter with a speculative look at further possible connections between time reversal and monopoles in general, both geometric and electromagne-

tic.

## 5.1 The Dirac monopole

It is fundamental to standard electromagnetism that while opposite electric charges (both positive and negative) exist independently, magnetic sources do not appear to be divisible into separate north and south poles. Put another way, lines of magnetic field are always closed, while lines of electric field are not. This is the reason for the asymmetry in Maxwell's equations, where  $\nabla \cdot \mathbf{B} = 0$  but  $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$ .

If such a magnetic monopole did exist it would, in analogy with the electric field of a point charge, have a magnetic field of the form

$$\mathbf{B} = \frac{g}{r^2} \hat{\mathbf{r}} = -g \nabla \frac{1}{r}, \quad (5.1)$$

where  $g$  is the strength of the magnetic charge. Since  $\nabla^2(1/r) = -4\pi\delta^3(r)$  we have

$$\nabla \cdot \mathbf{B} = 4\pi g \delta^3(r) \quad (5.2)$$

corresponding to a point charge, and analogous to the Maxwell equation for the electric field. Thus electricity and magnetism are treated on a completely equal footing and the symmetry between them is complete [207]. A very good elementary description of how standard electromagnetism changes if magnetic monopoles are included can be found in the article by Kocher [104].

A vector potential associated with (5.1) can taken as

$$\mathbf{A}_N = \frac{g(1 - \cos \theta)}{r \sin \theta} \hat{\phi}. \quad (5.3)$$

One point to note about the potential (5.3) is that it is not well defined over all space. It is singular along the line given by  $\theta = \pi$ . However, we know that electromagnetism is a  $U(1)$  gauge theory and thus has a certain arbitrary character to its potentials. That is, we can obtain an equally valid potential via the gauge transformation

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla \eta. \quad (5.4)$$

Consequently, letting  $\eta = -2g\phi$  we get another potential

$$\mathbf{A}_S = \frac{-g(1 + \cos \theta)}{r \sin \theta} \hat{\phi}. \quad (5.5)$$

This potential is singular along the line  $\theta = 0$  but is well defined everywhere else. Thus we see that to avoid these singularities one can cover the sphere  $S^2$

with two coordinate patches and define a non-singular vector potential in each patch. The potentials are linked by a gauge transformation in the region where the patches overlap. As the notation implies, in the present case the potentials  $\mathbf{A}_N$  are non-singular over all the surface of  $S^2$  except for the south pole  $\theta = \pi$ , and  $\mathbf{A}_S$  is non-singular over all the surface of except for the north pole  $\theta = 0$ .

In 1931 Dirac considered what would happen if such a magnetic monopole did exist, and by doing so ingeniously demonstrated that the existence of such an object would force both electric charge and magnetic charge to be quantized [60, 62]. A modern and considerably more concise method than Dirac's original proof to demonstrate this can be obtained by considering the phase factor of a particle of charge  $e$  transported in a closed path around a monopole of charge  $g$ . Suppose the monopole resides at the origin and the path taken is in the equatorial plane  $\theta = \pi/2$ . In §3.2 I showed that a charged particle acquired a phase factor when moving in an electromagnetic field that was given by

$$\exp[i\phi] = \exp\left[\frac{iq}{\hbar} \oint \mathbf{A} \cdot d\mathbf{R}\right]. \quad (5.6)$$

Now, because the potentials  $\mathbf{A}_N$  and  $\mathbf{A}_S$  are equivalent, the phase factor calculated by using either of them should be identical. Noting that the total magnetic flux from a monopole must be given by  $4\pi g$  and using Stokes' theorem we find

$$\begin{aligned} \exp\left[\frac{ie}{\hbar} \oint A_\mu^N dx^\mu\right] &= \exp\left[\frac{ie}{\hbar} \oint A_\mu^S dx^\mu\right] \\ \exp\left[\frac{ie}{\hbar} \oint (A_\mu^N - A_\mu^S) dx^\mu\right] &= 1 \\ \frac{e}{\hbar} 4\pi g &= 2n\pi \\ eg &= \frac{n\hbar}{2} \end{aligned} \quad (5.7)$$

which is the celebrated Dirac quantization condition for the U(1) monopole. It shows that were one magnetic monopole to exist anywhere, then electric charge must necessarily be quantized, as we observe.

When dealing with a monopole, two coordinate patches are, if not a necessity, a great convenience, as Wu and Yang elegantly pointed out [204]. This is because it is impossible to find a potential without the singularities noticed in (5.4) and (5.5). There must always be a line of singularities, called the Dirac string, stretching from the monopole to some point at infinity. The proof is simple: Take a closed surface enclosing the monopole and assume we have a potential that is *not* singular at any point on this surface. Then we use Stokes' theorem to split the surface in two at the equator and calculate the flux through the surface using two line integrals acting on the same potential. If the first line integral circles the equator anticlockwise for the northern hemisphere flux and the second clockwise for the southern hemisphere flux, then we must find that the total flux

is zero. This contradicts the expected value of  $4\pi g$  and consequently the surface must contain a singularity. By continuity of the potential the singularities must trace out a line to infinity.

Monopoles also have other interesting effects. For example they alter angular momentum in a peculiar way [7, 72]. Introducing a monopole to a system will shift the angular momentum spectrum by a factor of  $\frac{1}{2}\hbar$ , changing a fermionic spectrum into a bosonic one and vice versa. I will examine this effect in more detail in Chapter 7 when I consider time-odd Jahn-Teller systems and their anomalous angular momenta.

## 5.2 The $U(1)$ monopole and the geometric phase

It is possible to make a very general statement about the form of the vector potential generated by the adiabatic evolution of two-level systems dependent on some set of parameters,  $\mathbf{r}$ . Assume the two levels of the system become degenerate at some point in parameter space. Without loss of generality we can choose the point of degeneracy as  $\mathbf{r} = 0$ , and the energy at the degeneracy to be zero.

The most general Hamiltonian  $H(\mathbf{r})$  describing such a system can be represented by a  $2 \times 2$  Hermitian matrix [31]. Close to the point of degeneracy we can also choose  $H$  traceless to ensure that the two energy eigenvalues for the two states are equal and opposite. This can be justified by noting that if the form of the energy eigenvalues are given by a Taylor expansion about  $\mathbf{r} = 0$ , they will depend linearly on  $r = |\mathbf{r}|$  close to the point of degeneracy. With these conditions the most general Hamiltonian has three free parameters and takes the form

$$H(\mathbf{r}) = \begin{pmatrix} z & x - iy \\ x + iy & z \end{pmatrix}. \quad (5.8)$$

This demonstrates a result long known: for  $2 \times 2$  Hamiltonians it is necessary to vary three parameters in order to make a degeneracy occur accidentally, i.e. not on account of symmetry [193]. More generally, Mead has shown that for  $n \times n$  Hamiltonians the number of conditions required is  $n^2 - 1$  [123].

The Hamiltonian (5.8) has eigenvalues  $\lambda = \pm\sqrt{x^2 + y^2 + z^2} = \pm r$ . A pair of associated eigenvectors are given by

$$\psi_N^+ = \frac{1}{[2r(r+z)]^{\frac{1}{2}}} \begin{pmatrix} r+z \\ x+iy \end{pmatrix} \quad (5.9)$$

corresponding to the eigenvalue  $+r$ , and

$$\psi_N^- = \frac{1}{[2r(r+z)]^{\frac{1}{2}}} \begin{pmatrix} -x+iy \\ r+z \end{pmatrix} \quad (5.10)$$

corresponding to the eigenvalue  $-r$ .

The geometric vector potential  $A_i$  associated with these eigenvectors is given by  $A_i = \langle \psi | i \partial_i | \psi \rangle$ . In this case, we obtain

$$\mathbf{A}_N^\pm \equiv \langle \psi_N^\pm | i \nabla | \psi_N^\pm \rangle = \frac{\mp 1}{2r(r+z)}(-y, x, 0). \quad (5.11)$$

Equation (5.11) is a vector potential that describes a monopole of strength  $\mp \frac{1}{2}$ . This can be seen more clearly in a spherical polar coordinate system. Treating the parameters  $x, y, z$  as cartesian coordinates and recalculating the previous eigenvectors and potentials we obtain

$$\psi_N^+ = \begin{pmatrix} \cos \frac{\theta}{2} \\ e^{i\phi} \sin \frac{\theta}{2} \end{pmatrix} \quad (5.12)$$

and

$$\psi_N^- = \begin{pmatrix} -e^{-i\phi} \sin \frac{\theta}{2} \\ \cos \frac{\theta}{2} \end{pmatrix} \quad (5.13)$$

and an associated geometric vector potential

$$(\mathbf{A}_N^\pm)_\phi = \frac{\mp(1 - \cos \theta)}{2r \sin \theta}. \quad (5.14)$$

We compare with equation (5.3), and lo! a monopole, provided we take the strength as  $g = \frac{1}{2}$  and treat the coordinates in parameter space as real spatial variables. It is intriguing that the geometry of this system should throw up a field with the same functional form as a monopole, and as we shall see the analogy goes much further and is truly remarkable.

The potentials  $\mathbf{A}_N^\pm$  are evidently singular at  $\theta = \pi$ , and the corresponding eigenvectors are ill-defined at that point. As we have seen, this is a consequence of the fact that the potential for a monopole must be singular on at least one continuous line running from the monopole to infinity. To avoid this singularity one can proceed as with the Dirac case in §5.1 and cover the sphere  $S^2$  with two coordinate patches, defining a non-singular vector potential in each patch. As the notation implies, in the present case the potentials  $\mathbf{A}_N$  are non-singular over all the surface of  $S^2$  except for the south pole  $\theta = \pi$ , and thus can be freely used in the northern hemisphere. Correspondingly, one can obtain potentials which are non-singular except at the north pole  $\theta = 0$  by using the eigenvectors

$$\psi_S^+ = \frac{1}{[2r(r-z)]^{\frac{1}{2}}} \begin{pmatrix} x - iy \\ r - z \end{pmatrix} = \begin{pmatrix} e^{-i\phi} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} \end{pmatrix} \quad (5.15)$$

and

$$\psi_S^- = \frac{1}{[2r(r-z)]^{\frac{1}{2}}} \begin{pmatrix} -r + z \\ x + iy \end{pmatrix} = \begin{pmatrix} -\sin \frac{\theta}{2} \\ e^{i\phi} \cos \frac{\theta}{2} \end{pmatrix}. \quad (5.16)$$

The associated “S” potentials are

$$\mathbf{A}_S^\pm \equiv \langle \psi_S^\pm | i\nabla | \psi_S^\pm \rangle = \frac{\mp 1}{2r(r-z)}(y, -x, 0) \quad (5.17)$$

or

$$(\mathbf{A}_S^\pm)_\phi = \frac{\mp(-1 - \cos\theta)}{2r \sin\theta}. \quad (5.18)$$

$\mathbf{A}_N^\pm$  are therefore the potentials in the “northern hemisphere” patch, and  $\mathbf{A}_S^\pm$  those in the “southern hemisphere” patch. From the spherical polar forms of (5.12), (5.13), (5.15) and (5.16), we see immediately that the  $\psi_N^\pm$  are related to the  $\psi_S^\pm$  by a phase transformation

$$\psi_N^\pm = e^{\pm i\phi} \psi_S^\pm, \quad (5.19)$$

implying via equation (3.15) that  $\mathbf{A}_N^\pm$  and  $\mathbf{A}_S^\pm$  are related by the gauge transformation

$$\mathbf{A}_N^\pm - \mathbf{A}_S^\pm = \mp \nabla \phi, \quad (5.20)$$

which is consistent with (5.14) and (5.18). If we take the equator  $\theta = \pi/2$  as the overlap region between the  $N$  and  $S$  patches, we see that after a full circuit of the equator the geometrical phase  $\exp[i \oint \mathbf{A} \cdot d\mathbf{r}]$  matches smoothly (via (5.19)) from  $N$  to  $S$ , but the non-trivial nature of the gauge transformation (5.19) means that in mathematical language the  $U(1)$  bundle over  $S^2$  is non-trivial, and is indeed the monopole bundle, about which I have more to say in §5.4.

I feel this is a good point to stress the beautiful correspondence between the gauge theory of electromagnetism and the behaviour of the geometric vector potential. The link was mentioned in Chapter 3 but now we have a concrete example. In gauge theories we have the situation where different gauges correspond to the same measurable physical reality, an example of which is the northern and southern patches of the monopole vector potential, although different, yielding the same magnetic field. When the transition to the geometric potentials is made, the same result holds for a different reason. The two northern and southern potentials arise from the fact that the form we choose for the eigenvectors is to an extent arbitrary — one can choose a different basis by rephasing the old basis vectors and still preserve the physics. So the extra degree of freedom in the  $U(1)$  gauge theory of electromagnetism is mimicked by the arbitrariness of the phase in eigenvectors. This correspondence is, of course, at a deeper level due to the fact that while the vector potentials in gauge theories such as electromagnetism can be viewed as connections on a fibre bundle [170], the geometric vector potential given by  $\langle \psi | i\partial_\mu | \psi \rangle$  can also be viewed as a connection defined on a fibre bundle over a projective Hilbert space [169].

This analogy will be demonstrated in more depth in Chapter 6 where I consider nonabelian gauge theories, which are considerably richer than the  $U(1)$  case being considered here.

It should be noted that Dirac magnetic monopoles have never been experimentally detected, and there is no compelling reason to believe that they must exist. They can be included or left out of electromagnetism according to taste. As we have seen, if they did exist there would be a very nice mechanism for the quantization of charge (as Dirac said, somewhat tongue in cheek, “Under the circumstances one would be surprised if Nature had made no use of it”), but that is all. The monopole fields arising from geometric phase considerations, on the other hand, are quite real. Of course they do not correspond to actual particles, but nonetheless exhibit the same intriguing topological effects as we will see later.

### 5.3 A simple example

As we have seen, a general two-level crossing gives rise to a geometric vector potential identical to a monopole provided it is parameterized correctly. I will now give a brief example of a situation where such a situation occurs. This example is initially due to Berry, and was considered in his seminal paper on the geometric phase [31].

We consider a particle of spin  $s$  under the influence of an external magnetic field  $\mathbf{B}$ . The interaction Hamiltonian is given by

$$H(\mathbf{B}) = \kappa \hbar \mathbf{B} \cdot \mathbf{s} \quad (5.21)$$

where  $\kappa$  is a constant and  $\mathbf{s}$  is the spin operator. The associated energy eigenvalues are given by

$$E_n = \kappa \hbar B n \quad (5.22)$$

where  $n$  is the spin eigenvalue. If the system has spin  $\frac{1}{2}$  then the energy eigenvalues are  $\pm \frac{1}{2} \kappa \hbar B$ , and we see that at zero magnetic field the energy eigenspectrum is doubly degenerate. The situation is now as was considered in the previous section. We have a system that is dependent on a set of three parameters,  $\mathbf{B}$ , that can be adiabatically varied, with a point of degeneracy existing at the origin in parameter space.

Berry uses (3.13) in conjunction with some basic algebra of spin raising and lowering operators to obtain a geometric field (corresponding to the curl of the vector potential) given by

$$V_n(\mathbf{B}) = \frac{n\mathbf{B}}{B^3} \quad (5.23)$$

This is the field of a monopole, and if we choose a spin of one half, then (5.23) corresponds to the monopole generated by the general two level crossing. The

reason for this is more obvious if we do not follow Berry's derivation but rather note that Hamiltonian (5.8) can be written

$$H(\mathbf{r}) = \mathbf{r} \cdot \boldsymbol{\sigma} \quad (5.24)$$

where  $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$  are the Pauli spin matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.25)$$

To complete the comparison we note that the spin-magnetic field interaction Hamiltonian (5.21) can be written  $H(\mathbf{B}) = \frac{1}{2}\kappa\hbar\mathbf{B} \cdot \boldsymbol{\sigma}$ .

The associated geometric phase of this system is easy to calculate, as it is for all monopole systems. By Stokes' theorem the phase will simply be given by calculating the "magnetic" flux from the monopole through the closed loop of the path in parameter space. This, in turn, will simply be given by the solid angle subtended by the path at the origin, that is the point of degeneracy, multiplied by the strength of the monopole. Hence the geometric phase  $\gamma_n$  is given by

$$\exp[i\gamma_n] = \exp[-in\Omega] \quad (5.26)$$

where  $\Omega$  is the solid angle that the path subtends at  $\mathbf{B} = 0$ .

So far this example has been considered purely in the adiabatic case, but it is important to realize that the same conclusions hold if we generalize to the nonadiabatic regime. This has been accomplished by Anandan who has considered the example of a neutron in an external magnetic field that can be varied [15]. As we have seen the geometric phase is given, up to a constant, by the angle swept out by the magnetic field, at least in adiabatic evolutions. But what happens if the field is varied rapidly so at any time the neutron spin is not necessarily parallel or antiparallel to the field?

It is possible to consider this situation using the idea of spin-space rather than the magnetic field parameter space. We know from §3.3 that if the same path is traced out through Hilbert space then the same geometric phase must be generated. This is the essence of the nonadiabatic formalism. Consequently, even though the magnetic field is not parallel to the neutron spin, provided the spin moves through the same path as when the field was slowly varied the same geometric phase will be generated. From this we can see that instead of looking at the path the magnetic field traces out we can instead look at the solid angle the spin vector traces out on the spin sphere, which will give the same result and generalize to nonadiabatic evolutions.

This viewpoint relates back to §3.5, where the geometric phase was linked to the holonomy of parallel transport. In the present example we are parallel transporting a vector around a path on the space of spins, which is topologically a sphere. By the Gauss-Bonnet theorem the rotation of a vector undergoing

parallel transport is equal to the surface integral of the curvature of the area enclosed by the path. For a sphere, which has constant curvature, this is simply equivalent to the solid angle subtended by the path, which is the result we obtained in the adiabatic case using the gauge-field formalism. Thus we see that, as always, the adiabatic and nonadiabatic approaches are linked through geometry. It is important to realize that the generation of a monopole is due to the geometry of the problem, and not what the parameters are or how they are varied. This example of the nonadiabatic evolution of a neutron in an external time-varying magnetic field is an interesting one, and has been considered in more detail by other workers. See, for example, Zhang *et al.* [210, 209] and the comment of Wagh *et al.* on their result [194].

As was demonstrated in Chapter 4 the existence of a geometric phase can be seen as always being due to some form of time reversal symmetry breakdown. It is instructive to consider this result within the context of the geometric monopole of this chapter.

For the case of a spin in an external magnetic field the cause of the time reversal violation is easy to find: the Hamiltonian (5.21) is explicitly time-odd due to the presence of the time-odd spin operator appearing as a linear term, and consequently the system falls into Case 5 of Chapter 4, that is explicitly broken time reversal symmetry. I note that this operator is coupled with a time odd magnetic field, but it is important to realize that this term is *external* and we are interested in the time reversal character of the particle subsystem. If the time reversal operation is applied to the entire system, both  $\mathbf{B}$  and  $\mathbf{s}$  will reverse leaving the Hamiltonian invariant, which is to be expected due to the time reversal invariance of electrodynamics.

The case of the general two-level crossing is a little more complicated. Admittedly the general Hamiltonian (5.8) governing such a crossing contains complex elements and consequently it is tempting to claim that since the time reversal operator involves complex conjugation this Hamiltonian is not time reversal symmetric. However it must be borne in mind that (5.8) is the matrix representation of the Hamiltonian, and it is not obvious that the time reversal operator in such a representation includes complex conjugation. To demonstrate that the two-level crossing still falls under Case 5 I appeal to a result derived by Mead [123]: Given a general  $2 \times 2$  time-odd Hamiltonian one must impose three conditions on the Hamiltonian in order to ensure a non-accidental (i.e. not due to symmetry) level crossing and ensure a point of degeneracy. For our purposes this means that the Hamiltonian must have at least three degrees of freedom. Applying the condition of hermiticity and choosing the Hamiltonian to be traceless the only possible form it can take in this case is the one given by (5.8). Thus this is indeed the archetypal form for any two-level time-odd Hamiltonian.

It is also interesting to note what happens when the Hamiltonian (5.8) is made real by reducing the parameter space by one dimension, i.e. take  $H = x\sigma_1 + z\sigma_3$ .

Since the Hamiltonian is now certainly time reversal symmetric we expect phases that are multiples of  $\pi$ . Calculation of the single-valued eigenvectors and the associated geometric vector potential is easily done [7]. The potential is now singular and in polar coordinates is given by

$$\mathbf{A} = \frac{1}{2}\nabla\phi. \quad (5.27)$$

This is the equation of a flux tube, that is the curl of this potential yields a line singularity of “magnetic” flux along the  $z$  axis. Using Stokes’ theorem it is simple to see that any path encircling the flux tube will yield a phase of  $n\pi$ , where  $n$  is the winding number of the path, as defined in §4.4. This falls under Case 3, time reversal symmetric with a singular potential. Finally, it should be noted that it is possible to eliminate the singularity by going to a higher-dimensional parameter space but retaining the same paths [93]. Thus instead of the potential (5.27) one can use a monopole potential, and providing all paths remain the  $x$ - $z$  plane any phases generated will remain multiples of  $\pi$ . We have thus changed to Case 5, and now ascribe the phase to an the explicitly broken time reversal symmetry of the Hamiltonian.

## 5.4 Monopole mathematics

The purpose of this section is to examine the monopole in a more mathematical way, partly in order to develop results for the simpler abelian case before considering the nonabelian gauge theory in the next chapter, and partly to draw out the parallels between the electromagnetic Dirac monopole and the geometric monopole constructed in this chapter.

To this end it is useful to consider matters from the point of view of fibre bundle theory. As shown in §3.5 a gauge theory can be seen in terms of fibre bundles and differential geometry, with the connection on the bundle being given by the one-form  $A_\mu dx^\mu$  and corresponding to the gauge potential, and the associated field tensor being given by the two-form  $F_{\mu\nu} dx^\mu \wedge dx^\nu$ , namely the exterior derivative of the gauge potential.

In the case of electromagnetism the gauge group is  $U(1)$ . The base space of this  $U(1)$  bundle is usually taken to be some flat manifold such as  $\mathbb{R}^4$  corresponding to Minkowski space-time or  $\mathbb{R}^3$  in nonrelativistic electromagnetism. Now, it is a well known result that if the base space of a bundle is contractible then the bundle is trivial, namely it can be globally described as the product of two spaces [179]. This is *not* the case when dealing with monopoles however. In this case the base space is not  $\mathbb{R}^3$  but rather  $\mathbb{R}^3 - \{0\}$  ( $\mathbb{R}^3$  with the point at the origin removed), which is no longer contractible and hence the monopole bundle is not trivial.

This gives rise to a deeper and more topological way of understanding the quantization of monopole charge that I derived in §5.1. A Dirac monopole sitting at

the origin in  $\mathbb{R}^3$  can be seen as a principal  $U(1)$  bundle  $P$  over the base space  $\mathbb{R}^3 - \{0\}$ . But  $\mathbb{R}^3 - \{0\}$  can be retracted to the sphere  $S^2$  without changing the topology of the bundle, and thus  $P$  is equivalent to a  $U(1)$  bundle over  $S^2$ . It is thus classified by elements of the homotopy class  $\pi_1(U(1)) = \pi_1(S^1) = \mathbb{Z}$ , i.e. integers, and as the connection on the bundle corresponds to the electromagnetic vector potential this is quite suggestive. To make the link between quantization and monopoles concrete we turn to cohomology classes and characteristic class numbers. This way of looking at things, one which I will use in Chapter 6, exposes the quantization at a more obvious level, dealing as it does with the actual field strengths and potentials.

Characteristic classes are cohomology classes, and their use to us lies in the fact that they classify principal bundles, and that they can be defined using the connection on a bundle [135]. From a physicist's point of view this means that it is possible to divide various solutions to the gauge theory into classes. There are several types of characteristic classes, for example Chern, Pontrjagen, Euler and so on. The classes we will be interested in are Chern classes, as these refer to principal bundles with associated group  $U(n)$ . In terms of the curvature form  $\mathbf{F} = \frac{1}{2}F_{\mu\nu}^i X_i dx^\mu \wedge dx^\nu$ , where  $X_i$  are the generators of the gauge group, the Chern classes are the coefficients of the polynomial in  $t$  arising from the expansion of

$$\det \left[ tI + \frac{i}{2\pi} \mathbf{F} \right] = t^n + c_1(\mathbf{F})t^{n-1} + \dots + c_n(\mathbf{F}). \quad (5.28)$$

What is of interest to us is that the  $c_i$  actually determine integral cohomology classes, i.e. elements of  $H^{2i}(X; \mathbb{Z})$ , where  $X$  is the base space. In fact this integer can be related to the homotopy class [171]. Thus the integral of  $c_i$  over  $X$  yields an integer. The fact that the cohomology class must be integral is more obvious if we note that integrating the Chern class number is almost identical to using the Gauss-Bonnet theorem to show the total integral of the local curvature of a manifold is a multiple of  $2\pi$  [154].

For the case of the  $U(1)$  monopole we find that all the Chern class numbers are zero apart from  $c_1$  which is given by [89]

$$C_1 = \int_{S^2} \frac{\mathbf{F}}{2\pi}. \quad (5.29)$$

The one-form corresponding to the monopole potential is

$$\mathbf{A} = \frac{g}{r(r+z)}(-ydx + xdy) \quad (5.30)$$

and the associated curvature two-form is given by the exterior derivative:

$$\mathbf{F} = d\mathbf{A} = \frac{g}{r^3}(x dy \wedge dz + y dz \wedge dx + z dx \wedge dy). \quad (5.31)$$

Finally, in calculating the first Chern class number, we have to take note of the fact that when describing the electromagnetic potential as a connection, it will

differ by a factor of  $e/\hbar$  from the pure geometric form due to the existence of the units used to describe the physics. Thus we have

$$\begin{aligned}
 C_1 &= \frac{e}{2\pi\hbar} \int_{S^2} \frac{\mathbf{F}}{2\pi} \\
 &= \frac{eg}{2\pi\hbar} \int_{S^2} \frac{1}{r^3} (x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}) \cdot \hat{\mathbf{n}} dS \\
 &= \frac{eg}{2\pi\hbar} \iint \frac{1}{r^2} \sin\theta d\theta d\phi \\
 &= \frac{2eg}{\hbar} \tag{5.32}
 \end{aligned}$$

which must be an integer, demonstrating the quantization of the monopole strength.

It is natural to wonder to what extent this quantization rule carries over to the geometric monopole. The geometric gauge potential can be seen as a  $U(1)$  connection on a principal bundle over a base space that is a projective Hilbert space, so the same arguments should certainly apply. However when calculating the geometric phase factor there is no factor of  $e/\hbar$  in the exponential, unlike phase factors in electromagnetism. Consequently we obtain the simple quantization condition

$$g = \frac{n}{2} \tag{5.33}$$

where  $g$  is the strength, or magnetic charge, of the geometric monopole. The two-level crossing discussed in §5.2, for example, has a strength of  $g = \frac{1}{2}$ , corresponding to the case  $n = 1$ .

A more interesting example is the one considered in §5.3, which consisted of a particle of spin  $S$  interacting with an external magnetic field. This was shown to give rise to a monopole of strength  $s$ , where  $s$  is the spin eigenvalue of the particle projected along the magnetic field direction. Coupling this with the quantization rule developed from topological considerations we obtain

$$s = \frac{n}{2}. \tag{5.34}$$

Thus we have proved the well-known result that spin must be a multiple of one-half, at least when interacting with a magnetic field, using purely topological arguments. Assuming that spin is intrinsic to a particle and not altered by the external field, one can draw the extended conclusion that spin *always* comes in integer multiples of  $\hbar/2$ . While the fact that geometric monopole strengths are related to spins has been known since Berry's seminal paper [31], no one appears to have noticed this beautiful result.

Finally, a note on Hopf fibration. Throughout this chapter I have been using two coordinate patches to describe the monopole potential. As noted in §5.1 this

is because any three-dimensional potential has a line of singularities stretching from the origin to infinity. It is, however, possible to remove these singularities by going to a higher dimensional space via a mathematical technique called Hopf fibration. Thus instead of defining the monopole potentials on a the sphere  $S^2$  one defines them on the sphere  $S^3$ , via the Hopf map  $\eta : S^3 \rightarrow S^2$ , giving a set of regular, sourceless potentials [131, 154]. The mathematical reason behind this is that the total space of a  $U(1)$  bundle over  $S^2$  has the same topology as the 3-sphere  $S^3$ . Since the group  $U(1)$  has the same topology as the circle  $S^1$  a picture of this fibre bundle is given by the sequence

$$S^1 \rightarrow S^3 \rightarrow S^2 \quad (5.35)$$

corresponding to fibering spheres with spheres. The other two known cases of this are [89]

$$S^3 \rightarrow S^7 \rightarrow S^4 \quad (5.36)$$

where  $S^3$  has the topology of the group  $SU(2)$  and represents a principal bundle over  $S^4$  and

$$S^7 \rightarrow S^{15} \rightarrow S^8. \quad (5.37)$$

Equation (5.36) will be applicable when I consider the  $SU(2)$  monopole and instanton in the next chapter.

This technique of Hopf fibration can be applied to the monopole arising from the geometric phase as well. Aitchison has shown how to obtain the singularity-free fields on  $S^3$  using the Balachandran Lagrangian formalism and the spinors (5.12, 5.13) [8]. He also demonstrates what the Hopf map is in terms of these spinors, giving a nice demonstration of how the above esoteric (from the point of view of most physicists, anyway) mathematics translates to physical reality.

## 5.5 Time reversal and the monopole

This last section is somewhat speculative, but does open up the possibility of a wonderfully symmetric relationship between time reversal symmetry and the monopole. The central point I wish to consider is, what happens to a magnetic monopole under the operation of time reversal? Consider one of Maxwell's equations if we allow magnetic monopoles to exist:

$$\nabla \cdot \mathbf{B} = 4\pi g \rho_m \quad (5.38)$$

Normally all Maxwell's equations are invariant under the transformation  $\mathbf{B} \rightarrow -\mathbf{B}$  and  $t \rightarrow -t$ , demonstrating their time reversal symmetry. However, we see that (5.38) is not invariant under this transformation, demonstrating that electromagnetism appears to violate time reversal symmetry if we allow monopoles. What to do?

The simplest choice is just to treat the monopole field as any other magnetic field, and simply reverse its direction. This is equivalent to letting the monopole charge change sign under time reversal. The advantage to deciding that the magnetic charge will simply reverse under time reversal is that since  $g \rightarrow -g$  equation (5.38) becomes time reversal symmetric, and so does electromagnetism.

However, this choice is not quite as clear cut as it seems. In ordinary electromagnetism *sans* monopoles time reversal is equivalent to motion reversal, and since it is the motion of electric charges that generate magnetic fields motion reversal will obviously cause magnetic fields to change sign. A magnetic monopole, however, has its charge as an intrinsic property and thus, avoiding questions on internal structure, motion reversal will not change the sign of its charge. So letting time reversal change the sign of a monopole requires the abandonment of the identification of time reversal with motion reversal even in classical physics. A few remarks considering the consequences of assuming the magnetic monopole is invariant under time reversal and thus ensuring that Maxwell's equations are not time reversal symmetric have been made by Schiff [79].

The reason this speculation is relevant is because if we let monopoles remain invariant under time reversal, it would give rise to a nice symmetry between time reversal and the monopoles of electromagnetism and those arising in the context of the geometric phase. We have seen that the archetypal Hamiltonian (5.8) with broken time reversal symmetry gives rise to a monopole-like geometric vector potential; the idea that electromagnetic monopoles also imply broken time reversal symmetry would make for an intriguing result. Note that the link here is stronger than it might appear at first glance. For example depending on equation (5.8) is not necessary. This is because a monopole must have a source and this must be point of singularity within any field description. As shown in Chapter 4 a singularity can only appear in the potential due to the breakdown of time reversal symmetry.

Whether this similarity between the two types of monopole is coincidental or reflects some deeper topological result stemming from their identical gauge description is not obvious. Nonetheless, it may be worth exploring this link in more detail.

## 6. The nonabelian monopole

This chapter is primarily an examination of the wonderfully rich topological consequences that follow from very simple beginnings, those of the geometric phase and two Kramers doublets. Along the way we will encounter nonabelian gauge fields, large numbers of subscripts and a symphony of exotic coordinate systems. But it's all fun, and some beautiful and quite surprising results struggle out of the messy metrics.

Most of the ideas in this chapter have been described in the paper by Ian Aitchison and myself [97]. The basic motivation was originally as follows: it is known that a pair of levels becoming degenerate at a point can give rise to a geometric vector potential identical to that of a monopole, provided a suitable parameterization is chosen. Is this monopole-like behaviour a function of all types of degeneracies? What would happen, for example, if two sets of degenerate levels became hyper-degenerate at a point? To begin with the vector potential would become nonabelian, of course, and this makes it harder to decide on what a monopole potential should look like.

If we are to investigate, with a view to a monopole perspective, a pair of degenerate levels crossing, the obvious choice is to consider two pairs of Kramers doublets. Firstly because as we have seen in Chapter 4 time reversal symmetry has an important role in the existence of non-trivial geometric phases; and secondly a Kramers doublet that is split by a magnetic field gives rise to an abelian monopole potential as we saw in §5.3. The geometric phases associated with pairs of Kramers doublets have been previously examined in the literature by several people, for example Mead and Koizumi *et al.* within the context of molecular physics [107, 126], and by Avron *et al.* [25] and Levay [116] who analysed the topology and geometry of such systems using differential geometry and a quaternionic formalism.

The purpose of this chapter, then, is to examine what happens when a pair of Kramers doublets is driven by some set of adiabatically varying parameters such that there is a point in parameter space where both pairs of levels become hyper-degenerate.

To do this I first require the Hamiltonian governing the evolution of the doublets. I derive this in §6.2 using time reversal considerations and demonstrate that the parameter space must be five-dimensional, unlike the more intuitive three-space that is required by the abelian monopole. Then in §6.3 I calculate eigenvectors

of the Hamiltonian, and then using these find the associated geometric vector potentials in both cartesian and spherical polar coordinates. The potentials thus found have  $SU(2)$  symmetry and are very similar those of the abelian monopole.

In §6.4 I first go over what one might expect a nonabelian monopole to look like, drawing on the work of Yang [205] who explicitly constructed two  $SU(2)$  monopoles by analogy from the abelian case. Interestingly, Yang's monopole also exists only in five dimensions, although he uses an peculiar set of projective coordinates to describe it. In the remainder of the section I recalculate my pair of vector potentials corresponding to the higher and lower energy Kramers doublets using Yang's coordinate system and identify them exactly with Yang's two monopoles. There is another way of looking at these fields, and that is from the perspective of a topological entity known as the instanton [29]. In §6.5 I explain what an instanton is, and show how it is possible, via yet another coordinate system (involving the projection of the cartesian monopole fields onto the four-dimensional hypersphere) to identify it with my Kramers doublets potentials. This connection between time reversal invariant Fermi systems and the instanton was first realized by Avron *et al.* [25].

As I spend most of the chapter constructing a nonabelian monopole and examining its topological consequences I include §6.6, which explains whether or not monopoles are actually likely to exist, and briefly touches on the topic of duality to demonstrate the subtlety of the concept of a monopole.

Finally in §6.7 I look at two cases in molecular physics where the situation considered in this chapter, that is a pair of colliding Kramers doublets, actually occurs, and a geometric vector potential corresponding to a monopole/instanton can actually be generated.

This chapter relies on nonabelian gauge theory, and so it is appropriate to include a very brief review of how abelian theory extends to the Yang-Mills case. This is the subject of the next section.

## 6.1 Nonabelian gauge theory

Nonabelian gauge fields are the stuff of this chapter, and it will be helpful to have a brief overview of how they behave and how nonabelian gauge fields differ from abelian ones. I generally consider only classical rather than quantized fields, although most results are immediately transferable.

It is possible to develop gauge theories from phase parallel transport considerations, and this is the approach I take, mostly following Chan and Tsou [89].

I first consider the abelian case. We have some field  $\psi(x)$  defined over a manifold, and we need a way to compare the phase of  $\psi(x)$  at different points. To do this we introduce a gauge field  $A_\mu$ , which specifies how the phase changes in each

direction at each point (looking at things this way makes it obvious that the gauge field is a connection). The phase of the wavefunction of a particle of charge  $e$  at point  $x$  is said to be parallel to the phase at the point  $x + dx$  if their phases differ by  $eA_\mu dx^\mu$ . Now, in order to give this definition an invariant meaning under local phase transformations  $A_\mu$  must transform in a certain way. If we perform a local gauge transformation such that

$$\psi(x) \rightarrow \psi'(x) = e^{ie\eta(x)}\psi(x) \quad (6.1)$$

then the phase change at  $x$  is  $e\eta(x)$  but at  $x + dx^\mu$  it is  $e(\eta(x) + \partial_\mu\eta(x))dx^\mu$ . To make the definition of parallelism the same as before the transformation we need to make  $A_\mu$  transform as

$$A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) + \partial_\mu\eta(x). \quad (6.2)$$

Using these definitions we see that the phase change  $\phi$  on transporting a particle along some path  $P$  from  $x_1$  to  $x_2$  is just

$$\phi = e \int_P A_\mu dx^\mu. \quad (6.3)$$

By Stokes' theorem transport around a closed path will yield a total phase change of

$$e \iint F_{\mu\nu}(x) d\sigma^{\mu\nu} \quad (6.4)$$

where  $\sigma^{\mu\nu}$  is the area element of some surface bounded by the path  $P$  and

$$F_{\mu\nu}(x) = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (6.5)$$

is the field tensor. If  $F_{\mu\nu}$  is considered as a measure of curvature of the area enclosed by the path  $P$ , the link to rotation of a vector undergoing parallel transport via the Gauss-Bonnet theorem as mentioned in §3.5 is obvious.

I now turn to the nonabelian case. Here the field or wavefunction we are considering has several internal degrees of freedom. By a change of phase I now mean a mixing of these internal degrees of freedom, namely

$$\psi(x) \rightarrow \psi'(x) = S\psi(x) \quad (6.6)$$

where  $S$  is a matrix. Once again, in order to describe the notion of parallel transport a gauge field is introduced, except this time it is matrix-valued:  $A_\mu = A_\mu^i X_i$ . The  $X_i$  are the generators of the group under which  $A_\mu$  is supposed to be invariant. For example if we have an  $SU(2)$  gauge theory  $X_i = -\frac{i}{2}\sigma_i$ .

The phase of a particle of charge  $e$  at point  $x$  is said to be parallel to the phase at the point  $x + dx$  if their phases differ by  $eA_\mu dx^\mu$ .

Analogous to the abelian case we define the phase of  $\psi(x)$  to be parallel to the phase of  $\psi(x+dx)$  if they differ by  $gA_\mu dx^\mu$ , where  $g$  is the charge of the particle. In order to impose local gauge invariance we require  $A_\mu$  to transform as [89]

$$A_\mu(x) \rightarrow A'_\mu(x) = S(x)A_\mu(x)S^{-1}(x) - \frac{i}{g}(\partial_\mu S(x))S^{-1}(x). \quad (6.7)$$

To obtain the definition of the field tensor consider Figure 6.1. Starting at  $A$  and parallel transporting  $\psi(x)$  to point  $C$  via  $B$  we have

$$\psi_{ABC} = \exp[igA_\nu(x+dx^\mu)dx^\nu] \exp[igA_\mu(x)dx^\mu] \psi(x) \quad (6.8)$$

and transporting via  $D$  we have

$$\psi_{ADC} = \exp[igA_\mu(x+dx^\nu)dx^\mu] \exp[igA_\nu(x)dx^\nu] \psi(x). \quad (6.9)$$

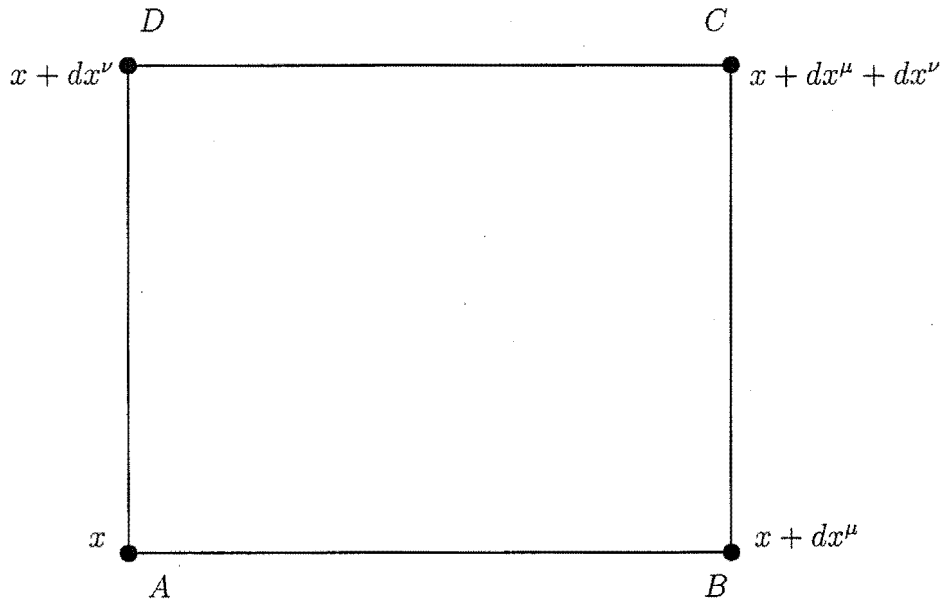


Fig. 6.1: Parallel transport around a closed loop

Expanding (6.8) and (6.9) to first order in  $dx^\mu dx^\nu$  and taking their difference we get

$$ig(\partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + (ig)^2[A_\nu(x), A_\mu(x)])dx^\mu dx^\nu \quad (6.10)$$

where the square brackets denote the commutator. As in the abelian case we define  $gF_{\mu\nu}dx^\mu dx^\nu$  to be the difference in phase obtained by parallel transporting along the paths  $ABC$  and  $ADC$  and hence find

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) - ig[A_\mu(x), A_\nu(x)]. \quad (6.11)$$

Unlike the abelian case the field tensor also transforms under a gauge transformation:

$$F_{\mu\nu}(x) \rightarrow F'_{\mu\nu}(x) = S(x)F_{\mu\nu}(x)S^{-1}(x). \quad (6.12)$$

It should be noted that the above derivation of the field tensor and behaviour of  $A_\mu$  under a gauge transformation is more often derived a different way, using a Lagrangian approach [72]. To do this one begins with the free-field Lagrangian that exhibits some global symmetry. The demand is then made that this symmetry be made local, that is position dependent. In order for this to work it is necessary to introduce an extra massless vector field  $A_\mu$ , and this extra field transforms exactly as in (6.7). In some ways I feel this approach is more beautiful, as it highlights the astonishing fact that merely demanding local symmetry accurately predicts the existence of other fields and also correctly specifies their interaction and free field terms.

However this approach stresses the gauge fields and field tensors as more important than the phase factors themselves. Not only is a phase factor approach more suitable to geometric phases, but the phase factor is actually the more fundamental object. This is because  $A_\mu$  overdescribes the situation since different  $A_\mu$  related by a gauge transformation describe the same physics, and  $F_{\mu\nu}$ , while the correct quantity in the classical abelian case, underdescribes the physics in quantum mechanics (for example the Aharonov-Bohm effect). When dealing with nonabelian gauge fields,  $F_{\mu\nu}$  can actually underdescribe the physics even in the classical case — explicit examples exist of gauge-inequivalent  $A_\mu$  that have the same  $F_{\mu\nu}$  [204]. The phase factor on the other hand is what any physical experiment will actually measure, it is the truly fundamental quantity. There is no ambiguity; it does not over- or underdescribe the physics.

Another reason that the phase factor is more useful is that in nonabelian gauge theories there exists no useful analogue to the concept of flux. This can be seen as follows: when dealing with abelian electromagnetic theory, provided there are no monopoles then we have  $\nabla \cdot \mathbf{B} = 0$ . This is a consequence of the more general Bianchi identity which states

$$\partial^\nu *F_{\mu\nu} = 0 \quad (6.13)$$

which holds for any abelian gauge field derivable from a potential as in (6.5). That is, only if a closed surface contains a monopole, and (6.5) fails along the line of singularities that is the Dirac string, can (6.13) be nonzero, and by Stokes theorem generate a net flux through the surface. Things are different when dealing with a nonabelian gauge field, however. An equivalent to the Bianchi identity still exists, but it uses the covariant derivative and becomes [89]

$$D^\nu *F_{\mu\nu} = 0 \quad (6.14)$$

where

$$D_\mu = \partial_\mu - ig[A_\mu, \ ]. \quad (6.15)$$

Now, if we use the abelian approach and define flux as

$$\text{flux} = \iint F_{\mu\nu} d\sigma^{\mu\nu} \quad (6.16)$$

then this will still be equivalent to the integral of the divergence  $\partial^\nu {}^*F_{\mu\nu}$  and by (6.14) this need not be zero even if the absence of an enclosed monopole. Thus effectively we have a “flux” even though there is no source inside the surface to create it. There have been attempts to create some nonabelian analogue to flux but essentially they all lead back to the concept of the phase factor [89].

## 6.2 The two-Kramers doublet Hamiltonian

Kramers doublets arise in time reversal symmetric systems with half-integral spin. If we denote the time reversal operator by  $T$ , this corresponds to the case where  $THT^{-1} = H$  and  $T^2 = -1$ . In this case every state is doubly degenerate with another, and the two states are related via time reversal:

$$T|\phi\rangle = |\bar{\phi}\rangle \quad (6.17)$$

and thus

$$T|\bar{\phi}\rangle = T^2|\phi\rangle = -|\phi\rangle. \quad (6.18)$$

In the construction of the geometric vector potential associated with two Kramers doublets I shall proceed in a manner similar to that which I used for the general two-level crossing in §5.2. In order to use this method to calculate the geometric vector potential it is first necessary to obtain the most general Hamiltonian describing such a system. It turns out that it is possible to do this from purely time reversal symmetry considerations.

To construct the matrix representation of the Hamiltonian we will need its matrix elements both between the basis vectors within a doublet and between different doublets. I will denote the basis vectors of the first doublet as  $|\phi\rangle, |\bar{\phi}\rangle$  and the second as  $|\psi\rangle, |\bar{\psi}\rangle$  where the overbar represents time reversal. Now, using time reversal symmetry and noting that the operation of time reversal on a matrix element is equivalent to complex conjugation we have

$$\langle\phi|H|\phi\rangle = \langle T\phi|THT^{-1}|T\phi\rangle^* = \langle T\phi|H|T\phi\rangle^* = \langle\bar{\phi}|H|\bar{\phi}\rangle \quad (6.19)$$

since diagonal matrix elements of hermitian operators are always real. Similarly

$$\begin{aligned} \langle\phi|H|\bar{\phi}\rangle &= \langle\phi|H|T\phi\rangle \\ &= \langle T\phi|THT^{-1}|T^2\phi\rangle^* \\ &= \langle T\phi|H|T^2\phi\rangle^* \\ &= -\langle\bar{\phi}|H|\phi\rangle^*. \end{aligned} \quad (6.20)$$

To proceed further we need the result that the matrix element of an arbitrary operator satisfies

$$\langle i|O|j\rangle^* = \langle j|O^\dagger|i\rangle. \quad (6.21)$$

Using this and the fact that  $H$  is hermitian we have

$$\langle \phi|H|\bar{\phi}\rangle = \langle \bar{\phi}|H^\dagger|\phi\rangle^* = \langle \bar{\phi}|H|\phi\rangle^*. \quad (6.22)$$

Combining (6.20) and (6.22) we obtain the result

$$\langle \phi|H|\bar{\phi}\rangle = \langle \bar{\phi}|H|\phi\rangle = 0. \quad (6.23)$$

Turning to matrix elements between different doublets we have

$$\langle \phi|H|\psi\rangle = \langle T\phi|THT^{-1}|T\psi\rangle^* = \langle T\phi|H|T\psi\rangle^* = \langle \bar{\phi}|H|\bar{\psi}\rangle \quad (6.24)$$

and

$$\langle \phi|H|\bar{\psi}\rangle = \langle T\phi|THT^{-1}|T\bar{\phi}\rangle^* = \langle T\phi|H|T\bar{\phi}\rangle^* = -\langle \bar{\phi}|H|\psi\rangle. \quad (6.25)$$

Within the basis  $\{|\phi\rangle, |\bar{\phi}\rangle, |\psi\rangle, |\bar{\psi}\rangle\}$  equations (6.19), (6.23), (6.24) and (6.25) define the Hamiltonian for the pair of Kramers doublets up to a set of five undetermined parameters. Specifically

$$H = \begin{pmatrix} q_5 & 0 & q_3 + iq_4 & q_1 + iq_2 \\ 0 & q_5 & -q_1 + iq_2 & q_3 - iq_4 \\ q_3 - iq_4 & -q_1 - iq_2 & -q_5 & 0 \\ q_1 - iq_2 & q_3 + iq_4 & 0 & -q_5 \end{pmatrix}. \quad (6.26)$$

Notice the more than passing similarity to the  $2 \times 2$  Hamiltonian (5.8) describing the two-level crossing in Chapter 5 which was shown to lead to a geometric vector potential describing a  $U(1)$  monopole.

### 6.3 The geometric vector potential

The next step in the calculation of the geometric gauge potential associated with this pair of Kramers levels is to find the eigenvectors of (6.26). We find, as expected, that there are two sets of degenerate eigenvectors. Defining a pseudo-radial coordinate

$$q = \sqrt{q_1^2 + q_2^2 + q_3^2 + q_4^2 + q_5^2} \quad (6.27)$$

we have a pair of eigenvectors

$$\psi_1^+ = \frac{i}{\sqrt{2q(q - q_5)}} \begin{bmatrix} q_1 + iq_2 \\ q_3 - iq_4 \\ 0 \\ q - q_5 \end{bmatrix}, \quad \psi_2^+ = \frac{-i}{\sqrt{2q(q - q_5)}} \begin{bmatrix} q_3 + iq_4 \\ -q_1 + iq_2 \\ q - q_5 \\ 0 \end{bmatrix} \quad (6.28)$$

with the associated eigenvalue  $+q$ , and another pair

$$\psi_1^- = \frac{i}{\sqrt{2q(q+q_5)}} \begin{bmatrix} -q_1 - iq_2 \\ -q_3 + iq_4 \\ 0 \\ q + q_5 \end{bmatrix}, \quad \psi_2^- = \frac{-i}{\sqrt{2q(q+q_5)}} \begin{bmatrix} -q_3 - iq_4 \\ q_1 - iq_2 \\ q + q_5 \\ 0 \end{bmatrix} \quad (6.29)$$

with the associated eigenvalue  $-q$ .

I now choose the convention (see discussion in §3.4) that defines the gauge potential as

$$A_\mu^{ij} = \langle \psi_i | i\partial_\mu | \psi_j \rangle \quad (6.30)$$

where in this case the index  $\mu$  runs from 1 to 5 and the derivative operator acts on the  $q_\mu$ . Inserting (6.28) and (6.29) into (6.30) and performing a truly Herculean amount of algebra, one obtains

$$A_\mu^\pm = \frac{1}{2q(q \mp q_5)} \begin{bmatrix} -q_4\sigma_1 - q_3\sigma_2 + q_2\sigma_3 \\ q_3\sigma_1 - q_4\sigma_2 - q_1\sigma_3 \\ -q_2\sigma_1 + q_1\sigma_2 - q_4\sigma_3 \\ q_1\sigma_1 + q_2\sigma_2 + q_3\sigma_3 \\ 0 \end{bmatrix} \quad (6.31)$$

where the first row on the right hand side of (6.31) gives the matrix for  $A_1^\pm$  and so on, ending with  $A_5^\pm = 0$ , and the  $\sigma_i$  are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (6.32)$$

Equation (6.31) certainly bears more than a passing resemblance to the cartesian form of the Dirac monopole (5.11), except that now the gauge potential is SU(2) rather than U(1). It has the right  $1/q$  dependence and a very similar prefactor. Just how monopole-like is (6.31), really? Things become a little more obvious in spherical polars. Define

$$\begin{aligned} q_1 &= r \sin \theta \sin \phi \sin \eta \sin \gamma \\ q_2 &= r \sin \theta \sin \phi \sin \eta \cos \gamma \\ q_3 &= r \sin \theta \sin \phi \cos \eta \\ q_4 &= r \sin \theta \cos \phi \\ q_5 &= r \cos \theta \end{aligned} \quad (6.33)$$

which has the metric

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 + r^2 \sin^2 \theta \sin^2 \phi d\eta^2 + r^2 \sin^2 \theta \sin^2 \phi \sin^2 \eta d\gamma. \quad (6.34)$$

If the eigenvectors (6.28) and (6.29) are rewritten in terms of these new five-dimensional spherical polars and inserted into (6.30) we obtain

$$A^+ = \frac{1 + \cos \theta}{2r \sin \theta} \left[ \begin{array}{c} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} -\cos \eta & -ie^{i\gamma} \sin \eta \\ ie^{-i\gamma} \sin \eta & \cos \eta \end{pmatrix} \\ \begin{pmatrix} \cos \phi \sin \eta & -ie^{i\gamma}(\cos \eta \cos \phi - i \sin \phi) \\ ie^{-i\gamma}(\cos \eta \cos \phi + i \sin \phi) & -\cos \phi \sin \eta \end{pmatrix} \\ \begin{pmatrix} \sin \phi \sin \eta & -ie^{i\gamma}(\cos \eta \sin \phi + i \cos \phi) \\ ie^{-i\gamma}(\cos \eta \sin \phi - i \cos \phi) & -\sin \phi \sin \eta \end{pmatrix} \end{array} \right] \quad (6.35)$$

$A^-$  is identical aside from the prefactor being given by  $(1 - \cos \theta)/\sin \theta$ .

Once again, this is similar to the polar coordinate potentials (5.18) and (5.14) describing the  $U(1)$  monopole arising from the two-level crossing. As well as an identical prefactor, (6.35) also has both its  $r$ - and  $\theta$ -components vanishing.

Is this similarity enough to justify labelling these potentials as an  $SU(2)$  monopole? Do these potentials exhibit the interesting topological properties we have seen in the  $U(1)$  monopole? It is to these questions that I now turn.

## 6.4 Yang's $SU(2)$ monopole

We must first be clear on what an  $SU(2)$  monopole would actually look like. In 1978 Yang considered just this problem [205]. He tried to construct an  $SU(2)$  gauge field by extrapolating the properties of the  $U(1)$  monopole to the non-abelian case.

The first obvious thing to do is to note that for the  $U(1)$  monopole there exists an invariant, non-zero flux for all surfaces enclosing the monopole. This can be written as

$$\frac{1}{2} \int_{S^2} F_{\mu\nu} dx^\mu \wedge dx^\nu = 4\pi g \quad (6.36)$$

for a monopole of strength  $g$ . This is just the appearance of the first, and only non-zero, Chern class number as described in §5.4. When calculating Chern class numbers for an  $SU(2)$  gauge field we find that only the second class number is

non-zero and is given by

$$\begin{aligned} C_2 &= \frac{1}{16\pi^2} \int_{S^4} \mathbf{F}^i \wedge \mathbf{F}^i \\ &= \frac{1}{64\pi^2} \int_{S^4} F_{\mu\nu} F_{\alpha\beta} dx^\mu \wedge dx^\nu \wedge dx^\alpha \wedge dx^\beta \end{aligned} \quad (6.37)$$

since the curvature two-form is defined as  $\mathbf{F} = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu$ . Thus, one would expect an SU(2) monopole to have a non-zero second Chern class number.

One would also expect such a monopole to be rotationally symmetric. That is, if constructed in  $n$  dimensions, and SO( $n$ ) rotation should leave the gauge potential invariant up to a gauge transform. It might also be expected that the potential would be singular in places and require two overlapping patches for a full description.

Yang showed that there exist two gauge fields in five dimensions that satisfy these criteria, which I shall label as  $\alpha$  and  $\beta$ . Each of these two gauge fields needs two overlapping potentials for a full description, and I shall label these  $a$  and  $b$ . The coordinate system he used is given by

$$x_i = \frac{2r\xi_i \sin \theta}{1 + \xi^2} \quad i = 1, 2, 3 \quad (6.38)$$

$$x_4 = \frac{r(1 - \xi^2) \sin \theta}{1 + \xi^2} \quad (6.39)$$

$$x_5 = r \cos \theta \quad (6.40)$$

$$r = (x_\mu x^\mu)^{\frac{1}{2}} \quad (6.41)$$

where the  $x_\mu$  are five-dimensional cartesian coordinates,  $\mu$  running from 1...5. The metric associated with this coordinate system is given by

$$ds^2 = dr^2 + r^2 d\theta^2 + \frac{4r^2 \sin^2 \theta}{(1 + \xi^2)^2} d\xi^2. \quad (6.42)$$

In these coordinates Yang gives his fields as follows: Let the phase factor be given by

$$\Phi(x^\mu + dx^\mu) = 1 + \frac{i}{2} b_\mu^j \sigma_j dx^\mu. \quad (6.43)$$

Define

$$B_j^i \equiv \langle i|B|j\rangle, \quad D_j^i \equiv \langle i|D|j\rangle \quad (6.44)$$

and

$$B = -8(1 + \xi^2)^{-2} [\psi\psi^T + \frac{1}{2}(1 - \xi^2) + N] \quad (6.45)$$

$$D = -B^T \quad (6.46)$$

with

$$\psi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & -\xi_3 & \xi_2 \\ \xi_3 & 0 & -\xi_1 \\ -\xi_2 & \xi_1 & 0 \end{pmatrix}. \quad (6.47)$$

Using these definitions  $b_r^i = b_\theta^i = 0$  for both fields  $\alpha$  and  $\beta$  in both regions  $a$  and  $b$ . For the solution  $\alpha$

$$b_j^i(a) = \frac{1}{2}(1 - \cos \theta)D_j^i, \quad b_j^i(b) = \frac{1}{2}(1 + \cos \theta)B_j^i. \quad (6.48)$$

For the solution  $\beta$

$$b_j^i(a) = \frac{1}{2}(1 - \cos \theta)B_j^i, \quad b_j^i(b) = \frac{1}{2}(1 + \cos \theta)D_j^i. \quad (6.49)$$

The form of this definition is, to my mind, elegant, concise and totally opaque. Rewriting the potentials in a form familiar to most physicists we find that the  $\alpha$  field is given by

$$A_\mu^{\alpha,a} = \frac{(1 - \cos \theta)}{(1 + \xi^2)^2} \begin{bmatrix} 0 \\ 0 \\ (1 + \xi_1^2 - \xi_2^2 - \xi_3^2)\sigma_1 + 2(\xi_1\xi_2 - \xi_3)\sigma_2 + 2(\xi_1\xi_3 + \xi_2)\sigma_3 \\ 2(\xi_1\xi_2 + \xi_3)\sigma_1 + (1 - \xi_1^2 + \xi_2^2 - \xi_3^2)\sigma_2 + 2(\xi_2\xi_3 - \xi_1)\sigma_3 \\ 2(\xi_1\xi_3 - \xi_2)\sigma_1 + 2(\xi_2\xi_3 + \xi_1)\sigma_2 + (1 - \xi_1^2 - \xi_2^2 + \xi_3^2)\sigma_3 \end{bmatrix} \quad (6.50)$$

$$A_\mu^{\alpha,b} = \frac{-(1 + \cos \theta)}{(1 + \xi^2)^2} \begin{bmatrix} 0 \\ 0 \\ (1 + \xi_1^2 - \xi_2^2 - \xi_3^2)\sigma_1 + 2(\xi_1\xi_2 + \xi_3)\sigma_2 + 2(\xi_1\xi_3 - \xi_2)\sigma_3 \\ 2(\xi_1\xi_2 - \xi_3)\sigma_1 + (1 - \xi_1^2 + \xi_2^2 - \xi_3^2)\sigma_2 + 2(\xi_2\xi_3 + \xi_1)\sigma_3 \\ 2(\xi_1\xi_3 + \xi_2)\sigma_1 + 2(\xi_2\xi_3 - \xi_1)\sigma_2 + (1 - \xi_1^2 - \xi_2^2 + \xi_3^2)\sigma_3 \end{bmatrix}. \quad (6.51)$$

The  $\beta$  potentials are given as follows:  $A_\mu^{\beta,a}$  is identical to  $A_\mu^{\alpha,b}$  except with the numerator of the prefactor given by  $-(1 - \cos \theta)$ ; and  $A_\mu^{\beta,b}$  is identical to  $A_\mu^{\alpha,a}$  except with the numerator of the prefactor given by  $(1 + \cos \theta)$ . The region  $a$  includes the “north pole”  $\theta = 0$ , and the region  $b$  includes the “south pole”  $\theta = \pi$ . I call these regions  $N$  and  $S$  respectively. As is to be expected each field is related in the area of overlap between the  $a$  and  $b$  regions via a gauge transformation, and the  $\alpha$  and  $\beta$  fields are gauge-inequivalent [205].

It should be noted that Yang uses what he calls “tensor notation”, where he ignores the coefficients of the metric, e.g. takes the gradient operator in spherical polars as  $(\partial_r, \partial_\theta, \partial_\phi)$  rather than  $(\partial_r, 1/r\partial_\theta, 1/(r \sin \theta)\partial_\phi)$  and because of this his potentials differ by simple metric form factors from what may be expected. However, one never measures the gauge potential directly, but rather measures its effect by phase factors [204], and since these phase factors invariably arise within a line integral where the line element is modified the metric form factors,

i.e.  $g_{11}dx^1 + g_{22}dx^2 + \dots + g_{nn}dx^n$ , provided one ignores the metric factors in the integral as well everything cancels correctly.

Yang proved a number of interesting things about his two fields. Perhaps the most interesting is that these two fields are the *only* two SU(2) gauge fields that have a non-zero second Chern class number and SO(5) symmetry. Further, there exist no SU(2) fields in spaces of dimension  $n > 5$  with SO( $n$ ) symmetry, aside from the trivial case that is a pure gauge field, i.e. one that has  $F_{\mu\nu}^i = 0$ . In the case of  $n = 4$  there exists a one-parameter family of solutions. For  $n = 3$  there also exists an SU(2) field with SO(3) symmetry, but it is relatively trivial. It is described by  $A_i = A_i^{\text{dirac}}\sigma_3$  where  $A_i^{\text{dirac}}$  is the potential of the U(1) Dirac monopole, i.e.

$$\mathbf{A}^{\text{dirac}} = \frac{g}{2r} \frac{(1 - \cos\theta)}{\sin\theta} \sigma_z \hat{\phi} \quad (6.52)$$

Thus although it is SU(2) it only involves one generator and thus is abelian for all intents and purposes, and the full complexity of the group is not realized. It may appear that in the years since Yang published his paper counter-examples have appeared. For example, Biswas [37] and Bohm *et al.* [43] have monopoles defined by the potentials

$$\begin{aligned} A_i^{\text{bohms}} &= \frac{1}{2} \epsilon_{ijk} \frac{x^j}{r^2} \sigma_k \\ &= \frac{1}{2r^2} \mathbf{r} \times \boldsymbol{\sigma}. \end{aligned} \quad (6.53)$$

However such potentials are gauge-transformable to fields such as (6.52). If we let

$$S = \begin{pmatrix} \cos \frac{\theta}{2} & \sin \frac{\theta}{2} e^{i\phi} \\ -\sin \frac{\theta}{2} e^{-i\phi} & \cos \frac{\theta}{2} \end{pmatrix} \quad (6.54)$$

and apply a nonabelian gauge transformation we find

$$\begin{aligned} A_i^{\text{bohms}'} &= SA_i^{\text{bohms}}S^{-1} - i(\partial_i S)S^{-1} \\ &= A_i^{\text{dirac}}. \end{aligned} \quad (6.55)$$

In any case the potential described by (6.53) and equivalently (6.52) is not a true monopole since it does not fulfil the correct topological requirements unless  $g$  is a half-odd integer [89]. This arises from a similar argument to the one used in §5.1 to derive the quantization strengths for the Dirac monopole: The extra factor of one-half arises in this case because we are dealing with SU(2) rather than U(1). The upshot of this is that one can describe the Biswas or Bohm “monopole” using only one coordinate patch, and hence their potentials are topologically equivalent to the vacuum, i.e. the fibre bundle is trivial.

Taking Yang’s  $A_\mu^\alpha$  and  $A_\mu^\beta$  as the description of the SU(2) monopole, how do the Kramers doublet potentials (6.31) compare? To make the comparison it is

simplest to rewrite the basis vectors (6.28) and (6.29) in terms of Yang's coordinates and calculate the new geometric gauge potential using (6.30), omitting the metric form factors.

Letting  $\gamma^\mp = (i\sqrt{2(1 \mp \cos \theta)}(1 + \xi^2))^{-1}$  the eigenvectors become

$$\psi_1^\pm = \gamma^\mp \begin{bmatrix} \pm 2 \sin \theta (\xi_1 + i\xi_2) \\ \pm \sin \theta (2\xi_3 - i(1 - \xi^2)) \\ 0 \\ (1 + \xi^2)(1 \mp \cos \theta) \end{bmatrix}, \quad (6.56)$$

$$\psi_2^\pm = \gamma^\mp \begin{bmatrix} \mp \sin \theta (2\xi_3 + i(1 - \xi^2)) \\ \mp 2 \sin \theta (-\xi_1 + i\xi_2) \\ -(1 + \xi^2)(1 \mp \cos \theta) \\ 0 \end{bmatrix} \quad (6.57)$$

corresponding to the eigenvalues  $\pm r$ .

Inserting these basis vectors into (6.30) and comparing the resulting geometric potentials to Yang's fields (6.50) and (6.51) we find

$$A_\mu^+ = A_\mu^{(\alpha, S)} \quad (6.58)$$

$$A_\mu^- = A_\mu^{(\beta, N)}. \quad (6.59)$$

Astonishingly, we have an exact match between the geometric potential generated by the pair of Kramers doublets and the field of Yang's  $SU(2)$  monopole. However we can go still further. While we have identified the two geometric gauge potentials corresponding to the positive and negative energy levels of the doublets with the fields  $\alpha$  and  $\beta$ , we have only identified them in one hemisphere each. If the geometric gauge potentials are truly to describe a monopole it must be possible to obtain the potentials corresponding to the other hemispheres by using a different set of eigenvectors.

Since Yang's fields are nonabelian, the  $a$  and  $b$  patch fields are transformed between each other via a nonabelian gauge transformation of the form

$$A_\mu \rightarrow A'_\mu = SA_\mu S^{-1} - \frac{i}{g}(\partial_\mu S)S^{-1} \quad (6.60)$$

where  $S$  is an element of the gauge group, in this case  $SU(2)$ .

From §3.4 we know that when considering the nonabelian geometric phase, a gauge transformation of the potentials is associated with a unitary transformation  $\Lambda$  acting on the basis vectors in each degenerate subspace, which in this case are the Kramers doublets:

$$|\psi_i\rangle \rightarrow |\psi'_i\rangle = \Lambda_{ji}|\psi_j\rangle \quad (6.61)$$

$$\begin{aligned} A_\mu^{ij} \rightarrow A_\mu^{i'j'} &= \langle \psi'_i | i\partial_\mu | \psi'_j \rangle \\ &= \Lambda^{-1} A_\mu \Lambda + i\Lambda^{-1} \partial_\mu \Lambda. \end{aligned} \quad (6.62)$$

Note that Yang's fields  $A_\mu^{(\alpha,S)}$  and  $A_\mu^{(\beta,N)}$  are transformed to their other patch counterparts  $A_\mu^{(\alpha,N)}$  and  $A_\mu^{(\beta,S)}$  via (6.60) with [205]

$$S = (1 - \xi^2 + 2i\xi \cdot \sigma)/(1 + \xi^2). \quad (6.63)$$

Thus we apply the basis change  $|\psi'_i\rangle = |\psi_k\rangle \Lambda_{ki}$ , where

$$\begin{aligned} \Lambda &= S^{-1} \\ &= (1 - \xi^2 - 2i\xi \cdot \sigma)/(1 + \xi^2) \end{aligned} \quad (6.64)$$

and get the alternative basis set

$$\psi_1'^{\pm} = \gamma^{\mp} \begin{bmatrix} 0 \\ \pm \sin \theta (1 + \xi^2) \\ -2(\xi_1 + i\xi_2)(1 \mp \cos \theta) \\ i(1 - \xi^2 - 2i\xi_3)(1 \mp \cos \theta) \end{bmatrix}, \quad (6.65)$$

$$\psi_2'^{\pm} = \gamma^{\mp} \begin{bmatrix} \pm \sin \theta (1 + \xi^2) \\ 0 \\ -i(1 - \xi^2 + 2i\xi_3)(1 \mp \cos \theta) \\ 2(1 \mp \cos \theta)(\xi_1 - i\xi_2) \end{bmatrix} \quad (6.66)$$

using the previous definition of  $\gamma^{\mp}$ . When put into (6.30) these new vectors yield  $A_\mu'^+ = A_\mu^{(\alpha,N)}$  and  $A_\mu'^- = A_\mu^{(\beta,S)}$ .

I have now completely identified the two geometric potentials associated with the higher and lower energy Kramers doublets with Yang's two gauge-inequivalent monopoles:

$$A_\mu^+ = A_\mu^\alpha \quad (6.67)$$

$$A_\mu^- = A_\mu^\beta. \quad (6.68)$$

It should be noted that in [97] we demonstrated that the vector potentials arising from a pair of Kramers doublets was exactly the *negative* of Yang's fields. This was due to using the alternate definition of  $A_\mu^{ij}$ ; that is using (3.25) rather than (6.30). This effectively performs a transpose operation on the potentials. Under this operation  $\sigma_1$  and  $\sigma_3$  are unaffected since they are symmetric matrices, but  $\sigma_2 \rightarrow -\sigma_2$ . To tidy things up it is possible to perform a gauge transformation using (6.60) with  $S = i\sigma_2$  which changes the sign on the the  $\sigma_1$  and  $\sigma_3$  components. Thus the net effect has been  $A_\mu^{ij} \rightarrow -A_\mu^{ij}$ , accounting for the difference in sign.

There is another way of looking at these fields, however. Yang's monopole is closely related to the instanton, and it is to this object I now turn.

## 6.5 The $SU(2)$ instanton

The instanton is a particular solution to the Yang-Mills field equations first found by Belavin, Polyakov, Schwarz and Tyupkin [29]. These solutions were immediately further analysed by Jackiw and Rebbi [95] and by Callan, Dashen and Gross [48, 49] who explored their wonderfully rich topological character. They found that the solutions were the first evidence of non-trivial vacuum structure for nonabelian gauge field theory. More specifically they demonstrated that in Yang-Mills theories it can come about that there are a countable infinity of vacuum states, each separated by only a finite barrier, and consequently there is a finite probability of tunnelling between these vacuum states. The tunnelling between these states is conveniently described by certain Euclidean classical solutions of the Yang-Mills equations which are the instanton.

One of the first things that the instanton allowed was the solution of a long-standing puzzle in quantum field theory — the  $U(1)$  problem. Briefly speaking the  $U(1)$  problem arises from the fact that quantum chromodynamics has among its symmetries a chiral  $SU(2) \otimes SU(2)$  symmetry group generated by the strangeness-conserving weak-interaction currents and their parity transforms. Now, in the limit of massless up and down quarks this symmetry is exact, and a further chiral  $U(1)$  symmetry arises. If this  $U(1)$  symmetry was manifest, all non-massless hadrons would occur in parity doublets. This is obviously not the case, so the symmetry must be spontaneously broken. But by Goldstone's theorem this means there must be an associated isoscalar Goldstone boson. This was the  $U(1)$  problem: no one could find this particle, even though Weinberg showed that it must have a mass less than  $\sqrt{3}m_\pi$ , where  $m_\pi$  is the mass of the pion [196]. The instanton solution was quickly seized upon by 't Hooft as a method for explaining this problem via the Adler-Bell-Jackiw anomaly [184, 185, 186], demonstrating that this particle need not exist.

The classical way to construct an instanton is to find a solution that minimizes the Yang-Mills action:

$$S = \frac{1}{2} \int_M \text{Tr}(F_{\mu\nu} F^{\mu\nu}) d^4x \quad (6.69)$$

where the space under consideration is Euclidean and four-dimensional. The gauge group will be assumed to be  $SU(2)$  — if one wishes to use an arbitrary Lie group then the trace in the action is replaced by the Killing form. Minimizing (6.69) would appear to be very difficult but for a particular observation: The Euler-Lagrange equations

$$D^*F = 0 \quad (6.70)$$

or

$$[D_\mu, F_{\mu\nu}] = 0 \quad (6.71)$$

where  $D$  is the covariant derivative, are automatically satisfied if the curvature two-form is proportional to its dual. If

$$\mathbf{F} = \lambda^* \mathbf{F} \quad (6.72)$$

then the Bianchi identity

$$D\mathbf{F} = 0 \quad (6.73)$$

automatically implies that (6.70) is true. Further considerations can show that  $\lambda = \pm 1$  [135]. Finally one must apply boundary conditions. The action must be finite, so it is necessary for  $F_{\mu\nu}$  to vanish at infinity. However the gauge fields  $A_\mu$  need not decay as fast; in fact they need only decay to the zero connection:

$$\lim_{x \rightarrow \infty} A_\mu(x) = (\partial_\mu \Omega) \Omega^{-1} \quad (6.74)$$

where  $\Omega$  is an element of the gauge group (I have used  $\Omega$  rather than  $S$  in order to avoid confusion with the special unitary groups and hypersphere notation later).

Rather than continuing to work in  $R^4$  we move to a conformal compactification of this space by adding the point at infinity, exploiting the result that  $R^4 \cup \{\infty\}$  is isomorphic to the four-sphere  $S^4$  [135]. Now, from (6.74) we see that  $\Omega$  can be thought of as being defined by a sphere at infinity in  $R^4$ , that is on  $S^3$ , which will be our base space. Since the gauge group is  $SU(2)$ , each  $\Omega$  is a continuous map

$$\Omega : S^3 \rightarrow SU(2). \quad (6.75)$$

Such mappings fall into homotopy classes, and are the elements of  $\pi_3(SU(2))$ , where  $\pi_3$  is the three-dimensional homotopy group. However, topologically  $SU(2)$  is an  $S^3$ . Thus

$$\begin{aligned} \pi_3(SU(2)) &= \pi_3(S^3) \\ &= \mathbb{Z}. \end{aligned} \quad (6.76)$$

Thus every mapping  $\Omega$  is labelled by an integer  $q$ , and  $q$  is called the degree of  $S$ , or in this context the instanton number. This integer  $q$  therefore classifies principal fibre bundles with the group  $SU(2)$  over  $S^4$ . This integer also labels the different vacua that are connected to each other via a tunnelling transition. Each class of connections described the same  $q$  are reducible to each other by a continuous gauge transformation, whereas those in different classes cannot be so reduced.

This is worth comparing to the  $U(1)$  monopole. Recall that in the monopole case, we needed at least two coordinate patches to avoid singularities in the vector potential, and that the potentials were connected at the  $S^1$  boundary

between the patches by a non-trivial gauge transformation. Indeed, the associated transition function [204]  $\exp[\pm i\phi]$  defines a map from the  $S^1$  equator to the  $U(1)$  (structure) group, with winding number  $\pm 1$  (and similarly for monopoles of higher magnetic charge). In the instanton case,  $S^4$  can be covered by two patches with an overlap region which is  $S^3$ , and the gauge transformation which connects the two corresponding potentials in this  $S^3$  provides a map  $\Omega$  via (6.75). The topological integer  $q$  characterizing these maps is analogous to the magnetic charge carried by the  $U(1)$  monopole, but while the latter is defined via a two-dimensional surface integral of the second-rank field strength tensor, the former involves a four-dimensional surface integral of a fourth-rank tensor, namely  $\text{Tr}(F_{\mu\nu}F_{\rho\sigma}^*)$ , and is related to the second Chern class number.

Solutions to (6.72) and (6.74) that are spherically symmetric and have  $q = \pm 1$  are called BPST instantons after Belavin, Polyakov, Schwartz, Tyupkin [29]. It is these that we will be primarily concerned with. Instantons with  $|q| > 1$  (multi-instantons) do exist and have second Chern class number  $-q$  [179], but it is generally difficult to find the explicit form of their connection. The form of the BPST instanton is normally given in one of two forms, either [97, 150]

$$A_\mu^{\text{inst}} = \frac{2}{x^2 + \lambda^2} \Sigma_{\mu\nu} x_\nu \quad (6.77)$$

where  $\Sigma_{\mu\nu} = \eta_{i\mu\nu} \sigma_i / 2$  with  $\eta_{i\mu\nu} = -\eta_{i\nu\mu} = \epsilon_{i\mu\nu}$  for  $\mu, \nu = 1, 2, 3$  and  $\eta_{i\mu\nu} = \delta_{\mu\nu}$  for  $\nu = 4$ , or [29]

$$A_\mu^{\text{inst}} = \frac{x^2}{x^2 + \lambda^2} \Omega^{-1}(x) \partial_\mu \Omega(x), \quad (6.78)$$

$$\Omega(x) = \frac{1}{\sqrt{x^2}} (x_4 + i(x_1 \sigma_1 + x_2 \sigma_2 + x_3 \sigma_3)). \quad (6.79)$$

$\lambda$  is simply a parameter giving the “size” of the instanton.

So what is the relevance of the instanton to our  $SU(2)$  monopole? Yang demonstrated that his fields describing the  $SU(2)$  monopole also minimize the action, and the quantization of monopole strength, spherical symmetry and the fibre-bundle structure of those fields bear more than a passing resemblance to those of the instanton.

Going back to (6.50,6.51) we see that  $SU(2)$  monopole fields are independent of the radius,  $r$ , just as was the case for the  $U(1)$  monopole in Chapter 5. Thus the nonabelian potentials (6.31) are naturally defined on the sphere  $S^4$ . To exploit this we project from the five-dimensional potentials (6.31) onto the surface of the unit four-dimensional hypersphere via the coordinate transformation

$$\begin{aligned} q_\mu &= \frac{2x_\mu}{1+x^2} \\ q_5 &= \frac{1-x^2}{1+x^2} \end{aligned} \quad (6.80)$$

where  $\mu$  runs from 1 to 4 and  $x^2 = x_\mu x_\mu$ . Next we need to find how the components of the vector potential change under this transformations. The easiest way to do this is to compare distance elements in the two coordinate systems and in order to do this we need the metric form factors  $h_i$ . We have

$$h_i^2 = \left(\frac{\partial q_1}{\partial x_i}\right)^2 + \left(\frac{\partial q_2}{\partial x_i}\right)^2 + \left(\frac{\partial q_3}{\partial x_i}\right)^2 + \left(\frac{\partial q_4}{\partial x_i}\right)^2 + \left(\frac{\partial q_5}{\partial x_i}\right)^2 \quad (6.81)$$

For the coordinate transformation (6.80) this gives

$$h_1^2 = h_2^2 = h_3^2 = h_4^2 = \frac{4}{(1+x^2)^2} \quad (6.82)$$

and consequently comparing distance elements

$$\begin{aligned} ds &= \frac{2}{1+x^2} (dx_1 \hat{x}_1 + dx_2 \hat{x}_2 + dx_3 \hat{x}_3 + dx_4 \hat{x}_4) \\ &= dq_1 \hat{q}_1 + dq_2 \hat{q}_2 + dq_3 \hat{q}_3 + dq_4 \hat{q}_4 + dq_5 \hat{q}_5. \end{aligned} \quad (6.83)$$

Varying each  $x_i$  yields

$$\begin{aligned} \hat{x}_\mu &= \frac{1+x^2}{2} \frac{\partial q_j}{\partial x_\mu} \hat{q}_j \\ &= \frac{2}{1+x^2} \left[ \frac{1}{2} ((1+x^2) - x_1^2) \hat{q}_\mu - x_\mu x_\nu \hat{q}_\nu + x_\mu^2 \hat{q}_\nu - x_\mu \hat{q}_5 \right]. \end{aligned} \quad (6.84)$$

Denoting the four-dimensional potential by an overbar this gives the relations

$$A_\mu(q) = \bar{A}_\mu(x) - \frac{2}{1+x^2} x_\mu x_\nu \bar{A}_\nu(x) \quad (6.85)$$

$$A_5(q) = \frac{-2}{1+x^2} x_\nu \bar{A}_\nu(x) \quad (6.86)$$

along with the inverse relation

$$\bar{A}_\mu(x) = A_\mu(q) - x_\mu A_5(q). \quad (6.87)$$

Note that from (6.80) we have  $A_\mu(q) = 2(1+x^2)^{-1} A_\mu(x)$ . Thus our monopole potentials projected onto  $S^4$  have become

$$\bar{A}_\mu^+ = \frac{1}{x^2(1+x^2)} \begin{bmatrix} -x_4 \sigma_1 - x_3 \sigma_2 + x_2 \sigma_3 \\ x_3 \sigma_1 - x_4 \sigma_2 - x_1 \sigma_3 \\ -x_2 \sigma_1 + x_1 \sigma_2 - x_4 \sigma_3 \\ x_1 \sigma_1 + x_2 \sigma_2 + x_3 \sigma_3 \end{bmatrix} \quad (6.88)$$

while  $\bar{A}_\mu^- = x^2 A_\mu^+$ .

This should be compared with the instanton field (6.77). We see that upon performing the coordinate transformation  $x_4 \rightarrow -x_4$  our monopole potentials

$\bar{A}_\mu^-$  are exactly the BPST instanton of size  $\lambda = 1$ ! Furthermore, the anti-instanton is obtained by inverting  $A_\mu^{\text{inst}}$  [49], and thus we find that  $\bar{A}_\mu^+$  is the anti-instanton!

This is truly remarkable, and it is worth stepping back slightly and seeing what has transpired. From the prosaic starting point of a pair of Kramers doublets in non-relativistic quantum mechanics, a pair of nonabelian vector potentials has been generated that have not only been identified as the only two allowable  $SU(2)$  monopoles; but also perform double duty as the topological delights of the instanton/anti-instanton pair. Who would have thought that two Kramers doublets would know so much about exotic topology and classical nonabelian gauge field theories?

It should be noted that in the present case, of course, the field configurations are entirely in Euclidean space, and so there is no question of interpreting them as tunnelling events in Minkowski space.

As mentioned at the beginning of this chapter, the link between time reversal invariant Fermi systems and the instanton was first noticed by Avron *et al.* [25] in their examination of topological invariants for such systems. They noted that the associated bundles had a non-zero second Chern class number of  $\pm 1$  and that the connection had self-dual curvature, and was thus related to the instanton. They did not perform explicit calculations or derive potentials, however. This situation was also investigated by Lévy, who examined time reversal symmetric, half-integer spin systems using a quaternionic formalism and differential geometry [116, 117]. He reached the same conclusions as Avron *et al.*, anticipating some of the results in this chapter, but once again did not explicitly calculate the instanton or monopole potentials.

Finally, it is worth noting that another interesting link between instanton-like topological curiosities and time reversal arises from the fact that it has been shown that in the non-relativistic limit Thomas precession of electron spin can be understood as arising from a  $SU(2)$  geometric vector potential [122]. This was shown by considering the spinors that are the solutions to the free Dirac equation, where the four levels at each momentum split into two spin-degenerate pairs with equal and opposite energies, related by time reversal. Later Shankar and Mathur claimed to have demonstrated that the object generating this  $SU(2)$  potential was a meron [165], a solution to the Yang-Mills equations with an instanton number of  $\frac{1}{2}$  [57]. This claim has been disputed by Samuel, who maintains that although locally this identification can be made, globally it fails because the electron four-momentum is not an arbitrary parameter, but rather must fall within the particle's light cone [159]. Samuel does, however, show that abelian merons do exist within the context of the geometric phase, arising when a two-dimensional equatorial slice of a monopole potential is considered. This is simply the case discussed at the end of §5.3 where the three-dimensional monopole space is reduced by one by forcing the Hamiltonian to be time-even.

In this case the potential we are left with is effectively that of a fluxtube which is known to give rise to a simple sign change in the phase factor. Thus the sign change initially found in the polyatomic systems considered by Herzberg and Longuet-Higgins were, unbeknownst to them, generated by a meron! These results underscore the fact that the results in this chapter are purely topological: even using the Dirac equation instanton-like solutions arise from time-even, half-integral spin systems.

## 6.6 Duality and existence of monopoles

It is perhaps worthwhile to pause at this point and consider the question of whether monopoles are actually expected or believed to exist.

The Dirac U(1) monopole of electromagnetism appears to be optional to all theories. Admittedly it does symmetrize the laws of electromagnetism [207] and provide a reason for the existence of quantized electric charge but these points do not demand its existence. One can construct theories with or without the Dirac monopole and either way one has a theory that describes the physics we measure equally well. Because of this we are forced to turn to experimental methods of verification, and so far a Dirac monopole has never been experimentally found, although there have been at least a couple of briefly exciting false alarms [47, 149].

The situation becomes more complex when we consider the existence of nonabelian monopoles. While monopoles are not found in the standard model, many of its extensions predict them. To admit monopoles a compact gauge group is required, and while the standard model (gauge group  $SU(2) \times U(1)$ ) does not meet this criterion other models do. 't Hooft [183] and Polyakov [148] were the first to find that a simple extension of the standard model admitted monopole solutions. The reason I have not considered the 't Hooft-Polyakov monopole until now is that it does not exist in isolation, but arises upon introducing a triplet of scalar Higgs field exhibiting spontaneous symmetry breaking into a gauge theory. The solutions to this theory include monopole-like gauge fields that have the correct topological properties (i.e. require two patches for a complete description and so on), although it can be argued that they are actually U(1) monopoles embedded in a larger SO(3) group and hence not really nonabelian [89]. A major difference between this type of monopole and the Dirac monopole is that solutions like the 't Hooft-Polyakov monopole are *required* by the gauge theory, and are not optional. Such monopoles occur in grand unified theories and the simplest of such theories estimate a monopole mass of  $\sim 10^{13} - 10^{16} \text{ GeV}$  [153]. Of course such a mass is far beyond accelerator capabilities but they may exist as relics from the big bang and many experiments, so far fruitless, have tried to detect them. The current lower bound on the monopole flux is the so-called Parker bound of  $10^{-16} (m/10^{17} \text{ GeV}) \text{ cm}^{-1} \text{ sr}^{-1} \text{ s}^{-1}$  [2]. A good reference to the history of

monopole detection, both abelian and nonabelian, can be found in the resource paper by Goldhaber and Trower [81].

There is a point of some subtlety that arises in considering the existence of monopoles, and it is due to duality [89]. First take ordinary electromagnetism. It can be considered from two perspectives that correspond to using  $F_{\mu\nu}$  or  $*F_{\mu\nu}$ . In the  $F_{\mu\nu}$  formulation of electromagnetism electric charges are the sources and magnetic monopoles have the interesting topological properties, and in the  $*F_{\mu\nu}$  formulation things are reversed — electric charges are the monopoles and magnetic charges are the sources. That is we get the same equations of motion if we perform the transformation

$$F_{\mu\nu} \rightarrow *F_{\mu\nu} \tag{6.89}$$

$$e \rightarrow \tilde{e} \tag{6.90}$$

where  $e$  and  $\tilde{e}$  are the electric and magnetic charges respectively.

A sketch argument for this is as follows: From the definition of the  $\epsilon$ -dual in electromagnetism and Minkowski space we have [135]

$$F_{\mu\nu} = -*(F_{\mu\nu}). \tag{6.91}$$

The Bianchi identity can be written

$$\partial^\nu *F_{\mu\nu} = 0. \tag{6.92}$$

Now, at a point where spacetime is source-free we have  $\partial^\nu F_{\mu\nu} = 0$ . This is the Bianchi identity for the dual field  $*F_{\mu\nu}$ . Consequently the dual field is derivable from a potential by the Poincare lemma, i.e.  $*F_{\mu\nu} = \partial_\mu \bar{A}_\nu - \partial_\nu \bar{A}_\mu$ , and all of electromagnetism follows happily in this new picture.

Thus because of dual symmetry in the abelian case the dynamics of monopoles and sources are the same in either description. However things become different when considering Yang-Mills gauge theories. In this case we can't say that there exists an analogous potential for the dual field tensor as we could for the abelian formulation ([89], chapter 5).

Consequently although we can automatically consider monopoles to exist in the abelian case, such a statement is trivial because of the duality. In the nonabelian case though, there is a genuine difference in the dynamics between a monopole and a source, and so it is possible to really say whether a particle is a monopole or a source regardless of what prescription one takes. It should be noted, however, that most evidence would suggest that the charges we see in Yang-Mills theories are actually sources rather than monopoles [89].

## 6.7 Applications

The reader may well by now be very impressed with the lovely topological consequences that follow from a pair of Kramers doublets, and intrigued at the curious appearance of gauge field entities such as the instanton turning up from such humble origins, but be experiencing a slightly nagging worry along the line of “does it have any relevance for real physical systems?” Although I personally feel that exploring the mathematics and discovering beautiful relationships in the theory is motivation enough to explore the elegance of this formalism, this chapter does indeed have some relevance to real physical problems. To this end, let us now descend from the rarefied heights of fibre bundles, Chern classes and excessively multi-dimensional spheres to the more concrete area of molecular physics in order to consider a pair of actual examples.

The geometric phase becomes relevant in molecular physics because of the adiabatic or Born-Oppenheimer approximation. I will examine this in more detail when I study Jahn-Teller systems in Chapter 7, but loosely speaking this approximation considers the molecule in two parts. The motion of the nuclei is considered separately from the electronic motion, and because the nuclear positions move much slower than the electronic positions they can be considered to be adiabatic parameters upon which the electronic states depend. We thus have a tailor-made system for geometric phases and, as might be expected, a vector potential makes an appearance in the Hamiltonian governing the nuclear motion. This vector potential is exactly the geometric vector potential calculated via  $A_\mu = \langle \psi(Q) | i\partial_\mu | \psi(Q) \rangle$  where the  $Q$  are the nuclear coordinates.

We therefore look for molecular systems where there are two pairs of doublet levels, each pair degenerate because of time reversal considerations. The first such case that I will consider is the  $\text{CH}_4^+$  ion near its  $T_d$  configuration, which is studied by Mead in his excellent review article on the geometric phase in molecular systems [126]. For simplicity Mead considers only the five bending modes of the molecule. These five normal coordinates for the molecule are analogous to the five-parameter Hamiltonian derived in §6.2 and consequently the five-dimensional space the  $\text{SU}(2)$  monopole exists in. The space of electronic states that we are interested in and which will be coupled to the five bending modes transforms as  $U_{3/2}$  which is four dimensional and transforms like the angular momentum eigenkets corresponding to quantum number  $\frac{3}{2}$ . Mead consequently takes as his basis states the combination

$$\begin{aligned}
 |1\rangle &= \frac{1}{\sqrt{2}}(|\tfrac{1}{2}\rangle - i|-\tfrac{3}{2}\rangle) \\
 |2\rangle &= \frac{1}{\sqrt{2}}(-|-\tfrac{1}{2}\rangle + i|\tfrac{3}{2}\rangle) \\
 |3\rangle &= \frac{1}{\sqrt{2}}(|\tfrac{1}{2}\rangle + i|-\tfrac{3}{2}\rangle) \\
 |4\rangle &= \frac{1}{\sqrt{2}}(-|-\tfrac{1}{2}\rangle - i|\tfrac{3}{2}\rangle).
 \end{aligned} \tag{6.93}$$

However, since in this system the angular momentum kets transform under time

reversal as

$$\begin{aligned} T|\frac{1}{2}\rangle &= |-\frac{1}{2}\rangle, & T|-\frac{1}{2}\rangle &= -|\frac{1}{2}\rangle \\ T|\frac{3}{2}\rangle &= -|-\frac{3}{2}\rangle, & T|-\frac{3}{2}\rangle &= |\frac{3}{2}\rangle \end{aligned} \quad (6.94)$$

we see that  $T|4\rangle = |1\rangle$  and  $T|2\rangle = |3\rangle$ . Thus relabelling the states via  $|1\rangle \rightarrow |2\rangle$ ,  $|2\rangle \rightarrow |3\rangle$ ,  $|3\rangle \rightarrow |4\rangle$ ,  $|4\rangle \rightarrow |1\rangle$ , we obtain the basis used in §6.2 and used to calculate the Hamiltonian (6.26). Finally we have to allow for the different coordinate systems used. Mead uses one that splits the five bending mode normal coordinates into two sets: a three-dimensional cartesian set parameterized by  $r_i$ ,  $i = 1 \dots 3$  and a two-dimensional polar set  $\rho, \epsilon$ . Allowing for the relabelling of the basis as above and the different coordinate system we find that Mead's Hamiltonian is identical to (6.26) if we identify

$$\begin{aligned} q_1 &= r_1 \\ q_2 &= r_2 \\ q_3 &= -\rho \cos \epsilon \\ q_4 &= -\rho \sin \epsilon \\ q_5 &= r_3. \end{aligned} \quad (6.95)$$

The second case is that studied by Apsel *et al.* [22]. They examine the  $\Gamma_8 \otimes (\tau_2 \oplus \epsilon)$  Jahn-Teller system. This is of interest because the non-distorted electronic state transforming as the four-dimensional representation  $\Gamma_8$  has half-integral angular momentum, and the Kramers degeneracy cannot be removed by any interaction with the normal modes. Consequently we have a pair of Kramers levels coupled to the normal phonon modes  $\tau_2$  and  $\epsilon$  which are respectively three and two dimensional representations. Thus once again we have five adiabatic parameters governing the evolution of a pair of Kramers doublets. Apsel *et al.* also choose combinations of  $S_z$  spin eigenstates as a basis, taking

$$\begin{aligned} |1\rangle &= \frac{1}{\sqrt{2}}(-i|\frac{1}{2}\rangle + |-\frac{3}{2}\rangle) \\ |2\rangle &= \frac{1}{\sqrt{2}}(i|-\frac{1}{2}\rangle - |\frac{3}{2}\rangle) \\ |3\rangle &= \frac{1}{\sqrt{2}}(-i|\frac{1}{2}\rangle - |-\frac{3}{2}\rangle) \\ |4\rangle &= \frac{1}{\sqrt{2}}(i|-\frac{1}{2}\rangle + |\frac{3}{2}\rangle). \end{aligned} \quad (6.96)$$

Thus, interchanging two states via  $|2\rangle \leftrightarrow |4\rangle$  give the basis I used in §6.2. Apsel *et al.* also use a different coordinate system for the nuclear positions, and to identify their Hamiltonian we take

$$\begin{aligned} q_1 &= V_T \cos \beta \sin \theta \cos \phi \\ q_2 &= -V_T \cos \beta \sin \theta \sin \phi \\ q_3 &= -V_E \sin \beta \cos \chi \\ q_4 &= -V_E \sin \beta \sin \chi \\ q_5 &= V_T \cos \beta \cos \theta. \end{aligned} \quad (6.97)$$

It may appear that (6.97) uses more than five parameters — in addition to the normal mode coordinates  $\{q, \theta, \phi, \chi, \beta\}$  there are two other factors:  $V_T$  and  $V_E$ . These correspond to the coupling strengths between the electronic states and the nuclear states. It must be remembered, however, that these factors are usually taken as constants, and do not change under the adiabatic evolution of the nuclear positions and thus have no effect on the geometric vector potential. Further, if one should create a model where these coupling constants vary with nuclear position one can still consider the equal coupling case were  $V_T = V_E$ . In this case the Hamiltonian acquires  $SO(5)$  symmetry as desired, and there is only one extra parameter which can be “swept under the rug” by identifying it with a straightforward adjustment to the harmonic restoring term. In this case it does not appear in the vector potential.

Although I have shown that these two systems can be written in such a way as to produce an instanton/monopole geometric vector potential, the aforementioned workers have not done so. At first glance it may seem that their potential should differ from mine by only a gauge transformation; after all they describe the same physical situation. This is not the case, however. In the nonabelian formulation of the geometric phase a gauge transformation is associated with a basis transformation within a degenerate level. In the situation we are considering this means a linear combination of the two degenerate states that make up a Kramers doublet. In order to put their basis in the form  $\{|\phi\rangle, |\bar{\phi}\rangle, |\psi\rangle, |\bar{\psi}\rangle\}$  Mead and Apsel *et al.* need a basis change that mixes levels outside this degeneracy. Finally I would like to stress that despite this these two problems, and indeed many others, *can* be written in such a form as to develop an instanton geometric vector potential.

To my knowledge most workers in the field of molecular physics are unaware of the instanton result. It is certainly useful to be aware of it, and as I have for the first time given the explicit forms of the geometric vector potential arising from a reasonable choice of basis, it is now a simple matter to include these potentials into the effective nuclear Hamiltonian for any molecular system with pair of Kramers degenerate levels. The inclusion of vector potentials into molecular vibronic systems will be more fully justified in Chapter 7.

## 7. Jahn-Teller systems and time-odd coupling

As has been hinted in earlier chapters, the geometric phase plays an important role within the field of molecular physics. This is to be expected as molecular physics generally deals with the interaction between two widely separate systems — the first being the low energy, slow-moving nuclei and the second the much lighter, faster and higher energy electrons. The electrons are slaved to the slower nuclei and tend to follow their motion adiabatically. Thus we have a system that depends adiabatically on a set of slowly moving parameters, a situation tailor-made for geometric phases. The only subtlety is that in this case the adiabatic parameters are not generally under external experimental control, but rather are internal variables themselves.

The standard tool for dealing with the interactions between nuclei and electrons in a molecular setting is the Born-Oppenheimer approximation [42, 45], dating back to 1927. I review the Born-Oppenheimer and adiabatic approximations in §7.1, and demonstrate how the geometric gauge potential arises within its description. It is interesting to note that despite the obviousness of the potential when molecular problems were analysed in this way, it was still persistently ignored in the mistaken belief that it would vanish under the appropriate choice of phase. This situation continued until Herzberg and Longuet-Higgins showed that the Born-Oppenheimer electronic wavefunction underwent a sign change if the nuclear coordinates followed a closed path around a conical intersection of two electronic potential-energy surfaces [87]. These things stood until Mead and Truhlar examined the behaviour of the electronic wavefunction in a more general way and showed that the multivaluedness of the wavefunction could be removed at the expense of introducing a vector potential [127]. Mead later coined the term “molecular Aharonov-Bohm effect” to describe the effect of this vector potential [124], a name which was remarkably prophetic given that in four years the work of Berry [31] and Simon [169] would unify the real and molecular Aharonov-Bohm effects via the more general construct of the geometric phase.

It is interesting to note that practically the only situation in which molecular physicists and chemists did not throw out the geometric vector potential was in the analysis of the diatom [133]. In fact, Van Vleck’s 1929 Born-Oppenheimer analysis of the diatom treated precisely the effects of the the geometric vector

potential, without being remarked on [191]. The geometric phase was almost discovered over fifty years early!

Another place the Born-Oppenheimer approximation has enjoyed success is in the analysis of the Jahn-Teller effect [39, 96]. The Jahn-Teller effect is the name given to the result that many molecular complexes do not have their minimum energy state in their position of highest symmetry. Consequently the ligands distort and remove the symmetry. It is of interest to us because the geometric phase has been shown to play a large part in Jahn-Teller systems [59, 83, 84]. More specifically, most formulations of the Jahn-Teller problem involve a simplification where the electronic subsystem generates an effective potential in which the nuclei move, and it is assumed that the nuclei remain on the lowest energy sheet of this effective potential. The paths the nuclei trace out on this lowest energy surface can be looked on as trajectories through an adiabatic parameter space that generate geometric phases. Furthermore, the the geometric vector potential corresponds to the operators of non-adiabaticity that were often dropped from the Born-Oppenheimer approximation [147], so the geometric phase formalism gives a new way to interpret the physics. In §7.2 I review the Jahn-Teller effect and demonstrate its applicability to the  $E \otimes \epsilon$  system, with particular emphasis on the way the geometric phase can be seen to be responsible for a half-integral shift in the angular momentum spectrum for the vibronic wavefunctions.

With this background I examine the case of time-odd coupling in §7.3. That is, I take the coupling between the lattice vibrations and the electronic subsystem to be time-odd. This essentially means coupling the electronic position states to the momentum of the lattice rather than the nuclear positions as is usually done. Time-odd coupling has been almost completely ignored in the literature, right from the beginnings of the Born-Oppenheimer approximation. The reason for this is that it is generally dismissed as negligible [76, 180], although Moore and Stedman have suggested that particularly in cases where the usual interactions are suppressed such a mechanism could contribute significantly [134]. It would thus seem a fertile area for exploration. What makes it particularly intriguing from the perspective of the geometric phase is that we have seen how time-oddness of various types are at the root of all non-trivial geometric phases, and thus one could expect interesting geometric structures to arise within a time-odd Jahn-Teller system. Further, as noted in §7.2, in the  $E \otimes \epsilon$  Jahn-Teller system the geometric phase is responsible for a half-integral shift in angular momenta, and this can be traced to a geometric phase of  $\pi$ . In a time-odd system arbitrary phases are possible, not merely multiples of  $\pi$ , so an intriguing question is raised: What happens to the angular momentum in time-odd systems? Could there be an analog with the anyonic systems that exist in two dimensions and exhibit fractional statistics [200]?

To explore these questions I consider in which point groups a time-odd Jahn-Teller effect could exist and proceed to construct the Hamiltonian for such a

system using the  $E \otimes (b_1 \oplus b_2 \oplus a_2)$  system as an example. I solve this exactly by use of a novel canonical transformation, and obtain the wavefunctions for the system and the form of the geometric vector potentials. Interestingly, the potential that arises is that of the ubiquitous monopole, topological exemplar of time-oddity.

Having solved the  $E \otimes (b_1 \oplus b_2 \oplus a_2)$  system, §7.4 is an examination of how the angular momentum is modified in this system due to the presence of the geometric monopole.

## 7.1 The adiabatic and Born-Oppenheimer approximations

The adiabatic approximation (often called the Born-Oppenheimer approximation) first arose in molecular physics [45] but is applicable to any physical system with two widely separated energy scales. In molecular physics the two energy scales are those of the heavy nuclei and those of the much lighter electrons. The energy levels associated with transitions between electronic states are much larger than those associated with nuclear vibrational and rotational modes. Thus for low energy excitations of the molecule it should be possible to arrive at a description that involves only nuclear degrees of freedom. The electrons are effectively forced to remain in only one state, and are enslaved to the nuclear motion. Another way to see this is to obtain an order of magnitude estimate for the relative velocities of the electrons and the nuclei by applying the uncertainty principle and the virial theorem. If the masses of the nuclei and electrons are  $M$  and  $m$  respectively,  $V$  and  $v$  are the magnitudes of their average velocities, and  $Q$  and  $q$  represent the average distance from equilibrium and the extent of the electron cloud, then we find that [35]

$$\frac{V}{v} \sim \left(\frac{m}{M}\right)^{\frac{3}{4}}, \quad Q \sim q \left(\frac{m}{M}\right)^{\frac{1}{4}}. \quad (7.1)$$

As the mass of the nuclei is  $\sim 10^5$  times larger than the mass of the electrons, it is evident that the nuclei move very slowly compared to the electrons. Consequently the electronic wavefunction has enough time to constantly adapt to the instantaneous position of the nuclei, and we see that that this is a good argument for the enslavement of the electrons to the nuclei. For this reason the degrees of freedom describing the subsystem with low energy gaps are often called the “slow” variables and those describing the large energy gap subsystem are called the “fast” variables.

How applicable the adiabatic approximation is depends on the average relative energy gaps in the fast and slow subsystems. A simple argument [35] shows that

$$\frac{\Delta E_{\text{nucl}}}{\Delta E_{\text{el}}} \sim \left(\frac{m}{M}\right)^{\frac{1}{2}}. \quad (7.2)$$

The approximation is valid provided  $\Delta E_{\text{slow}} \ll \Delta E_{\text{fast}}$  which is certainly the case for molecular systems, provided there is no electronic degeneracy. Unfortunately electronic degeneracy is common, a good example being Jahn-Teller systems, and at this point the adiabatic approximation breaks down.

With this background I now give the derivation of the adiabatic approximation as applied to molecular systems. We assume we have a system that is composed of a massive, slowly varying part (the nuclei in the molecular system) and a lighter part (the electronic subsystem) that quickly adjusts to change in the slow part. Denoting the nuclear position variables by  $\mathbf{Q}$  and the electronic variables by  $\mathbf{q}$  the total Hamiltonian for the system can be written as

$$H = T_n(\mathbf{Q}) + T_e(\mathbf{q}) + V(\mathbf{q}, \mathbf{Q}) = T_n(\mathbf{Q}) + H_e, \quad (7.3)$$

where  $T_e(\mathbf{q}) = -(\hbar^2/2m)\nabla_{\mathbf{q}}^2$  and  $T_n(\mathbf{Q}) = -(\hbar^2/2M)\nabla_{\mathbf{Q}}^2$  are the electronic and nuclear kinetic energies respectively, and  $V(\mathbf{q}, \mathbf{Q})$  is the interaction energy between them. As a first step we solve the electronic Hamiltonian  $H_e$  to obtain a set of electronic basis states  $\psi_n$  for each nuclear position  $\mathbf{Q}$ :

$$H_e\psi_n(\mathbf{q}, \mathbf{Q}) = \epsilon_n\psi_n(\mathbf{q}, \mathbf{Q}). \quad (7.4)$$

This basis is known as the adiabatic basis.

We look for approximate solutions to the total Hamiltonian of the form

$$\Psi(\mathbf{q}, \mathbf{Q}) = \sum_n \psi_n(\mathbf{q}, \mathbf{Q})\chi_n(\mathbf{Q}), \quad (7.5)$$

where

$$H\Psi(\mathbf{q}, \mathbf{Q}) = E\Psi(\mathbf{q}, \mathbf{Q}). \quad (7.6)$$

Substituting (7.5) into (7.6) and integrating out the electronic wavefunctions gives

$$(T_n + \epsilon_n)\chi_n(\mathbf{Q}) - \hbar^2/2M \sum_m c_{nm}(\mathbf{Q})\chi_m(\mathbf{Q}) = E\chi_n(\mathbf{Q}), \quad (7.7)$$

where

$$c_{nm} = \langle \psi_n | \nabla_{\mathbf{Q}} | \psi_m \rangle \cdot \nabla_{\mathbf{Q}} + \frac{1}{2} \langle \psi_n | \nabla_{\mathbf{Q}}^2 | \psi_m \rangle. \quad (7.8)$$

Once again we see the appearance of our old friend the vector potential:

$$\mathbf{A}_{nm} = \langle \psi_n | i \nabla_{\mathbf{Q}} | \psi_m \rangle. \quad (7.9)$$

Using the identity

$$-\langle \psi_n | \nabla_{\mathbf{Q}}^2 | \psi_m \rangle = \sum_k \mathbf{A}_{nk} \cdot \mathbf{A}_{km} + i \nabla_{\mathbf{Q}} \cdot \mathbf{A}_{nm} \quad (7.10)$$

it is possible to rewrite (7.7) as

$$H_{nm}^{\text{eff}}\chi_m = E\chi_n \quad (7.11)$$

where the effective nuclear matrix Hamiltonian is given as [166]

$$H_{mn}^{\text{eff}} = -\frac{\hbar^2}{2M} \sum_k (\delta_{mk} \nabla_Q - i\mathbf{A}_{mk}(Q)) \cdot (\delta_{kn} \nabla_Q - i\mathbf{A}_{kn}(Q)) + \delta_{mn} \epsilon_n(Q). \quad (7.12)$$

In the Born-Oppenheimer approximation the off-diagonal terms are ignored (a reasonable assumption due to the much larger energy difference between electronic states compared to the nuclear states, but one that breaks down for Jahn-Teller systems) and we can rewrite (7.12) as

$$H_n^{\text{BO}} = -\frac{\hbar^2}{2M} (\nabla - i\mathbf{A}_n(Q))^2 + \epsilon_n = E_n. \quad (7.13)$$

Notice the astonishing resemblance to the minimal coupling Hamiltonian describing a charged particle subject to electromagnetic fields. The motion of the nuclei can be looked at as particles of unit charge moving in an electromagnetic potential  $\mathbf{A}$  generated by the topological behaviour of the electrons and a scalar potential corresponding to the energy eigenvalues of the electronic subsystem. Prior to the paper of Mead and Truhlar [127] the terms arising from the geometric vector potential were generally ignored since  $\langle \psi | \nabla^2 | \psi \rangle$  was small, and it was assumed that  $\mathbf{A}$  could be made zero by a suitable choice of phase for the electronic basis vectors. However, as shown in Chapter 4 this is not always the case. For example, in three-dimensional parameter space if  $\mathbf{A}$  has non-zero curl such a basis cannot be chosen with a single-valued choice of phase. More recent treatments of vibronic interactions include these terms and they are known as operators of nonadiabaticity [35].

Except in certain simple geometries it is very difficult to find the adiabatic basis given by (7.4). That is, it is difficult to analytically define how the electronic basis vectors depend on the nuclear positions [152]. If this is the case what is generally done is employ what is called the *crude adiabatic approximation*.

The crude adiabatic approximation involves solving (7.3) with the expansion

$$\Psi^0(q, Q) = \sum_n \psi_n^0 \phi_n(Q) \quad (7.14)$$

where the electronic wavefunctions  $\psi_n^0 = \psi_n^0(q, Q_0)$  are solutions to

$$[H(Q_0) - E_n(Q_0)]\psi_n^0 = 0, \quad (7.15)$$

i.e. for the specialized geometry  $Q = Q_0$ . This is usually taken to be the point of highest nuclear symmetry. Using this approximation (7.7) becomes

$$(T_n + H_{nn'} - E)\phi_n(Q) = 0 \quad (7.16)$$

where  $H_{nn'}(Q)$  are the electronic matrix elements of the potential  $V = V(Q) - V(Q_0)$  expanded in a Taylor series about  $Q_0$ :

$$H_{nn'}(Q) = \sum_i \langle \psi_n^0 | \left( \frac{\partial V}{\partial Q_i} \right)_{Q_0} | \psi_{n'}^0 \rangle Q_i + \sum_i \sum_j \langle \psi_n^0 | \left( \frac{\partial^2 V}{\partial Q_i \partial Q_j} \right)_{Q_0} | \psi_{n'}^0 \rangle Q_i Q_j + \dots \quad (7.17)$$

Except in simple systems this crude adiabatic approach is the one that is normally used. However it is sometimes possible, as we shall see in §7.2, to then diagonalize this effective potential and thereby create a  $\mathbf{Q}$ -dependent adiabatic basis at the expense of introducing off-diagonal terms into the kinetic energy operator in (7.16).

Having now reviewed the Born-Oppenheimer approximation, the workhorse of molecular physics, we are in a position to consider the Jahn-Teller effect, and this will be the focus of the next section.

## 7.2 Jahn-Teller systems

In 1937 Jahn and Teller published a paper which examined the effect of nuclear motion on the electronic states of a molecule [96]. They concluded that “orbital electronic degeneracy and stability of the nuclear configuration are incompatible unless all the atoms of a molecule lie on a straight line.” What this essentially means is that although a molecule in a state of high symmetry will usually have degenerate electronic states due to the symmetry, this symmetric configuration is unstable and the molecule will distort to lower the symmetry and split the energy degeneracies. This process is called the Jahn-Teller effect and molecules that exhibit it are called Jahn-Teller active. The origin of this effect comes from the fact that a displacement of atoms from a symmetric configuration usually produces splittings of a degenerate electronic energy level that are linear with respect to the displacement, while all elastic restoring forces are derived from potentials that are quadratics. Equilibrium is therefore achieved for a non-zero value of the displacement.

Jahn and Teller originally proved their theorem by laboriously considering each of the possible molecular point groups in turn and examining the possible vibrations for the minimal set of atoms which are required to form each particular point group. Since then more general and elegant proofs have been offered [39].

Jahn and Teller’s work was immediately used by van Vleck to examine transition metal complexes [192] and since then has been used in a wide variety of situations in condensed matter physics. More recently it was pointed out by Fletcher that in some conditions it is possible for the degenerate electronic states to be coupled to the nuclei through the momentum of the normal vibrational modes rather than their positions [69, 70]. Moore and Stedman have considered this form of coupling as an example of time-odd coupling [134], which has normally been

ignored in the literature. It is this type of coupling I shall be concerned with later.

Soon after Berry's original paper [31] it became obvious that the geometric phase entered the Jahn-Teller effect in a natural way. It is not evident who was the first to notice this, but the earliest references include Delacrétaz *et al.* [59], Ham [83] and Zwanziger and Grant [211]. It is particularly relevant to the  $E \otimes \epsilon$  system, affecting as it does the orbital angular momentum spectrum [6, 211].

As an illustration of the Jahn-Teller effect will I consider the  $E \otimes \epsilon$  system, where a doubly degenerate electronic  $E$  state is coupled to a pair of vibrational modes transforming with  $E$  symmetry. This case is the best studied of all Jahn-Teller systems, and is beloved of theoreticians everywhere due to its amenability to analytic solution [34, 35, 147, 182], and will set the scene for the cases I will consider later. A physical example of such a system is given by a  $\text{Cu}^{2+}$  ion in octahedral surroundings [139, 142].

The first point to notice is that the Born-Oppenheimer approximation is not strictly applicable as there is electronic degeneracy. Thus we must include at least two electronic states in the expansion given by (7.14). Two nuclear modes can be excited, and their normal coordinates are shown in Figure 7.1.

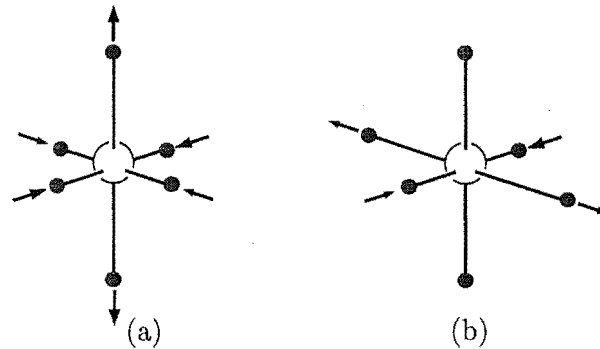


Fig. 7.1: Two E-type normal modes in a lattice with octahedral symmetry. (a) the  $Q_\theta$  mode, (b) the  $Q_\epsilon$  mode

Letting the normal coordinates of these two vibrational modes be denoted by  $Q_1$  and  $Q_2$ , the effective nuclear Hamiltonian in the space of crude adiabatic states  $\psi_1^0, \psi_2^0$  is given by [68]

$$H = -\frac{\hbar^2}{2m} \left( \frac{\partial^2}{\partial Q_1^2} + \frac{\partial^2}{\partial Q_2^2} \right) + \frac{1}{2} m \omega^2 (Q_1^2 + Q_2^2) + \frac{1}{2} L \begin{pmatrix} -Q_2 & Q_1 \\ Q_1 & Q_2 \end{pmatrix}. \quad (7.18)$$

Changing to polar coordinates with

$$Q_1 = Q \cos \phi, \quad Q_2 = Q \sin \phi \quad (7.19)$$

the effective potential governing nuclear motion (equivalent to the effective Hamiltonian (7.12)) is

$$V_{\text{eff}}(\mathbf{Q}) = \frac{1}{2}m\omega^2\mathbf{Q}^2 + \frac{1}{2}LQ \begin{pmatrix} -\cos\phi & \sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \quad (7.20)$$

This effective potential has eigenvalues

$$\epsilon_{\pm}(\mathbf{Q}) = \frac{1}{2}m\omega^2Q^2 \pm \frac{1}{2}LQ. \quad (7.21)$$

These eigenvalues are plotted in figure 7.2. The lower-energy eigenvalue  $\epsilon_-$  has a minimum  $\epsilon_-^0 = L/2m\omega^2$  in the shape of a trough of radius  $Q = L/2\hbar\omega$ . The two surfaces described by  $\epsilon_-$  and  $\epsilon_+$  are called adiabatic potential energy surfaces, and are of great importance in Jahn-Teller calculations. In this case, supposing there is not enough energy for a transition from the lower energy surface to the higher, then the low energy motion of the nuclei will be around the bottom of this trough. The equilibrium point about which the nuclei oscillate is removed from the configuration of highest symmetry  $Q = 0$ .

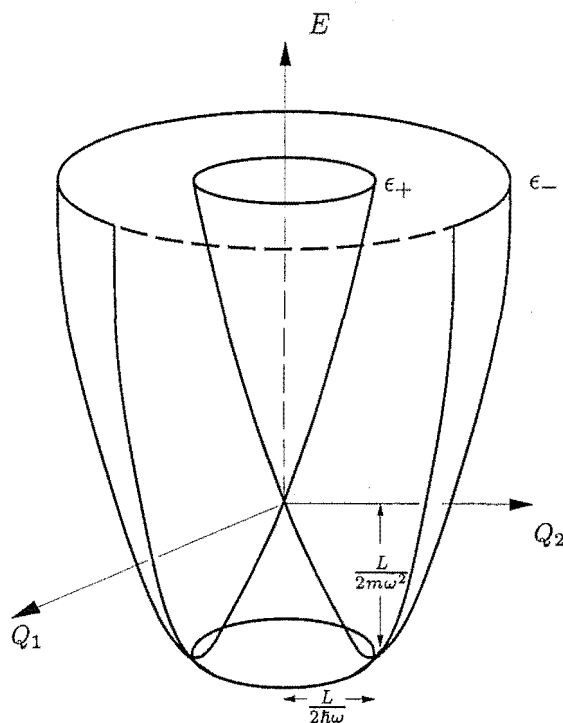


Fig. 7.2: Adiabatic potential surfaces for the  $E \otimes e$  Jahn-Teller system

The next step in our analysis is to obtain the adiabatic basis, which we do by diagonalizing the effective potential (7.20). The eigenvectors corresponding to

$\epsilon_{\pm}$  are given by

$$\psi_- = \cos(\phi/2)\psi_1^0 - \sin(\phi/2)\psi_2^0 \quad (7.22)$$

$$\psi_+ = \sin(\phi/2)\psi_1^0 + \cos(\phi/2)\psi_2^0. \quad (7.23)$$

In this adiabatic basis the effective potential has been diagonalized at the expense of introducing off-diagonal terms into the nuclear kinetic energy operator. Provided  $\epsilon_-$  and  $\epsilon_+$  are well separated this coupling can be ignored, and I shall assume this is the case. Note that the eigenvectors are double-valued in  $\phi$ , changing sign on a complete circuit around the trough.

From this point on I consider only the lower energy sheet, corresponding to  $\epsilon_-$  as this has more interesting dynamics. Solving the effective Schrödinger equation for  $\chi_-$  we find solutions of the form [68]

$$\chi_- = e^{ij\phi} f(Q). \quad (7.24)$$

The total vibronic wavefunction  $\Psi_- = \chi_- \psi_-$  must be single-valued, so due to the double-valued nature of  $\psi_-$  we must require the quantum number  $j$  to be half-odd integer, giving the nuclear angular momentum a fermionic structure. This fractional quantization of the pseudorotation quantum number has been experimentally observed by Delacrétaz *et al.* in the spectrum of gas-phase  $\text{Na}_3$  [59]. It is interesting to note that if the linear coupling terms used above are taken as zero and purely quadratic coupling terms are used, then one has the Renner-Teller Hamiltonian and the angular momentum is shifted by  $\hbar$  rather than  $\frac{\hbar}{2}$  [211].

Another way of understanding this half-integral shift is to consider the angular momentum operator that commutes with the effective Hamiltonian. In the limit of purely linear coupling, the operator

$$J = \begin{pmatrix} -i\frac{\partial}{\partial\phi} + \frac{1}{2} & 0 \\ 0 & -i\frac{\partial}{\partial\phi} - \frac{1}{2} \end{pmatrix} \quad (7.25)$$

commutes with the effective nuclear Hamiltonian (7.18). Considering only the  $z$  component, this angular momentum can be seen as consisting of two parts:  $J_z = L_z + \sigma_z$ . The  $L_z$  term represents the nuclear angular momentum, and consists of nuclear distortions rotating clockwise about the molecule; and the  $\sigma_z$  operator is the “energy-spin” term that represents an electronic charge density wave circling the  $z$ -axis [35]. In the absence of vibrational coupling  $\mathbf{L}$  and  $\sigma/2$  are separately conserved, but in the presence of coupling the angular momentum is free to swap between the two.

This fermionic structure can be explained in terms of the geometric phase [6]. To see this notice that the adiabatic basis was chosen to be real and hence was double-valued. This choice of real basis vectors can always be made for a time-even system. Because the basis vectors given by (7.23) are real the geometric

vector potential  $\langle \psi_{\pm} | i \nabla_Q | \psi_{\pm} \rangle$  in (7.12) vanishes. However, we are not compelled to choose a real, double-valued adiabatic basis. We are free to rephase the basis vectors to make them single-valued:

$$\tilde{\psi}_{-} = e^{i\phi/2} \psi_{-}. \quad (7.26)$$

However now the geometric vector potential does *not* vanish and we are left with a term in the effective nuclear Hamiltonian that looks just like a electromagnetic vector potential.

Calculating this geometric vector potential gives

$$\mathbf{A} = \langle \tilde{\psi}_{-} | i \nabla | \tilde{\psi}_{-} \rangle = \frac{1}{2} \nabla_Q \phi \quad (7.27)$$

which is the potential of a fluxtube of strength  $\pi$ , since  $\nabla \times \mathbf{A} = \pi \delta(\mathbf{r}) \hat{\mathbf{z}}$  in cylindrical polars. (It should be noted that this idea of strength is not universal. For example Aitchison [7] assigns  $\mathbf{A}$  strength of  $\frac{1}{2}$ , which gives the angular momentum shift from the fluxtube rather than the phase change. I feel that my definition is better however, as it gives the phase imparted on circling the fluxtube with a particle of unit charge, and this is the definition I will use for the concept of fluxtube strength from here on.) As in the analogous situation in electromagnetism, the Aharonov-Bohm effect, a circuit about this fluxtube will induce an extra phase of  $\pi$  in the nuclear wavefunction. More formally, the introduction of a vector potential in the effective nuclear Hamiltonian will modify the solution to the old (in this case double-valued) nuclear Hamiltonian  $\chi$  as follows:

$$\tilde{\chi}_{-} = e^{i \int \mathbf{A} \cdot d\mathbf{x}} \chi_{-} = e^{i\pi} \chi_{-}. \quad (7.28)$$

Thus  $j$  in (7.24) must still be chosen as half-odd integer, although now the extra phase is associated with the geometric phase rather than the double-valued electronic wavefunctions.

Finally one can also consider the anomalous angular momentum as arising from the fluxtube. If one takes a fluxtube of strength  $\Phi$  and a particle of charge  $q$  orbiting around it then its associated potential is  $A_{\phi} = \Phi/2\pi$  and the Lagrangian for the system is [166]

$$L = \frac{1}{2} m r^2 \dot{\phi}^2 + \frac{q\Phi}{2\pi} \dot{\phi}. \quad (7.29)$$

The canonical momentum is

$$p_{\phi} = m r^2 \dot{\phi} + \frac{q\Phi}{2\pi} \quad (7.30)$$

giving a Hamiltonian

$$\frac{1}{2m r^2} \left( p_{\phi} - \frac{q\Phi}{2\pi} \right)^2. \quad (7.31)$$

As can be seen, the canonical angular momentum operator has become

$$L_z = -i\frac{\partial}{\partial\phi} - qA_\phi \quad (7.32)$$

since  $\mathbf{r} \times \mathbf{A}|_z = -\Phi/2\pi$ . A number of other methods reaching the same conclusion of an  $L_z$  angular momentum shift of  $qA_\phi$  are given by Wilczek [198].

In the  $E \otimes \epsilon$  Jahn-Teller system instead of an electromagnetic vector potential we are dealing with a geometric vector potential, and hence we take  $q = 1$ . Consequently for a fluxtube potential of the form (7.27) there is a phase shift of  $\pi$  and an associated angular momentum shift of  $\frac{1}{2}$ .

The interesting thing about considering the angular momentum to be altered by the fluxtube is that it allows, in principle, for the angular momentum to take on any value, not just the canonically accessible  $n\hbar/2$ . This is the basis of the existence of anyons, objects with non-half-integral spin and statistics different from the ordinary Fermi-Dirac kind [9, 24, 56, 73, 115, 199]. The reason that these fractional-spin particles can slip around the half-integral quantization rule is that they exist only in two dimensions. In two dimensions the angular momentum algebra has only a single generator and hence is commutative, unlike the three-dimensional case where three generators are required and arbitrary rotations do not commute. Thus the usual appeal to commutation relations to derive the  $l = n\hbar/2$  rule is not valid. Anyons have been applied to a wide variety of two-dimensional physical systems, for example the fractional quantum hall effect [23, 118], high temperature superconductivity [10, 52, 66], and P and T violations in field theories [66, 102]. Their links to the geometric phase have also been explored [91].

The reason this is of interest is that we have seen that a geometric phase of  $\pi$  has given rise to a half-integral shift to the nuclear angular momentum in the  $E \otimes \epsilon$  Jahn-Teller system. The phase shift of  $\pi$  is linked to the fact that this system is time-even but includes a point of degeneracy, or, in the language of Chapter 4, a singularity in the vector potential. However, also from Chapter 4 we know that a time-odd system is capable of exhibiting arbitrary geometric phases. It is thus a natural question to wonder what time-odd coupling in a Jahn-Teller system would do to the nuclear angular momenta. As we are dealing with three-dimensional systems the overall angular momenta certainly must be integral or half-integral, provided the generators for the total vibronic angular momenta satisfy the usual commutation relations, but the possibility exists that the angular momenta may be peculiar in each of the electronic and nuclear subsystems. Indeed, this even seems likely since one of the features of a Jahn-Teller molecule is that the angular momenta of the electrons and nuclei are not separately conserved, although the sum of them, the vibronic angular momentum, is conserved and can be used as a label for a state [105].

To this explore this idea I now turn to Jahn-Teller systems exhibiting time-odd coupling.

### 7.3 Time-odd coupling in Jahn-Teller systems

Since the 1930s the Van Vleck model of electron-lattice coupling has reigned more or less supreme. This model is a quasistatic theory and considers the interactions of atoms with the static coulomb field generated by the ligands. Consequently the interactions are governed purely by electric fields and are time-even. It has been recognized that purely electrostatic calculations of the ligand fields are inadequate for some situations such as the lanthanides, leading to the extension of the Van Vleck model [136], and modifications to dynamic ligand field theories [178]. Such modifications are at best quasistatic, however, and still ignore time-odd coupling. Review articles that even mention time-odd coupling between lattice vibrations and electronic states generally do so only to dismiss the effect as negligible [76, 180]. Among the very few workers to seriously consider time-odd coupling in topics such as spin-lattice relaxation are Fletcher [69], Fletcher and Pooler [70] and in the context of breakdown of sum rules in Ham reduction factors in Jahn-Teller systems Payne and Stedman [146, 144, 145]. More recently the importance of time-odd lattice-electron coupling was reexamined by Moore and Stedman, who concluded that in some instances it could contribute significantly [134], and by Riley and Furlan, who demonstrated which point groups momentum and Barnett coupling can occur in [152].

The simplest way to include time-odd terms in a vibronic system is to allow momentum coupling, that is allow the momentum of the nuclei to have an effect on the electronic states. This scheme is disallowed in the crude adiabatic approximation. Furthermore, not all point groups allow the possibility of momentum coupling, since in some cases a minimum complexity of the molecule is required [152]. If momentum coupling is allowed then in addition to the electric field from the ligands there also exists a magnetic field. A simple example is shown in Figure 7.3. Here the total system has  $D_4$  symmetry, and the electronic states with  $E$  symmetry (in this case given as a pair of electronic  $p$  orbitals) are coupled to an  $A_2$  rotational mode of the lattice ions, shown as shaded black circles. Generally the lattice vibrational modes couple only electrostatically via the Stark effect, but the rotational mode will generate a magnetic field as shown and allow a time-odd Zeeman interaction. We thus see how time-odd coupling is possible outside a quasistatic approximation.

To obtain interesting geometric phases it is desirable to have at least a three-dimensional parameter space, which requires normal modes with at least three degrees of freedom. It is also helpful to have the electronic state only two-fold degenerate, as this limits the complexity of any analysis. Two obvious candidates satisfying these criteria are systems with coupling symmetries of the form  $E \otimes (\epsilon + a_2)$  and  $E \otimes (b_1 \oplus b_2 + a_2)$ . To begin I shall choose the latter, and will solve it exactly.

I label the normal coordinates by the obvious notation  $Q_{b_1}, Q_{b_2}, Q_{a_2}$ . In this

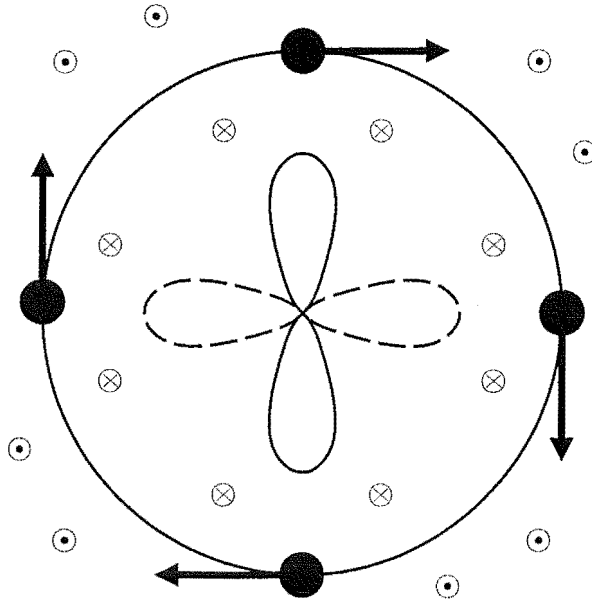


Fig. 7.3: A rotational  $A_2$  mode coupled to a pair of E-type electronic levels, shown here as  $p_x$  and  $p_y$  orbitals. Note the generation of a magnetic field.

notation the total Hamiltonian for the system will be given by [174]

$$H = \frac{1}{2\mu}(P_{b_1}^2 + P_{b_2}^2 + P_{a_2}^2) + \frac{1}{2}m(\omega_b^2(Q_{b_1}^2 + Q_{b_2}^2) + \omega_a^2 Q_{a_2}^2) + K_b(\sigma_1 Q_{b_1} + \sigma_3 Q_{b_2}) + K_a \sigma_2 P_{a_2} \quad (7.33)$$

where the  $\sigma_i$  are the Pauli spin matrices, the  $\omega$  are the vibrational frequencies of the modes,  $\mu$  is the effective mass of the nuclear displacement,  $P_\alpha = -i\hbar\nabla_\alpha$  and  $V_a$  and  $V_b$  are coupling constants. The crude adiabatic approximation cannot immediately be employed with this Hamiltonian, because the matrix elements of the  $\sigma_2 P_{a_2} = -i\hbar\sigma_2 \nabla_{a_2}$  term will vanish in the crude adiabatic basis.

To circumvent this problem I employ a canonical transformation. Canonical transformations are not often considered in quantum mechanics texts, which is a pity as they can be quite useful and offer an insightful look into the transition between classical and quantum mechanics [21, 112, 129]. Canonical transformations have been used previously in Jahn-Teller problems, although not extensively [64, 65, 120, 132, 147].

The quantum mechanical analogue of the classical mechanical time-independent canonical transformation is generally taken to be the unitary transformation

$$H \rightarrow H' = UHU^{-1}. \quad (7.34)$$

This apparent equivalence between unitary and canonical transformations can be traced back to Dirac [61]. Actually, however, unitary transformations represent

only a sub-class of valid canonical transformations [20]. Another point to be aware of in carrying over transformations from the classical to the quantum is extra subtleties arise regarding the ordering of operators since they no longer need commute.

For the current problem we choose, after some trial and error and inspired guesswork,

$$U = \exp \left[ \frac{-i\pi}{4\hbar} \left( \frac{1}{\mu\omega_a} P_a^2 + \mu\omega_a Q_a^2 \right) \right]. \quad (7.35)$$

Using the commutation relations

$$[Q, P^2] = 2i\hbar P, \quad [P, Q^2] = -2i\hbar Q \quad (7.36)$$

along with the identity

$$e^{-S} A e^S = A + \frac{1}{1!} [A, S] + \frac{1}{2!} [[A, S], S] + \dots \quad (7.37)$$

after some algebra we find that under the unitary transformation (7.35) the operators in the Hamiltonian (7.33) transform as

$$\begin{aligned} Q_{b_1} &\rightarrow Q_{b_1} & P_{b_1} &\rightarrow P_{b_1} \\ Q_{b_2} &\rightarrow Q_{b_2} & P_{b_2} &\rightarrow P_{b_2} \\ Q_{a_2} &\rightarrow \frac{-1}{m\omega_a} P_{a_2} & P_{a_2} &\rightarrow \mu\omega_a Q_{a_2} \end{aligned} \quad (7.38)$$

Thus the standard commutation relations still hold, as is required for a canonical transformation [151]. Letting  $Q_1 = Q_{b_1}$ ,  $Q_2 = Q_{a_2}$ ,  $Q_3 = Q_{b_2}$ , under this unitary transformation the old Hamiltonian (7.33) becomes

$$H = \frac{1}{2\mu} (P_1^2 + P_2^2 + P_3^2) + \frac{1}{2} \mu (\omega_b^2 (Q_1^2 + Q_3^2) + \omega_a^2 Q_2^2) + K \sigma_i Q_i \quad (7.39)$$

where I have assumed  $K_b = \mu\omega_a K_a = K$  for convenience. The transformation has removed the momentum dependence in the coupling term and we can now apply the crude adiabatic approximation.

We change to spherical polar coordinates

$$Q_1 = Q \sin \theta \cos \phi, \quad Q_2 = Q \sin \theta \sin \phi, \quad Q_3 = Q \cos \theta \quad (7.40)$$

and proceed as before. The effective potential is given by

$$V_{\text{eff}} = \frac{1}{2} \mu Q^2 (\omega_b^2 (\sin^2 \theta \cos^2 \phi + \cos^2 \theta) + \omega_a^2 \sin^2 \theta \sin^2 \phi) \quad (7.41)$$

$$+ K Q \begin{pmatrix} \cos \theta & \sin \theta e^{-i\phi} \\ \sin \theta e^{i\phi} & -\cos \theta \end{pmatrix} \quad (7.42)$$

which has eigenvalues

$$\epsilon_{\pm} = \frac{1}{2}\mu Q^2(\omega_b^2(\sin^2 \theta \cos^2 \phi + \cos^2 \theta) + \omega_a^2 \sin^2 \theta \sin^2 \phi) \pm \frac{1}{2}KQ. \quad (7.43)$$

The higher energy eigenvalue  $\epsilon_+$  is uninteresting having a minimum at  $Q = 0$  and will not lead to any Jahn-Teller activity. The lowest adiabatic potential energy surfaces corresponding to the other energy eigenvalue  $\epsilon_-$  is more complex. If  $\omega_a < \omega_b$  then there exist two points that are global minima. They lie on the  $Q_2$  axis at  $Q_2 = \pm K/2\mu\omega_a^2$ , and again give rise to no interesting dynamics. If  $\omega_a > \omega_b$  then the global minimum is a ring in the  $Q_2 = 0$  plane, radius  $(Q_1^2 + Q_3^2)^{\frac{1}{2}} = K/2\mu\omega_b^2$ . Thus in this case low energy motions will be confined to a toroidal region with effectively only one degree of freedom. This situation is similar to the  $E \otimes \epsilon$  system where the low-energy motions were also in a ring at the bottom of the trough of the surface in Figure 7.2. Following the calculations through for this case shows that this motion will give rise to phase changes of  $\pm\pi$  upon circling the  $Q_2$  axis, as might have been expected. The most interesting situation exists when  $\omega_a = \omega_b$ . In this case the lowest energy eigenvalue has the form

$$\epsilon_- = \frac{1}{2}\mu\omega^2 Q^2 - \frac{1}{2}KQ. \quad (7.44)$$

Consequently the lowest adiabatic potential energy surface is a sphere of radius  $Q = K/2\mu\omega^2$ . A sphere allows a rich variety of paths in parameter space and consequently we choose  $\omega_a = \omega_b$ .

Next we change to an adiabatic basis. This is accomplished by diagonalizing the effective potential by changing to a new basis given by its eigenvectors

$$\psi_+ = \begin{pmatrix} e^{-i\phi} \cos \frac{\theta}{2} \\ \sin \frac{\theta}{2} \end{pmatrix} \quad \psi_- = \begin{pmatrix} -\sin \frac{\theta}{2} \\ e^{i\phi} \cos \frac{\theta}{2} \end{pmatrix}. \quad (7.45)$$

Notice that these are identical to the basis vectors (5.15, 5.16) in §5.2 which arose from the general two-level crossing and gave rise to a monopole geometric vector potential.

In this new basis the nuclear kinetic operator is no longer diagonal. After some calculation we find its components are given by

$$\langle \psi_+ | \nabla^2 | \psi_+ \rangle = -\frac{1}{Q^2} \left( \frac{1}{4} + \frac{1 + \cos \theta}{2 \sin^2 \theta} \right) \quad (7.46)$$

$$\langle \psi_+ | \nabla^2 | \psi_- \rangle = -\frac{e^{i\phi}}{2Q^2} (\cot \theta + \csc \theta) \quad (7.47)$$

$$\langle \psi_- | \nabla^2 | \psi_+ \rangle = \frac{e^{-i\phi}}{2Q^2} (\cot \theta + \csc \theta) \quad (7.48)$$

$$\langle \psi_- | \nabla^2 | \psi_- \rangle = -\frac{1}{Q^2} \left( \frac{1}{4} + \frac{1 + \cos \theta}{2 \sin^2 \theta} \right) \quad (7.49)$$

Assuming low energy motion, corresponding to massive nuclei compared to the much lighter electrons, the nuclear motion will remain on the lowest adiabatic potential energy surface, as given by (7.44).

Writing the total vibronic wavefunction as  $\Psi_{\pm} = \psi_{\pm}\chi_{\pm}$  we have

$$\left(-\frac{\hbar^2}{2\mu}\nabla^2 + \epsilon_{\pm}^{min}\right)\Psi_{\pm} = E\Psi_{\pm}. \quad (7.50)$$

Denoting the matrix representation of an operator in the adiabatic basis by enclosing it in square brackets (7.50) becomes

$$\left[-\frac{\hbar^2}{2\mu}([\nabla^2] + 2[\nabla]\cdot\nabla + \nabla^2) + \begin{pmatrix} \epsilon_+^{min} & 0 \\ 0 & \epsilon_-^{min} \end{pmatrix}\right] \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix} = E \begin{pmatrix} \chi_+ \\ \chi_- \end{pmatrix}. \quad (7.51)$$

We note the appearance of the geometric vector potential:  $[\nabla] = -i\mathbf{A}$ . In the adiabatic basis we obtain

$$\mathbf{A}_- = i\langle\psi_+|\nabla|\psi_+\rangle = \frac{-(1 + \cos\theta)}{2r \sin\theta} \hat{\phi} \quad (7.52)$$

This is the vector potential for a monopole of charge +1.

We ignore transitions between electronic levels enabling us to neglect the off-diagonal parts of the kinetic energy operator above. The pair of differential equations can thus be uncoupled and considered separately. Since we are considering low-energy motion, the system will remain on the spherical lowest adiabatic potential energy surface, in this case a sphere, and  $Q$  will remain constant. With this in mind and considering only the more interesting lower energy surface, (7.51) becomes

$$\begin{aligned} -\frac{\hbar^2}{2\mu} \frac{1}{Q^2} \left( -\frac{1}{4} - \frac{1 + \cos\theta}{2 \sin\theta} + \frac{\partial^2}{\partial\theta^2} + \cot\theta \frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial\phi^2} \right. \\ \left. + \frac{i(1 + \cos\theta)}{\sin^2\theta} \frac{\partial}{\partial\phi} \right) \chi_- = (E - \epsilon_-^{min})\chi_-. \end{aligned} \quad (7.53)$$

The way to separate variables is to take the  $\phi$  dependence as  $\chi_- \propto \exp(im\phi)$ . Because we have included the vector potential explicitly, we must choose  $m$  to be an integer in order to preserve the single-valuedness of the total vibronic wavefunction  $\Psi_-$ . Odd angular momenta will still result however, as the inclusion of the vector potential affects the form of the angular momentum operators.

Separating the variables in this way yields the following equation in  $\theta$ :

$$\left( \frac{\partial^2}{\partial\theta^2} + \cot\theta \frac{\partial}{\partial\theta} - \frac{2m^2 + (1 + 2m)(1 + \cos\theta)}{2 \sin^2\theta} - \frac{1}{4} + \frac{2\mu Q^2}{\hbar^2} (E - \epsilon_-^{min}) \right) \chi_- = 0. \quad (7.54)$$

An identical equation arises in Apse *et al.*'s study of the  $\Gamma_8 \otimes (\tau_2 \oplus \epsilon)$  Jahn-Teller system.

The solution to (7.54) can be obtained using the following method. First, employ the change of variable

$$x = \frac{1}{2}(1 + \cos \theta) \quad (7.55)$$

transforming (7.54) into the form

$$\left[ x(1-x) \frac{d^2}{dx^2} - (2x-1) \frac{d}{dx} - \frac{m^2 + (1+2m)x}{4x(1-x)} - \frac{1}{4} + E' \right] \chi_-(x) \quad (7.56)$$

where for simplicity I have rewritten the total energy as  $E' = 2\mu Q^2(E - \epsilon_{-}^{\min})/\hbar^2$ . (7.56) is not in the form of any standard differential equations, so in order to progress we introduce the dependent function  $X(x)$  via

$$\chi_-(x) = x^{\lambda_1}(x-1)^{\lambda_2}X(x) \quad (7.57)$$

where  $\lambda_1$  and  $\lambda_2$  are arbitrary parameters. After considerable algebra one obtains the following, and apparently uglier differential equation:

$$\left[ x(1-x) \frac{d^2}{dx^2} + (2\lambda_1 + 1 - 2(\lambda_1 + \lambda_2 + 1)x) \frac{d}{dx} + \frac{\lambda_1 + \lambda_1(\lambda_1 - 1)(1-x)}{x} + \frac{(\lambda_2^2 + \lambda_2)x - \frac{1}{4}(m+1)^2 - \lambda_2}{1-x} - 2\lambda_1\lambda_2 - 2\lambda_1 + E' \right] X(x) = 0. \quad (7.58)$$

However, in this form we can see that

$$\lambda_1 = \frac{m}{2}, \quad \lambda_2 = \frac{1}{2}(1+m) \quad (7.59)$$

reduces (7.58) to

$$\left[ x(1-x) \frac{d^2}{dx^2} + [(1+m) - (2m+3)x] \frac{d}{dx} - [(1+m)^2 - E'] \right] X(x) = 0. \quad (7.60)$$

In this form it is recognizable as a specialization of the hypergeometric differential equation which is defined as [88]

$$\left[ z(1-z) \frac{d^2}{dz^2} + [\gamma - (\alpha + \beta + 1)z] \frac{d}{dz} - \alpha\beta \right] w(z) = 0 \quad (7.61)$$

where  $z$  is complex and  $\alpha, \beta, \gamma$  are constants. The hypergeometric differential equation has solutions based on combinations of the Gauss hypergeometric series [1]

$$F(\alpha, \beta; \gamma; z) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{n=0}^{\infty} \frac{\Gamma(\alpha+n)\Gamma(\beta+n)}{\Gamma(\gamma+n)} \frac{z^n}{n!} \quad (7.62)$$

which has a radius of convergence  $|z| < 1$ . On the circle  $|z| = 1$  convergence of the series is conditional, and depends on the values of the constants  $\alpha, \beta$  and  $\gamma$ .

Comparing (7.60) with (7.61) we identify

$$\gamma = m + 1, \quad \alpha + \beta = 2(m + 1), \quad \alpha\beta = (1 + m)^2 - E'. \quad (7.63)$$

Solving these equations for  $\alpha$  and  $\beta$  yields a solution

$$\alpha = m + 1 - \sqrt{E'} \quad (7.64)$$

$$\beta = m + 1 + \sqrt{E'}. \quad (7.65)$$

Returning to the first change of variable (7.55), we see that the range of the original variable  $0 \leq \theta \leq \pi$  corresponds to  $0 \leq x \leq 1$ . In order for the wavefunction to be finite throughout this range it is necessary that (7.62) converge at  $x = 1$ . This can only be accomplished if the series terminates after a finite number of terms, which requires  $\alpha$  to be either zero or a negative integer. Setting  $\alpha = -n$  and using (7.64) yields the energy condition

$$E' = (n + m + 1)^2. \quad (7.66)$$

Because (7.61) is second order, two linearly independent solutions exist. Since  $\gamma$  is an integer the first solution has the form [137]

$$\begin{aligned} X_1(x) &= F(\alpha, \beta; \gamma; x) \\ &= F(-n, 2(m + 1) + n; m + 1; x) \\ &= \frac{n!m!}{(n + m)!} P_n^{(m, m+1)}(1 - 2x) \end{aligned} \quad (7.67)$$

where I have used the identity (15.4.6) in Abramowitz and Stegun [1] to rewrite the hypergeometric series in terms of Jacobi polynomials. The second solution is of the form

$$X_2(x) = x^{-m} \frac{(n + m)!(1 - n)!}{(1 - m)!} P_{n+m}^{(-m, m+1)}(1 - 2x) \quad (7.68)$$

and should be discarded as unphysical due to it becoming singular for  $n, m > 1$ .

The full solution can thus consist only of  $X_1$  and can be written, after some simplifications, as

$$\begin{aligned} \chi_- &= e^{im\phi} x^{\frac{m}{2}} (x - 1)^{\frac{1}{2}(m+1)} \frac{n!m!}{(n + m)!} P_n^{(m, m+1)}(1 - 2x) \\ &= (-1)^{n+1} \frac{n!m!}{(n + m)!} e^{im\phi} \cos^m\left(\frac{\theta}{2}\right) \sin^{m+1}\left(\frac{\theta}{2}\right) P_n^{(m+1, m)}(\cos\theta). \end{aligned} \quad (7.69)$$

Finally, the spectrum (7.66) agrees with that obtained by O'Brien and Pooler in their study of the dynamic  $\Gamma_8 \otimes \tau_2$  Jahn-Teller system [140], providing further evidence that (7.66) and (7.69) do indeed provide the correct solution.

## 7.4 Angular momenta in the $E \otimes (b_1 \oplus b_2 \oplus a_2)$ system

As might be expected, due to the vibronic interaction between the nuclei and electrons the standard angular momentum relations have become tangled, just as in the  $E \otimes \epsilon$  system. Once again the vibronic angular momentum has acquired a half-integral spectrum, although this time the cause can be looked at differently. The simplest way to see that the angular momentum has shifted in this way is to note that the geometric vector potential present in this system is that of a monopole, and it has long been known that the presence of a magnetic monopole does peculiar things to standard angular momentum relations [156]. The simplest way to see this in a classical context is to note that the equation of motion for a particle of charge  $q$  moving in the background field of a monopole of the form (5.1) is given by

$$m\ddot{\mathbf{r}} = q\dot{\mathbf{r}} \times \frac{g\mathbf{r}}{r^3}. \quad (7.70)$$

Using this we find that the standard mechanical angular momentum is no longer conserved:

$$\begin{aligned} \frac{d}{dt}(\mathbf{r} \times m\dot{\mathbf{r}}) &= \mathbf{r} \times m\ddot{\mathbf{r}} \\ &= \frac{qg}{r^3}\mathbf{r} \times (\dot{\mathbf{r}} \times \mathbf{r}) \\ &= \frac{d}{dt}(qg\hat{\mathbf{r}}). \end{aligned} \quad (7.71)$$

Thus the standard conserved quantity now has an extra term consisting of angular momentum running between the monopole and the electric charge, with its magnitude quantized in units of the monopole charge. This extra angular momentum is of course carried by the electromagnetic field. The simple argument outlined above can be carried out much more rigorously via canonical arguments and can be shown to hold quantum mechanically [78], with the correct angular momentum operators being given by [167]

$$J_i = -i\epsilon_{ijk}r_j(\nabla_k - iA_k) - \frac{1}{2}r_i\epsilon_{lmn}r_lF_{mn} \quad (7.72)$$

where  $F_{mn}$  are components of the electromagnetic field tensor. Using (5.1) as the magnetic field in (7.72) and evaluating the  $J_z$  component we find

$$J_z = -i\left(x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}\right) - g \quad (7.73)$$

showing that the angular momentum spectrum is shifted by an amount equal to the monopole strength.

As we have seen, the monopoles arising from geometric phases have their strengths quantized in units of spin, and the archetypal version which arises in the  $E \otimes$

$(b_1 \oplus b_2 \oplus a_2)$  Jahn-Teller system I studied in §7.3 has strength one-half. Thus this system exhibits a shift of vibronic angular momentum of one-half also, just as in the  $E \otimes \epsilon$  system. Another way of seeing this shift is to note that extended angular momentum operator

$$J_z = L_z + \frac{1}{2}\sigma_z \quad (7.74)$$

$$= -i\frac{\partial}{\partial\phi} + \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (7.75)$$

commutes with the effective Hamiltonian (7.39).

It must be remembered that I solved the  $E \otimes (b_1 \oplus b_2 \oplus a_2)$  Hamiltonian via a canonical transformation, and as such the  $Q_{a_2}$  position and momentum coordinates have effectively been interchanged. Since a canonical transformation does not affect eigenvalues our reasoning about the angular momentum spectrum still holds for the original, untransformed problem. When constructing excursions on the spherical minimal energy surface for geometric phase considerations, however, it is important to realize that the  $Q_2$  coordinate is actually the momentum of the  $A_2$  vibrational mode. Thus, for example, if one wished to consider evolving the system on an equatorial path in the  $Q_1 - Q_3$  plane in order to examine the phases generated, it must be borne in mind that this corresponds to a zero momentum  $a_2$  mode, even though it need not be in its equilibrium  $Q_{a_2} = 0$  position.

## 8. Conclusion

Time reversal symmetry in quantum mechanics is an area in which much confusion exists, particularly regarding the form and action of the time reversal operator. I have surveyed the use of time reversal symmetry in the literature and pointed out the main sources of confusion and disagreement. My conclusion is that two major forms of the time reversal operator are used in the literature and that both are valid, although in various situations one or the other may prove more useful. I have also discussed the extension of time reversal symmetry to time-dependent Hamiltonians.

The primary result of this thesis is the formulation of various selection rules and existence conditions on the geometric phase by utilizing the ideas of time reversal symmetry. Building on the work of Ihm [93] I have proved that the existence of a non-zero geometric phase can be traced to the breakdown of time reversal symmetry in some form (Chapter 4). Specifically, every system can be broken down into one of five cases:

- Case 1: The system in question is time reversal symmetric (TRS), non-degenerate, and the geometric gauge potential generated by the system contains no singularities. There is no non-zero geometric phase.
- Case 2: The system is TRS, has no singularities, and is degenerate for reasons other than time reversal symmetry (that is, it does not exhibit Kramers degeneracy). Once again there is no non-zero geometric phase.
- Case 3: The system is apparently TRS, but gives rise to singularities in the geometric gauge potential. The singularities can be attributed to the existence of a “geometric flux” analogous to a magnetic flux, that breaks the TRS of the Hamiltonian and can generate phases that are multiples of  $\pi$ .
- Case 4: The system has spontaneous broken time reversal symmetry. In other words it is degenerate due to time reversal symmetry (Kramers degeneracy). In this case the phase is not constrained and arbitrary geometric phases are possible.
- Case 5: The Hamiltonian for the system is explicitly TRS-violating. Again, in this situation there is no constraint on the value of the geometric phase.

These results are completely general, holding for both nonadiabatic evolutions and arbitrary-dimensional parameter spaces. After proving this theorem I devote the remainder of this thesis to exploring some of the consequences of applying time reversal symmetry violation to the geometric phase.

I review the general two-level crossing, a two-level system whose states become can become degenerate for some value of a parameter dependent Hamiltonian, and show that a geometric potential functionally identical to the Dirac monopole is generated (§5.2). The two-level crossing conditions are exactly met by a Kramers doublet, falling into Case 4 above. I examine this situation mathematically, introducing tools I use in Chapter 6. From the mathematical description of the Kramers doublet I derive the half-integral spin quantization condition in a novel way in §5.4, using monopole topology and the geometric phase rather than the usual group theory or algebraic approach.

I extend the link between monopoles, Kramers doublets and the geometric phase in Chapter 6 by considering a pair of Kramers doublets that have the potential to become four-fold degenerate for some value of the parameters that govern their evolution. The Hamiltonian describing such a system is constructed from time reversal symmetry considerations, and depends on five parameters. From this Hamiltonian I derive explicit forms for the resulting pair of five-dimensional nonabelian gauge potentials, and show that they correspond to the two classes of the SU(2) monopole described by Yang [205]. Furthermore, utilizing a conformal transformation I demonstrate in §6.5 that the potentials are equivalent to those describing the SU(2) instanton. As a consequence I show the potentials describing the SU(2) monopole and instanton arise in molecular physics, and demonstrate this using the examples of a  $\text{CH}_4^+$  ion near its  $T_d$  configuration and the  $\Gamma_8 \otimes (\tau_2 \oplus \epsilon)$  Jahn-Teller system.

Finally, in Chapter 7 I examine the relevance of the geometric phase to the Jahn-Teller effect in the case of time-odd coupling between the ion and the lattice. Time-odd coupling has been mostly ignored in the literature, with its effects generally written off as small [76, 180]. Nonetheless, in some cases these effects can become significant [134], and in such an environment interesting gauge structures can arise and are certainly worth studying. I construct such a time-odd Jahn-Teller system, the  $E \otimes (b_1 \oplus b_2 \oplus a_2)$  system, and use the Born-Oppenheimer approximation to find the wavefunctions, energy spectrum and geometric potentials associated with it. The gauge potentials take the form of the U(1) monopole, a structure apparently ubiquitous in the field of time-odd geometric phases. As a consequence of the monopole potentials, the orbital (lattice) angular momentum spectrum is altered by a factor of  $\frac{1}{2}$ , analogous to the similar shift in the  $E \otimes \epsilon$  Jahn-Teller system, despite the arbitrariness of phase suggesting that other anomalous angular momentum shifts may result.

There is certainly room for extensions to the results reported here. It would be interesting to examine other level-crossing systems, say a triplet of Kramers

---

pairs becoming degenerate six-fold degenerate, and see if still higher-dimensional monopole analogues are produced, perhaps based on third Chern class numbers.

Further work is definitely needed on the interplay between the geometric phase and molecular physics. Mead, who has perhaps studied this field more than any other, has stated that “the study of the effects of the nonabelian geometric gauge potential in molecules is . . . in its infancy” [126] and this is entirely true. Very few problems in molecular physics are currently amenable to analytic solution, and while numerical *ab initio* calculations are reaching astonishing levels, development of new theoretical techniques is always valuable. The introduction of the geometric phase into molecular physics is a step in this direction, and there remains much to explore in the overlap between the two.

The effects of time-odd coupling are particularly neglected. The relationship between geometric phase and the angular momentum shift is especially interesting, and I remain convinced that a link between anyons and geometric phases can be made in a Jahn-Teller context, perhaps in a system with two nuclear degrees of freedom rather than three. Even with these anyon questions set aside, time-odd coupling remains an interesting area with much unexplored potential. My analysis in §7.3, while not pretending to numerical accuracy, does give the flavour of the physical effects produced, and suggests that further work within this topic would be worthwhile.



# Bibliography

- [1] Abramowitz M and Stegun I A 1970 *Handbook of Mathematical Functions* (Dover Publications).
- [2] Adams F C, Fatuzzo M, Freese K, Tarlé G and Watkins R 1993 *Extension of the Parker bound on the flux of magnetic monopoles* Phys. Rev. Lett. **70** 2511–2514.
- [3] Aharonov Y and Anandan J 1987 *Phase change during a cyclic quantum evolution* Phys. Rev. Lett. **58** 1593–1596.
- [4] Aharonov Y and Bohm D 1959 *Significance of electromagnetic potentials in the quantum theory* Phys. Rev. **115** (August) 485–491.
- [5] Aharonov Y, Coleman S, Goldhaber A, Nussinov S, Popescu S, Reznik B, Rohrlich D and Vaidman L 1994 *Aharonov-Bohm and Berry phases for a quantum cloud of charge* Phys. Rev. Lett. **73** 918–921.
- [6] Aitchison I J R August 1987 *Berry's topological phase in quantum mechanics and quantum field theory* (Helsinki) Seventh General Conference of the European Physical Society.
- [7] Aitchison I J R 1988 *Berry's topological phase in quantum mechanics and quantum field theory* Physica Scripta **T23** 12–20.
- [8] Aitchison I J R 1989 *Berry phases, magnetic monopoles and Wess-Zumino terms or, How the skyrmion got its spin* Acta Phys. Polon. **B18** 207.
- [9] Aitchison I J R and Mavromatos N E 1991 *Anyons* Contemporary Physics **32** 219–233.
- [10] Aksenov V L, Plakida N M and Yuschkhai Y Y (eds) 1990 *The search for anyon superconductivity* (Proceedings of the International Seminar on High Temperature Superconductivity) (JINR, Dubna).
- [11] Alekseev V A Ya, Zel'dovich B and Sobel'man I I 1976 *Parity non-conservation effects in atoms* Sov. Phys. Usp. **19** 207–219.
- [12] Altarev I S *et al.* 1992 *New Measurement of the electric dipole moment of the neutron* Phys. Lett. B **276** 242–246.

- [13] Anandan J 1988 *Comment on geometric phases for classical field theories* Phys. Rev. Lett. **60** 2555.
- [14] Anandan J 1988 *Non-adiabatic non-Abelian geometric phase* Phys. Lett. A **133** 171–175.
- [15] Anandan J 1992 *The geometric phase* Nature **360** (November) 307–313.
- [16] Anandan J 1994 *Topological and geometrical phases due to gravitational field with curvature and torsion* Phys. Lett. A **195** 284–292.
- [17] Anandan J and Aharonov Y 1988 *Geometric quantum phase and angles* Phys. Rev. D **38** 1863–1870.
- [18] Anandan J and Mazur P O 1993 *Geometric phase for the relativistic Klein-Gordon equation* Phys. Lett. A **173** 116–120.
- [19] Anandan J and Stodolsky L 1987 *Some geometrical considerations of Berry's phase* Phys. Rev. D **35** 2597–2600.
- [20] Anderson A 1993 *Quantum canonical transformations. Physical equivalence of quantum theories* Phys. Lett. B **305** 67–70.
- [21] Anderson A 1994 *Canonical transformations in quantum mechanics* Ann. Phys. **232** 292–331.
- [22] Apsel S E, Chancey C C and O'Brien M C M 1992 *Berry phase and the  $\Gamma_8 \otimes (\tau_2 \oplus \epsilon)$  Jahn-Teller system* Phys. Rev. B **45** 5251–5261.
- [23] Arovas D P, Schrieffer J R and Wilczek F 1984 *Fractional statistics and the quantum Hall effect* Phys. Rev. Lett. **53** 722–723.
- [24] Arovas D P, Schrieffer J R, Wilczek F and Zee A 1985 *Statistical mechanics of anyons* Nucl. Phys. **B251** 117–126.
- [25] Avron J E, Sadun L, Segert J and Simon B 1988 *Topological invariants in Fermi systems with time-reversal invariance* Phys. Rev. Lett. **61** 1329–1332.
- [26] Avron J E, Seiler R and Yaffe L G 1987 *Adiabatic theorems and applications to the quantum Hall effect* Commun. Math. Phys. **110** 33–49.
- [27] Barron L D 1982 *Molecular light scattering and optical activity* (Cambridge University Press).
- [28] Bates D R (ed.) 1962 *Quantum theory - radiation and high energy physics* vol. 3 (Academic Press, New York).
- [29] Belavin A A, Polyakov A M, Schwartz A S and Tyupkin Yu S 1975 *Pseudoparticle solutions of the Yang-Mills equations* Phys. Lett. B **59** 85–87.

- [30] Bernstein H J and Phillips A V 1981 *Fibre bundles and quantum theory* Sci. Am. **245** 122–137.
- [31] Berry M V 1984 *Quantal phase factors accompanying adiabatic changes* Proc. R. Soc. Lond. A **392** 45–57.
- [32] Berry M V 1985 *Classical adiabatic angles and quantal adiabatic phase* J. Phys. A **18** 15–27.
- [33] Berry M V 1990 *Anticipations of the geometric phase* Physics Today (Dec) 34–40.
- [34] Bersuker I B 1984 *The Jahn-Teller effect and Vibronic Interactions in Modern Chemistry* (Plenum, New York).
- [35] Bersuker I B and Polinger V Z 1989 *Vibronic Interactions in Molecules and Crystals* Springer Series in Chemical Physics 49. (Springer-Verlag).
- [36] Bhandari R 1997 *Polarization of light and topological phases* Physics Reports **281** 2–64.
- [37] Biswas S N 1989 *A realization of Berry's non-abelian gauge field* Phys. Lett. B **228** 440–442.
- [38] Blatt J M 1956 *Time reversal* Sci. Am. (August).
- [39] Blount E I 1971 *The Jahn-Teller theorem* J. Math. Phys. **12** 1890–1896.
- [40] Böhm A 1993 *Quantum Mechanics* (Springer Verlag, New York).
- [41] Bohm A, Antoniou I and Kielanowski P 1994 *The preparation-registration arrow of time in quantum mechanics* Phys. Lett. A **189** 442–448.
- [42] Bohm A, Kendrick B and Loewe M E 1992 *The Berry phase in molecular physics* Int. J. Quant. Chem. **41** 53.
- [43] Bohm A, Kendrick B, Loewe M E and Boya L J 1992 *The Berry connection and the Born-Oppenheimer method* J. Math. Phys. **33** 977–989.
- [44] Bohm A and Mostafazadeh A 1994 *The relation between the Berry phase and the Anandan-Aharonov connections for  $U(N)$  bundles* J. Math. Phys. **35** 1463–1470.
- [45] Born M and Oppenheimer J J 1927 *Quantum theory of molecules* Ann. Phys. Lpz. **48** 457–484.
- [46] Busch P, Lahti P and Mittelstaedt P 1991 *The quantum theory of measurement* (Springer).
- [47] Cabrera B 1982 *First results from a superconductive detector for moving magnetic monopoles* Phys. Rev. Lett. **48** 1378–1381.

- [48] Callan C G, Dashen R and Gross D J 1976 *The structure of the gauge theory vacuum* Phys. Lett. **63B** 334–336.
- [49] Callan C G, Dashen R and Gross D J 1978 *Toward a theory of the strong interactions* Phys. Rev. D **17** 2717–2741.
- [50] Capri A Z 1985 *Nonrelativistic Quantum Mechanics* (Benjamin/Cummings Publishing Company Inc).
- [51] Ceulemans A and Szopa M 1991 *The Berry phase for a threefold degenerate state* J. Phys. A **24** 4495–4509.
- [52] Chen Y H, Wilczek F, Witten E and Halperin B I 1989 *On anyon superconductivity* Int. J. Mod. Phys. **B3** 1001–1067.
- [53] Choi H-M and Ji C-R 1997 *Non-Abelian Berry's phase in a slowly rotating electric field* J. Math. Phys. **38** 1281–1288.
- [54] Christenson J H, Cronin J W, Fitch V L and Turley R 1964 *Evidence for the  $2\pi$  decay of the  $K_2^0$  meson* Phys. Rev. Lett. **13** 138–140.
- [55] Cina J A 1986 *Classical adiabatic angle and geometrical phase in spin precession* Chem. Phys. Lett. **132** 393–395.
- [56] Córtes J L, Gamboa J and Velázquez L 1994 *Electromagnetic interaction of anyons in nonrelativistic quantum field theory* Int. J. Mod. Phys. A **9** 953–967.
- [57] De Alfaro V, Fubini S and Furlan G 1976 *A new classical solution of the Yang-Mills field equations* Phys. Lett. **65B** 163–166.
- [58] Debrus J and Hirschfeld A C 1991 *Geometry and Theoretical Physics* (Springer-Verlag, Berlin).
- [59] Delacrétaz G, Grant E T, Whetten R L, Wöster L and Zwanziger J W 1986 *Fractional quantization of molecular pseudorotation in  $Na_3$*  Phys. Rev. Lett. **56** 2598–2601.
- [60] Dirac P A M 1931 *Quantised singularities in the electromagnetic field* Proc. Roy. Soc. **A133** 60.
- [61] Dirac P A M 1958 *Principles of Quantum Mechanics* (Oxford University Press).
- [62] Dirac P A M 1978 *Directions in Physics* (John Wiley and Sons).
- [63] Doughty N A 1990 *Lagrangian Interaction* (Addison Wesley).
- [64] Dunn J L 1988 *Strongly coupled orbital triplet Jahn-Teller systems* J. Phys. C **21** 383–399.

- [65] Dunn J L and Bates C A 1995 *Analysis of the  $T_{1u} \otimes h_g$  Jahn-Teller system as a model for  $C_{60}$  molecules* Phys. Rev. B **52** 5996–6005.
- [66] Dzyaloshinskii I E 1991 *Space and time parity violation in anyonic and chiral systems* Phys. Lett. A **155** 62–64.
- [67] Elliot C J and Small J G 1984 *Optical activity and parity violation* Proc. SPIE Int. Soc. Opt. Eng. **487** 128–137.
- [68] Englman R 1972 *The Jahn-Teller effect in Molecules and Crystals* (Wiley Interscience).
- [69] Fletcher J R 1981 *Reduction factors for a Jahn-Teller cluster model with momentum coupling* J. Phys. C: Solid State Phys. **14** L491–L492.
- [70] Fletcher J R and Pooler D R 1982 *The Jahn-Teller and Barnett effects* J. Phys. C **15** 2695–2707.
- [71] Fock V 1928 *Generalisation and solution of Dirac's statistical equation* Z. Phys. **49** 323–357.
- [72] Frampton P H 1987 *Gauge Field Theories* (Benjamin/Cummings).
- [73] Frölich J and Marchetti P-A 1988 *Quantum field theory of anyons* Letters in Mathematical Physics **16** 347–358.
- [74] Garrison J C and Chiao R Y 1988 *Geometrical phases from global gauge invariance of nonlinear classical field theories* Phys. Rev. Lett. **60** 165–168.
- [75] Gell-Mann M and Pais A 1955 *Behaviour of neutral particles under charge conjugation* Phys. Rev. **97** 1387–1389.
- [76] Gill J C 1975 *The establishment of thermal equilibrium in paramagnetic crystals* Rep. Prog. Phys. **38** 91–150.
- [77] Giller S, Gonera C and Kosiński P 1994 *Berry's phase and T invariance* Acta Phys. Polon. **B 25** 799–807.
- [78] Goddard P and Olive D I 1976 *Magnetic monopoles in gauge theories* Rep. Prog. Phys. **41** 1357–1437.
- [79] Gold T (ed.) 1967 *The Nature of Time* ch. XIII (Cornell University Press).
- [80] Goldberger M L and Watson K M 1975 *Collision Theory* (Kreiger, New York).
- [81] Goldhaber A S and Trower W P 1990 *Resource letter MM-1: Magnetic monopoles* Am. J. Phys. **58** 429–439.

- [82] Gottfried K 1966 *Quantum Mechanics* vol. 1 (Benjamin, New York).
- [83] Ham F S 1987 *Berry's geometrical phase and the sequence of states in the Jahn-Teller effect* Phys. Rev. Lett. **58** 725–728.
- [84] Ham F S 1990 *The role of Berry's phase in ordering the low-energy states of a  $T \otimes \tau_2$  Jahn-Teller system in strong coupling* J. Phys. Condens. Matter **2** 1163–1177.
- [85] Hannay J H 1985 *Angle variable holonomy in adiabatic excursion of an integrable Hamiltonian* J. Phys. A **18** 221–230.
- [86] Heine V 1960 *Group Theory in Quantum Mechanics* (Pergamon Press).
- [87] Herzberg G and Longuet-Higgins H C 1963 *Intersection of potential energy surfaces in polyatomic molecules* Discuss. Faraday Soc. **35** 77–79.
- [88] Hildebrand F B 1976 *Advanced Calculus for Applications* (Prentice-Hall, New Jersey).
- [89] Hong-Mo C and Tsun T S 1993 *Some Elementary Gauge Theory Concepts* (World Scientific, Singapore).
- [90] Hutchison K 1995 *Temporal asymmetry in classical mechanics* Brit. J. Phil. Sci. **46** 219–234.
- [91] Ihm J 1991 *Berry's phase originated from the broken time-reversal symmetry: Theory and application to anyon superconductivity* Phys. Rev. Lett. **67** 251–254.
- [92] Ihm J 1992 *Origins of Berry phase* AAPPS Bulletin **2** 2–6.
- [93] Ihm J 1993 *Broken time-reversal symmetry and Berry's Phase* Int. J. Mod. Phys. B **7** 2109–2146.
- [94] Ihm J 1993 *Symmetry structures and existence criteria of geometric phases* J. Kor. Phys. Soc. **26** 552–556.
- [95] Jackiw R and Rebbi C 1976 *Vacuum periodicity in a Yang-Mills quantum theory* Phys. Rev. Lett. **37** 172–175.
- [96] Jahn H A and Teller E 1937 *Stability of polyatomic molecules in degenerate electronic states* Proc. Roy. Soc. London, Ser. A **161** 220–235.
- [97] Johnsson M T and Aitchison I J R 1997 *The  $SU(2)$  instanton and the adiabatic evolution of two Kramers doublets* J. Phys. A **30** 2085–2092.
- [98] Jordan T F 1969 *Linear operators in quantum mechanics* (John Wiley & Sons, Inc, New York).
- [99] Kaempffer F A 1965 *Concepts in Quantum Mechanics* (Academic Press).

- [100] Kapitulnik A, Dodge J S and Fejer M M 1994 *High-resolution magneto-optical measurements with a Sagnac interferometer* J. Appl. Phys. **75** 6872–6877.
- [101] Kendrick B and Pack R T 1997 *Geometric phase effects in the resonance spectrum, state-to-state transition probabilities and bound state spectrum of HO<sub>2</sub>* J. Chem. Phys. **106** 3519–3539.
- [102] Kitazawa Y 1990 *P, T-breaking properties of anyons* Phys. Rev. Lett. **65** (September) 1275–1278.
- [103] Kivelson S A and Rokhsar D S 1988 *Quasiparticle statistics in time-reversal invariant states* Phys. Rev. Lett. **61** 2630–2631.
- [104] Kocher C A 1993 *Interactions of magnetic monopoles with electric currents* Am. J. Phys. **61** 875–878.
- [105] Koizumi H 1996 *Geometric phase in Jahn-Teller crystals* Phys. Rev. Lett. **76** 2370–2373.
- [106] Koizumi H and Sugano S 1994 *The geometric phase in two electronic level systems* J. Chem. Phys. **101** 4903–4913.
- [107] Koizumi H and Sugano S 1995 *Geometric phase in two Kramers doublets molecular systems* J. Chem. Phys. **102** 4472–4481.
- [108] Kozlov M G and Porsev S G 1989 *The possibility to study the break of time reversal invariance in atoms* Phys. Lett. A **142** 233–236.
- [109] Kunstatter G 1984 *Monopole charge quantization and the Aharonov-Bohm effect* Can. J. Phys. **62** 737–740.
- [110] Kuratsuji H 1988 *Geometric canonical phase factors and path integrals* Phys. Rev. Lett. **61** 1687–1690.
- [111] Kuratsuji H and Iida S 1985 *Effective action for adiabatic processes* Prog. Theo. Phys. **74** 439–445.
- [112] Lee H and l'Yi W S 1995 *Non-Hermitian techniques of canonical transformations in quantum mechanics* Phys. Rev. A **51** 982–988.
- [113] Lee H K, Nowak M A, Rho M and Zahed I 1993 *Nonabelian Berry phases in baryons* Ann. Phys. **227** 175–205.
- [114] Lee T D 1981 *Particle physics and introduction to field theory* (Harwood Academic Publishers).
- [115] Leinaas J M and Myrheim J 1977 *On the theory of elementary particles* Nuovo Cimento **37B** (January) 1–23.

- [116] Lévay P 1991 *Quaternionic gauge fields and the geometric phase* J. Math. Phys. **32** 2347–2357.
- [117] Lévay P 1992 *Non-Abelian Born-Oppenheimer electric gauge force and the natural metric on Hilbert subspaces* Phys. Rev. A **45** 1339–1346.
- [118] Li D 1993 *Anyons and quantum Hall effect on the sphere* Nucl. Phys. **B396** 411–428.
- [119] Littlejohn R G and Reinsch M 1997 *Gauge fields in the separation of rotations and internal motions in the  $n$ -body problem* Rev. Mod. Phys. **69** 213–275.
- [120] Liu Y M, Bates C A, Dunn J L and Polinger V Z 1996 *A scale transformation in  $T \otimes t_2$  Jahn-Teller systems* J. Phys. Condens. Matter **8** L523–L529.
- [121] Mantke W J 1996 *Time reversal of quantum experiments* Int. J. Th. Phys. **35** 1323–1339.
- [122] Mathur H 1991 *Thomas precession, spin-orbit interaction, and Berry's phase* Phys. Rev. Lett. **67** 3325–3327.
- [123] Mead C A 1979 *The noncrossing rule for electronic potential energy surfaces: The role of time-reversal invariance* J. Chem. Phys. **70** 2276–2283.
- [124] Mead C A 1980 *The molecular Aharonov-Bohm effect in bound states* Chem. Phys. **49** 23–32.
- [125] Mead C A 1987 *Molecular Kramers degeneracy and non-Abelian adiabatic phase factors* Phys. Rev. Lett. **59** 161–164.
- [126] Mead C A 1992 *The geometric phase in molecular systems* Rev. Mod. Phys. **64** 51–85.
- [127] Mead C A and Truhlar D G 1979 *On the determination of Born-Oppenheimer nuclear motion wavefunctions including complications due to conical intersections and identical nuclei* J. Chem. Phys. **70** 2284–2296.
- [128] Mendaš I 1997 *Pancharatnam phase for ordinary and squeezed states* Phys. Rev. A **55** 1514–1517.
- [129] Merzbacher E 1970 *Quantum Mechanics* (John Wiley and Sons, Inc, New York).
- [130] Messiah A 1966 *Quantum Mechanics* (John Wiley and Sons, Inc).
- [131] Minami M 1979 *Dirac's monopole and the Hopf map* Prog. Theor. Phys. **62** 1128–1142.

- [132] Moate C P, O'Brien M C M, Dunn J L, Bates C A, Liu Y M and Polinger V Z 1996  *$H \otimes h$ : A Jahn-Teller coupling that really does reduce the degeneracy of the ground state* Phys. Rev. Lett. **77** 4362–4365.
- [133] Moody J, Shapere A and Wilczek F 1986 *Realizations of magnetic-monopole gauge fields: Diatoms and spin precession* Phys. Rev. Lett. **56** 893–896.
- [134] Moore D J and Stedman G E 1990 *Effects of time-odd electron-phonon coupling in ligand field theory* J. Phys.: Condes. Matter **2** 2559–2577.
- [135] Nash C and Sen S 1983 *Topology and Geometry for Physicists* (Academic Press Ltd.).
- [136] Newman D J 1971 *Theory of lanthanide crystal fields* Adv. Phys. **20** 197–256.
- [137] Nikiforov A F and Uvarov V B 1988 *Special Functions of Mathematical Physics* (Birkhäuser).
- [138] Noether E 1918 *Invariate Variationsprobleme (Invariance in variation problems)* Nachr. kgl. Ges. Wiss. Göttingen Math. physik. Kl. 235–237. [§9.6].
- [139] O'Brien M C M 1964 *The dynamic Jahn-Teller effect in octahedrally co-ordinated  $d^9$  ions* Proc. Roy. Soc. **A281** 323–339.
- [140] O'Brien M C M and Pooler D R 1979 *Reduction factors for dynamic Jahn-Teller systems in strong coupling* J. Phys. C **12** 311–320.
- [141] Omnès R 1992 *Consistent interpretations of quantum mechanics* Rev. Mod. Phys. **64** 339–382.
- [142] Öpik U and Pryce M H L 1957 *Studies of the Jahn-Teller effect I: A survey of the static problem* Proc. Roy. Soc. **A238** 425–447.
- [143] Pancharatnam S 1956 *Generalized theory of interference and its applications* Proc. Indian Acad. Sci. **44** 247–262. (Reprinted in *Geometric Phases in Physics* edited by A Shapere and F Wilczek).
- [144] Payne S H and Stedman G E 1983 *II:  $E_o \times \sum(\alpha_1 + \epsilon)$  in octahedral symmetry* J. Phys. C **16** 2705–2723.
- [145] Payne S H and Stedman G E 1983 *III:  $\Gamma_5, \text{Gamma}_8 \otimes (\sum(\alpha_1 + \epsilon + \tau_2))$  in octahedral symmetry* J. Phys. C **16** 2725–2748.
- [146] Payne S H and Stedman G E 1983 *A quantum field theoretic approach to the calculation of reduction factors in Jahn-Teller systems. I: General theory* J. Phys. C **16** 2679–2703.

- [147] Perlin Yu E and Wagner M 1984 *The dynamical Jahn-Teller effect in localized systems* (North-Holland).
- [148] Polyakov A M 1974 *Particle spectrum in quantum field theory* JETP Lett. **20** 194–195.
- [149] Price P B, Shirk E K, Osborne W Z and Pinsky L S 1975 *Evidence for detection of a moving magnetic monopole* Phys. Rev. Lett. **35** 487–490.
- [150] Rajaraman R 1982 *Solitons and Instantons* (North-Holland, Amsterdam).
- [151] Ramond P 1990 *Field Theory: A Modern Primer* 2nd edn (Addison-Wesley).
- [152] Riley M J and Furlan A 1996 *The Jahn-Teller and Barnett effects* Chem. Phys. **210** 389–400.
- [153] Roe B P 1996 *Particle physics at the new millenium* (Springer-Verlag, New York).
- [154] Ryder L 1980 *Dirac monopoles and the Hopf map  $S^3 \rightarrow S^2$*  J. Phys. A **13** 437–447.
- [155] Sachs R G 1987 *The Physics of Time Reversal* (University of Chicago Press).
- [156] Saha M N 1936 *Origin of mass in protons and neutrons* Ind. J. Phys. **10** 141–153.
- [157] Sakurai J J 1985 *Modern Quantum Mechanics* (Benjamin/Cummings Publishing Company, Inc.).
- [158] Salmistraro F 1996 *Family of geometric phases* Phys. Rev. A **54** 2755–2758.
- [159] Samuel J 1996 *Merons and Thomas precession* Phys. Rev. Lett. **76** 717–721.
- [160] Samuel J and Bhandari R 1988 *General setting for Berry's phase* Phys. Rev. Lett. **60** 2339–2342.
- [161] Segert J 1987 *Non-Abelian Berry's phase, accidental degeneracy, and angular momentum* J. Math. Phys. **28** 2102–2114.
- [162] Segert J 1987 *Non-Abelian Berry's phase effects and optical pumping of atoms* Ann. Phys. **179** 294–312.
- [163] Serebrennikov Yu A and Steiner U E 1994 *Adiabatic rotation of effective spin. II. Spin-rotational relaxation* J. Chem. Phys. **100** 7508–7514.

- [164] Serebrennikov Yu A and Steiner U E 1995 *Nuclear spin-rotation interaction and non-Abelian gauge potentials* Mol. Phys. **84** 627–632.
- [165] Shankar R and Mathur H 1994 *Thomas precession, Berry potential, and the meron* Phys. Rev. Lett. **73** 1565–1569.
- [166] Shapere A and Wilczek F 1989 *Geometric Phases in Physics* (World Scientific).
- [167] Shapere A and Wilczek F 1989 *Geometric Phases in Physics* ch. 4 (World Scientific).
- [168] Shapere A and Wilczek F 1989 *Geometry of self-propulsion at low Reynolds number* J. Fluid Mech. **198** 557–585.
- [169] Simon B 1983 *Holonomy, the quantum adiabatic theorem, and Berry's phase* Phys. Rev. Lett. **51** 2167–2170.
- [170] Singer I M 1982 *Differential geometry, fiber bundles and physical theories* Physics Today **35** (March).
- [171] Spanier E H 1966 *Algebraic Topology* (McGraw-Hill).
- [172] Spevak V, Auerbach N and Flambaum V V 1997 *Enhanced  $T$ -odd,  $P$ -odd electromagnetic moments in reflection asymmetric nuclei* Phys. Rev. C **56** 1357–1369.
- [173] Spielman S, Fesler K, Eom C B, Geballe T H, Fejer M M and Kapitulnik A 1990 *Test for nonreciprocal circular birefringence in  $YBa_2Cu_3O_7$  thin films as evidence for broken time-reversal symmetry* Phys. Rev. Lett. **65** (July) 123–126.
- [174] Stedman G E 1983 *Visualising higher continuous symmetries in the Jahn-Teller effect* Eur. J. Phys. **4** 156–161.
- [175] Stedman G E 1992 *Modern nonlinear optics, Part 2* Prigogine and Rice. (Wiley, New York).
- [176] Stedman G E 1997 *Ring-laser tests of fundamental physics and geophysics* Rep. Prog. Phys. **60** 615–688.
- [177] Stedman G E, Johnsson M T, Li Z, Rowe C H and Bilger H R 1995  *$T$  violation and microhertz resolution in a ring laser* Opt. Lett. **20** 324–326.
- [178] Stedman G E and Newman D 1971 *Spin-lattice relaxation and elastic properties for rare-earth substituted  $LaCl_3$*  J. Chem. Phys. **55** 152–161.
- [179] Steenrod N 1970 *The Topology of Fibre Bundles* (Princeton University Press).

- [180] Stevens K W H 1967 *The theory of paramagnetic relaxation* Rep. Prog. Phys. **30** 189–226.
- [181] Streater R F and Wightman A S 1964 *PCT, Spin & Statistics and All That* (Benjamin).
- [182] Szopa M and Ceulemans A 1997 *The canonical form of the  $E \otimes \epsilon$  Jahn-Teller system* J. Phys. A **1295**–1302.
- [183] 't Hooft G 1974 *Magnetic monopoles in unified gauge theories* Nucl. Phys. **B79** 276–284.
- [184] 't Hooft G 1976 *Computation of the quantum effects due to a four-dimensional pseudoparticle* Phys. Rev. D **14** 3432–3450.
- [185] 't Hooft G 1976 *Symmetry breaking through Bell-Jackiw anomalies* Phys. Rev. Lett. **37** 8–11.
- [186] 't Hooft G 1989 *How instantons solve the  $U(1)$  problem* Physics Reports **142** 375–387.
- [187] Tomita A and Chiao R Y 86 *Observation of Berry's topological phase by use of an optical fiber* Phys. Rev. Lett. **57** 937–940.
- [188] Tompkin W R, Malcuit M S, Boyd R W and Chiao R Y 1990 *Time reversal of Berry's phase by optical phase conjugation* J. Opt. Soc. Amer. B **7** 230–233.
- [189] Tycko R 1987 *Adiabatic rotational splittings and Berry's phase in nuclear quadrupole resonance* Phys. Rev. Lett. **58** 2281–2284.
- [190] Šimánek E 1996 *Non-Abelian adiabatic phases in vortex dynamics. Magnus force in Fermi superfluids* Phys. Lett. A **221** 277–281.
- [191] van Vleck J H 1929 *On  $\sigma$ -doubling and electron spin in the spectra of diatomic molecules* Phys. Rev. **33** 467–506.
- [192] van Vleck J H 1937 *On the magnetic susceptibilities of Ti, V and Cr alums* Phys. Rev. **52** 246.
- [193] von Neumann J and Wigner E P 1929 *Behaviour of eigen values in adiabatic processes* Phys. Z. **30** 467–470.
- [194] Wagh A G and Rakhecha V C 1994 *Comment on "Spin evolution of a neutral particle in an oscillating magnetic field"* Phys. Lett. A **196** 139–140.
- [195] Wagh A G, Rakhecha V C, Summerhammer J, Badurek G, Weinfurter H, Allman B E, Kaiser H, Hamacher K, Jacobson D L and Werner S A 1997 *Experimental separation of geometric and dynamical phases using neutron interferometry* Phys. Rev. Lett. **78** 755–759.

- [196] Weinberg S 1975 *The  $U(1)$  problem* Phys. Rev. **D11** 3583.
- [197] Weinberg S 1995 *The Quantum Field Theory of Fields: Volume I* (University of Cambridge).
- [198] Wilczek F 1982 *Magnetic flux, angular momentum, and statistics* Phys. Rev. Lett. **48** 1144–1146.
- [199] Wilczek F 1982 *Quantum mechanics of fractional-spin particles* Phys. Rev. Lett. **49** 957–959.
- [200] Wilczek F 1990 *Fractional statistics and anyon superconductivity* (World Scientific, Singapore).
- [201] Wilczek F and Zee A 1984 *Appearance of gauge structure in simple dynamical systems* Phys. Rev. Lett. **24** 2111–2114.
- [202] Winstein B and Wolfenstein L 1993 *The search for direct CP violation* Rev. Mod. Phys. **65** 1113–1147.
- [203] Wu C S, Ambler E, Hayward R W, Hoppes D D and Hudson R P 1957 *Experimental test of parity conservation in beta decay* Phys. Rev. **105** 1413–1415.
- [204] Wu T T and Yang C N 1975 *Concept of nonintegrable phase factors and global formulation of gauge fields* Phys. Rev. D **12** 3845–3857.
- [205] Yang C N 1978 *Generalization of Dirac's monopole to  $SU(2)$  gauge fields* J. Math. Phys. **19** 320–328.
- [206] Zee A 1988 *Non-Abelian gauge structure in nuclear quadrupole resonance* Phys. Rev. A **38** 1–6.
- [207] Zeleny W B 1991 *Symmetry in electrodynamics: A classical approach to magnetic monopoles* Am. J. Phys. **59** 412–415.
- [208] Zhang J, Feng T, Zhang S and Jin G 1992 *Measurement of magnetic fields by a ring laser* Appl. Optics **31** 6459–6462.
- [209] Zhang Y D, Badurek G, Rauch H and Summhammer J 1994 *Reply to the comment of A. G. Wagh and B. C. Rakhecha on "Spin evaluation of a neutral particle in an oscillating magnetic field"* Phys. Lett. A **196** 141–142.
- [210] Zhang Y D, Badurek G, Rauch H and Summhammer J 1994 *Spin evolution of a neutral particle in an oscillating magnetic field* Phys. Lett. A **188** 225–231.
- [211] Zwanziger J W and Grant E R 1987 *Topological phase in molecular bound states: Application to the  $E \otimes e$  system* J. Chem. Phys. **87** 2954–2964.

- 
- [212] Zwanziger J W, Koenig M and Pines A 1990 *Non-Abelian effects in a quadrupole system rotating around two axes* Phys. Rev. A **42** 3107–3110.
- [213] Zygelman B 1987 *Appearance of gauge potentials in atomic collision physics* Phys. Lett. A **125** 476–481.