THE INFLUENCE OF THE GEOMETRIC PHASE ON REACTION DYNAMICS

STUART C. ALTHORPE
Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, UK

JUAN CARLOS JUANES-MARCOS
Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

ECKART WREDE
Department of Chemistry, University of Durham, South Road, Durham, DH1 3LE, UK

CONTENTS

I. Introduction
II. Unwinding the Nuclear Wave Function
   A. Topology and Encirclement
   B. Symmetry Approach
   C. Feynman Path Integral Approach
   D. Consistency between the Symmetry and Feynman Approaches
III. Application to the Hydrogen-Exchange Reaction
   A. Reaction Paths and Potential Energy Surface
   B. Application of Topology to Reactive Scattering
   C. Details of the Calculation: The Use of Vector Potentials
   D. Solving the Cancellation Puzzle
IV. Further Aspects of Topology
   A. Including Particle-Exchange Symmetry
   B. Complete Unwinding of the Nuclear Wave Function
   C. Difference between Bound and Scattering Systems
V. Outlook and Conclusions
Acknowledgments
References

Advances in Chemical Physics, Volume 138, edited by Stuart A. Rice
Copyright © 2008 John Wiley & Sons, Inc.
I. INTRODUCTION

Research over the past 15 years has established that conical intersections (CIs) are much more common than previously thought, and that they play a central role in photochemistry [1, 2]. The main role of a CI is to act as a funnel, transferring population between the upper and lower adiabatic electronic states. This effect has been studied extensively, and is well known to be caused by the derivative coupling terms between the two electronic surfaces [3], which are singular at the CI. However, an accompanying effect, which has received less attention, is the so-called geometric (or Berry) phase (GP), in which the adiabatic electronic wave function changes sign upon following a closed loop around the CI [4–9]. This effect is particularly interesting when the system is confined to the lower adiabatic surface, since it is then the only nonadiabatic effect produced by the CI.

The GP affects the nuclear dynamics by introducing a corresponding sign change in the continuity boundary condition, which cancels out the sign change in the electronic wave function (in order to keep the total wave function single valued). In a model particle-on-a-ring system, with the CI at the center of the ring, the GP boundary condition changes the allowed values of \( M \) in the nuclear wave function, \( \exp(iM\phi) \), from integer to half-integer values. The GP brings about analogous changes in the quantum numbers and energy levels of more realistic systems, and some of these have been predicted and observed in a variety of Jahn–Teller molecules [10–13]. One can say with some confidence, therefore, that the effect of the GP on the nuclear wave function of a bound-state system is well understood.

Until very recently, however, the same could not be said for reactive systems, which we define to be systems in which the nuclear wave function satisfies scattering boundary conditions. It was understood that, as in a bound system, the nuclear wave function of a reactive system must encircle the CI if nontrivial GP effects are to appear in any observables [6]. Mead showed how to predict such effects in the special case that the encirclement is produced by the requirements of particle-exchange symmetry [14]. However, little was known about the effect of the GP when the encirclement is produced by reaction paths that loop around the CI.

Very recently, this state of affairs has changed, and there is now a good general understanding of GP effects in gas-phase reactions. This has come about mainly through detailed reactive-scattering studies, both theoretical [15–37] and experimental [30–39], on the prototype hydrogen-exchange reaction (H + H\(_2\) → H\(_2\) + H). This is the simplest reaction to possess a CI, and at high energies (> 1.8eV above the potential minimum) it is just possible for some reaction paths to encircle the CI. Reactive-scattering calculations are difficult, and the first calculations that included the GP boundary condition were...
reported by Kuppermann and co-workers in the early 1990s [15–19]. These calculations predicted large GP effects at high energies, but unfortunately these predictions were not reproduced by later calculations [20–29], nor, most crucially, by experiment [30–39]. Instead, the experiments found no evidence at all of GP effects in the H + H2 reaction. Detailed scattering data from the experiments agreed quantitatively with theoretical predictions that omitted the GP boundary condition.

This negative result seemed to imply that H + H2 could not be used to investigate the effect of the GP because the nuclear wave function does not encircle the CI. However, a series of calculations by Kendrick [20–22] yielded a surprising result, which we will refer to in this chapter as the “cancellation puzzle”. This is that GP effects appear in the scattering observables at specific values of the total angular momentum quantum number \( J \), but cancel on summing over \( J \) to give the full reactive scattering wave function (describing a rectilinear collision of the reagents). This is puzzling because the sum over \( J \) is a unitary transformation in the external, angular, degrees of freedom, which describe the scattering of the products, and there is no direct relation linking this space to the GP boundary condition (which acts on the internal degrees of freedom describing motion around the CI).

This chapter surveys the work we did [25–29] to solve the cancellation puzzle, and the general explanation of GP effects in reactive systems that came out of it [27,28]. The latter uses ideas that were introduced in the late 1960s in Feynman path-integral [40] work on the analogous Aharonov–Bohm effect [41–45]. These early papers seem to have passed unnoticed in the chemical physics community, perhaps because they are written in the language of path-integral theory and algebraic topology. However, the central result of this work is surprisingly simple, and can be derived without path-integrals [27,28]. It is that the nuclear wave function has two components, and that the sole effect of the GP is to change their relative sign. One component contains all the Feynman paths that loop an even number of times around the CI; the other contains all the paths that loop an odd number of times.

In Section II, we introduce this central result, deriving it first without path integrals, by using a diagrammatic representation of the nuclear wave function; we then give a heuristic summary of the ideas behind the early Aharonov–Bohm papers, and show how they can be combined with the diagrammatic approach. In Section III, we describe in detail the solution to the cancellation puzzle in H + H2, which demonstrates how to explain GP effects in a reaction in terms of the even- and odd-looping Feynman paths. In Section IV, we discuss some further aspects of the topology, explaining the effect of particle-exchange symmetry, and the difference between GP effects in bound and reactive systems. Section V concludes the chapter.
II. UNWINDING THE NUCLEAR WAVE FUNCTION

A. Topology and Encirclement

We will consider a system with $N$ nuclear degrees of freedom, which possesses one CI seam [1,2] of dimension $N - 2$. The topology of the nuclear space can then be represented schematically as shown in Fig. 1 [28]. The line at the center represents every CI point in the seam. Each circular cut through the cylinder represents the two degrees of freedom in the nuclear “branching space”, in which the adiabatic potential energy surfaces have the familiar double-cone shape, centred about the CI point. We assume that the seam line extends throughout the entire region of energetically accessible nuclear coordinates space. We also assume that the system is confined to the lower (adiabatic) electronic state, because it has insufficient energy to approach the region of strong coupling with the upper state close to the conical intersection. The CI seam line is therefore surrounded by a tube of inaccessible coordinate space.

We then define an internal coordinate $\phi$ such that $\phi = 0 \rightarrow 2\pi$ denotes a path that has described one complete loop around the CI in the nuclear branching space. Other than this, we need specify no further details about $\phi$. We do not even need to specify whether the complete set of nuclear coordinates give a direct product representation of the space. It is sufficient that $\phi$ permits us to count how many times a closed loop has wound around the CI. Using this definition of $\phi$, we can express the effect of the GP on the

\[
\phi_{N - 2}^c(b) - \phi_{N - 2}^a(a)
\]

Figure 1. (a) Diagram illustrating the topology of the $N$-dimensional nuclear coordinate space of a reactive system with a CI. The vertical line represents the $(N - 2)$-dimensional space occupied by the CI seam. The gray disk represents a two-dimensional (2D) branching-space cut through one point on the seam, with the angle $\phi$ describing internal rotation around the CI. (b) A nuclear wave function that wraps around the CI, but is not a torus, and thus exhibits only trivial GP effects. (c) A torus-shaped nuclear wave function encircling the CI. The “arms” are to be understood as extending to infinity, and are the portions of the wave function in the reagent and product channels.
adiabatic ground-state electronic wave function \( \Phi(\phi) \) and the nuclear wave function \( \Psi(\phi) \) as

\[
\begin{align*}
\Phi(\phi + 2n\pi) &= (-1)^n\Phi(\phi) \\
\Psi(\phi + 2n\pi) &= (-1)^n\Psi(\phi)
\end{align*}
\]  

The dependence on the other \( N-1 \) nuclear degrees of freedom has been suppressed.

The effects of the GP are therefore the differences between the nuclear dynamics described by the wave function

\[
\Psi_G(\phi) = (-1)^n\Psi_G(\phi + 2n\pi)
\]

which correctly includes the GP boundary condition, and the wave function

\[
\Psi_N(\phi) = \Psi_N(\phi + 2n\pi)
\]

which ignores it (and is therefore physically incorrect). It is well known in the literature that the GP will only produce a nontrivial effect on the dynamics when \( \Psi_G(\phi) \) encircles the CI. Otherwise the effect is simply a change in the phase of \( \Psi_N(\phi) \), which has no effect on any observables. Hence, throughout this chapter, we are seeking to explain how the dynamics described by \( \Psi_G(\phi) \) differs from the dynamics described by \( \Psi_N(\phi) \), when these wave functions encircle the CI.

It is worth clarifying what is meant by encirclement. As already mentioned, the nuclear coordinates need not form a direct product, and in fact the notion of taking a cut through the nuclear coordinate space, in order to see whether \( \Psi_G(\phi) \) encircles the CI in this cut, is not useful. Figure 1(b) shows a nuclear wave function which, if a certain choice of nuclear coordinates were used, could easily be made to “encircle” the CI if a suitable 2D cut were taken. However, this particular wave function would not show nontrivial GP effects, because it does not encircle the CI: it has unconnected “ends”. For nontrivial GP effects to appear, \( |\Psi_G(\phi)|^2 \) must have the form of a torus in the nuclear coordinate space, as shown in Fig. 1(c). If one were to take a series of branching-space cuts through this wave function, none of them would encircle the CI, and hence one might get the mistaken impression that this wave function would only show a trivial phase change upon inclusion of the GP boundary condition. However, the wave function of Fig. 1(c) would definitely show strong, nontrivial GP effects. There are various ways in which one can prove this and we will mention one below. It is important to emphasise that it is \( |\Psi_G(\phi)|^2 \), which has the form of a torus and not the wave function \( \Psi_G(\phi) \).
B. Symmetry Approach

To explain the effect of the GP on the nuclear dynamics [i.e., to explain the difference between the dynamics described by an encircling $\Psi_G(\phi)$ and an encircling $\Psi_N(\phi)$], we need to compare the topology of $\Psi_G(\phi)$ with the topology of $\Psi_N(\phi)$. In Section II. C, we review how this can be done using the homotopy of the Feynman paths [41–45] that make up these wave functions. But first, to demonstrate the simplicity of the problem, we use the diagrammatic approach developed in Refs. [27 and 28].

We represent the internal coordinate space occupied by the nuclear wave function as shown in Fig. 2. The gray area represents the energetically accessible region of the potential energy surface; the conical intersection is the point at the center; the arms represent the reagent entrance and product exit channels. To simplify the discussion, we place a restriction on $\phi$ (which will be relaxed later), stating that $\phi$ tends to a constant value as the system moves down the entrance or exit channel toward an asymptotic separation of the reagents or products. This places no restriction on the generality of the diagram, other than that the conical intersection should be located in the “strong-interaction region” of the potential energy surface, where all the nuclei are close together. Note that, although we have restricted the number of product channels to one, the diagram is immediately generalizable to systems with multiple product channels. We also assume that the reaction is bimolecular (leaving unimolecular reactions until Section II.D), which means that it is initiated at the asymptotic limit of the reagent channel, at the value of $\phi$ that is reached in this limit. We will define this to be $\phi = 0$.

![Figure 2](image_url)

**Figure 2.** (a) Schematic picture of the potential surface of a reactive system, indicating that there is an energetically accessible (gray) “tube” through the potential surface, permitting encirclement of the CI (dot at center). The “arms” are the reagent and product channels. (b) The same surface, represented in the $0 \rightarrow 4\pi$ cover space. The rectangle represents a $0 \rightarrow 2\pi$ sector that can be cut out of the double space so as to map back onto the single space, where the $\phi = 0$ and $\phi = 2\pi$ “edges” are joined together at the cut line (chains).
Figure 2(b) represents the potential surface of the identical system, mapped onto the double-cover space [28]. The latter is obtained simply by “unwinding” the encirclement angle $\phi$, from $0 \rightarrow 2\pi$ to $0 \rightarrow 4\pi$, such that two (internal) rotations around the CI are represented as one in the page. The potential is therefore symmetric under the operation $R_{2\pi}$ defined as an internal rotation by $2\pi$ in the double space. To map back onto the single space, one cuts out a $2\pi$-wide sector from the double space. This is taken to be the $0 \rightarrow 2\pi$ sector in Fig. 2(b), but any $2\pi$-wide sector would be acceptable. Which particular sector has been taken is represented by a cut line in the single space, so in Fig. 2(b) the cut line passes between $\phi = 0$ and $2\pi$. Since the single space is the physical space, any observable obtained from the total (electronic + nuclear) wave function in this space must be independent of the position of the cut line.

To construct a diagrammatic representation of the wave function, we start in the double space, as shown in Fig. 3a. The arrow at the top indicates that the incoming boundary condition is applied here, and the arrows at each of the other channels indicate outgoing boundary conditions. Note that we are treating the second appearance of the reagent channel (at $\phi = 2\pi$) as though it were a product channel, and are treating the second appearance of the product channel (in the $2\pi \rightarrow 4\pi$ sector) as though it were physically distinct from the first appearance of this channel, which is indicated by the use of wavy lines. Consequently, the wave function $\Psi_e$ is neither symmetric nor antisymmetric under $\phi \rightarrow \phi + 2\pi$, which means it cannot be mapped back onto the physical space independently of the position of the cut line. In other words $\Psi_e$ is the wave function of a completely artificial system.

To construct wave functions that can be mapped back onto the physical space, one needs to take symmetric and antisymmetric linear combinations of $\Psi_e(\phi)$ and $\Psi_o(\phi) = \Psi_e(\phi + 2\pi)$, and these are illustrated in Fig. 3b. It is then

![Figure 3](image_url)

**Figure 3.** (a) The unsymmetrised nuclear wave function $\Psi_e$ (solid line) in the double space. The arrows indicate the application of incoming and outgoing scattering boundary conditions, (b) The symmetrized linear combinations of $\Psi_e$ (solid) and $\Psi_o$ (dashed), which yield $\Psi_{N/G} = 1/\sqrt{2}[\Psi_e \pm \Psi_o]$, (c) The same functions mapped back onto the single space.
clear that these functions can be mapped onto the physical space (Fig. 3c), and that they correspond to $\Psi_N$ and $\Psi_G$, respectively. Thus we may write,

$$\Psi_G = \frac{1}{\sqrt{2}} [\Psi_e + \Psi_o]$$

$$\Psi_N = \frac{1}{\sqrt{2}} [\Psi_e - \Psi_o]$$  \hspace{1cm} (5)

This equation is the main result needed to explain the effect of the GP on the nuclear dynamics of a chemical reaction. Clearly, the sole effect of the GP is to change the relative sign of $\Psi_e$ and $\Psi_o$. Within each of these functions the dynamics is completely unaffected by the GP. We emphasize that, despite remaining unnoticed for so long in the chemical physics community, Eq. (5) is exact.

If we can compute $\Psi_G$ and $\Psi_N$ numerically (as described below), it is therefore trivial to extract $\Psi_e$ and $\Psi_o$ by evaluating

$$\Psi_e = \frac{1}{\sqrt{2}} [\Psi_N + \Psi_G]$$

$$\Psi_o = \frac{1}{\sqrt{2}} [\Psi_N - \Psi_G]$$  \hspace{1cm} (6)

Once one has extracted $\Psi_e$ and $\Psi_o$, an explanation of the GP effect on the nuclear dynamics will follow immediately. The dynamics in $\Psi_e$ is decoupled from the dynamics in $\Psi_o$, and thus any observable will show GP effects only if the corresponding operator samples $\Psi_e$ and $\Psi_o$ in a region of space where these functions overlap. In a nonencircling nuclear wave function, $\Psi_e$ and $\Psi_o$ never overlap, and this gives us a diagrammatic proof (Fig. 4) of the well-known result that a nonencircling wave functions shows no nontrivial GP effects.

**Figure 4.** (a) Single- and (b) double-space representations of $\Psi_e$ (solid) and $\Psi_o$ (dashed) for a system that does not encircle the CI.
C. Feynman Path Integral Approach

We now explain the physical significance of the two components $\Psi_c$ and $\Psi_o$ in terms of the path integral theory developed in Refs. [41–45]. This theory was developed originally to treat the Aharonov–Bohm system, in which an electron encircles, but does not touch, a magnetic solenoid. The vector potential of the solenoid has an effect that is exactly equivalent to the application of the GP boundary condition, and scattering boundary conditions are applied at long range. The Aharonov–Bohm system is therefore exactly analogous to a nuclear wave function in a reactive system that encircles a CI.

To avoid discouraging the reader, we point out that only a few, basic concepts of path integrals are required. We review these here in a heuristic manner, beginning with the celebrated result of Feynman and Hibbs [40], which is that the time-evolution operator or Kernel, $K = \exp(-i\hat{H}t/\hbar)$, can be constructed using

$$K(x, x_0|t) = \int \mathcal{D}x(t)e^{iS(x, x_0)/\hbar}$$

(7)

Here, $\mathcal{D}x(t)$ represents the sum over all possible paths connecting the points $x$ and $x_0$ in the time interval $t$, and $S$ is the classical action evaluated along each of these individual paths. It is useful to point out two properties of this expression: (1) the overall sign of the Kernel is arbitrary, because $S$ is only defined up to an overall constant (because $S$ is the time integral over the Lagrangian, and the latter is only defined up to a total derivative in $t$ [46]); (2) each path has equal weight, so the relative contribution of a given path to the sum is determined by the extent to which it is canceled out by its immediate neighbors.

Any prediction expressed in the language of path integrals must have an equivalent formulation in the language of wave functions. Point (1) is equivalent to saying that a wave function is only specified up to an overall phase factor. Point (2) can be thought of as saying that, when computing $K(x, x_0|t)$, all possible paths between $x$ and $x_0$ in time $t$ are coupled. If we start with one particular path between $x$ and $x_0$, then we need to know all of its immediate neighbors, in order to assess the extent to which this path is canceled out by them. These neighboring paths are obtained by all possible tiny distortions that can be applied to the first path. We then need to know all of the immediate neighbors of each of the latter paths (in order to assess the extent to which each of these is canceled out), and then we need to find out the immediate neighbors of the new paths, and so on. In other words, if we start with one particular path between $x$ and $x_0$, then this path is coupled (in the sense just described) to all the other paths into which it can be continuously deformed. This is equivalent to saying that one cannot accurately compute just part of a wave function; one must compute all of it, since all parts of the function are coupled by the Hamiltonian operator.
In their work on the Aharonov–Bohm system, Schulman, deWitt, and co-workers [41–44] found that points (1) and (2) must be modified when applying path integral theory in a multiply connected space. The term multiply connected simply means that the space contains an inaccessible region or obstacle, which gets in the way, such that a given path between \( x \) and \( x_0 \) cannot be continuously deformed into all other possible paths. It can only be deformed into a subset of such paths. This subset defines a homotopy class: paths that belong to different homotopy classes are called different homotopes. The concept that there exist different classes of paths, such that a path that belongs to one class cannot be continuously deformed into a path that belongs to another, is called homotopy.

The nuclear coordinate space shown in Fig. 1 is a multiply connected space, because there is an energetically inaccessible “tube” of space surrounding the CI seam. An explanation of the homotopy of such a space will be found in any elementary text on topology [47]. Let us take first a system with only two nuclear degrees of freedom, so that the CI is just a point at the center of the branching space, and there are no other degrees of freedom. The homotopy of a given path within this space is simply the number of entire loops it follows around the CI. We can thus classify each homotopic class according to a winding number \( n \), as defined in Fig. 5. Note that the sign of \( n \) indicates the sense of the path, and that it is useful to adopt the convention that even \( n \) refer to paths that make an even number of clockwise loops or an odd number of counterclockwise loops; and odd \( n \) vice versa. It is easy to prove that the set of all these homotopic classes forms an infinite group, which is called the “Fundamental Group” of a circle [47].

The same classification into winding numbers can be used in a system with \( N \) nuclear degrees of freedom, in which the CI seam is an \( (N - 2) \)-dimensional hyperline as in Fig. 1. For example, if we take \( N = 3 \), then the seam is a line; the

\[ \begin{align*}
\text{Figure 5.} & \quad \text{Examples of Feynman paths belonging to different homotopy classes, illustrating how the winding number } n \text{ is defined.}
\end{align*} \]
homotopy of this system is just the same as for the \( N = 2 \) system, since the number of loops made around the line can be represented by a winding number defined exactly as for the 2D case. Although it is difficult to visualize, the generalization continues to all higher \( N \) in the same way, so that one can always classify a path by its winding number around the \( (N - 2) \)-dimensional CI hyperline. In many systems, each class of paths designated by the winding number \( n \) will in fact include more than one homotopy class, because it will be possible to further classify the paths according to their winding about other energetically inaccessible regions in the potential surface, which may exist in addition to the tube around the CI seam. However, to understand the GP we do not need to consider these classes, and so, for shorthand, we will use the terms “homotopy class” and “winding number” interchangeably.

Retracing the argument used to justify point (2), it is clear that, in a multiply connected space, a given path is only coupled to those paths into which it can be continuously deformed. By definition, these are all the paths that belong to the same homotopy class. Paths belonging to different homotopy classes are thus decoupled from one another [41–45]. For a reactive system with a CI that has the space of Fig. 1, this means that a path with a given winding number \( n \) is coupled to all paths with the same \( n \), but is decoupled from paths with different \( n \). As a result, the Kernel separates into [41–45]

\[
K(x, x_0 | t) = \sum_{n=-\infty}^{\infty} e^{in\alpha} K_n(x, x_0 | t)
\] (8)

where

\[
K_n(x, x_0 | t) = \int \mathcal{D}_n x(t) e^{iS(x, x_0)/\hbar}
\] (9)

and \( \mathcal{D}_n x \) denotes the sum over all paths linking \( x_0 \) to \( x \) that have winding number \( n \).

Each \( K_n \) in Eq. (8) has a different overall phase, which arises because a different Lagrangian can be used for each value of \( n \). However, there is a strong constraint on these phases, which arises because the set of Kernels \( K_n \) must form an irreducible representation of the Fundamental Group of the circle. As a result, the phases have the form \( e^{in\alpha} \) [given in Eq. (8)], so that there is only one parameter \( \alpha \) that can be varied. To determine possible values of \( \alpha \), let us consider the operation \( \phi \to \phi + 2\pi \) on \( K_n \). This operation is equivalent to rotating the end points of all the paths around the CI by \( 2\pi \), thus increasing the winding number of each path from \( n \) to \( n + 1 \). As a result, \( K_n \to K_{n+1} \), which means that, overall, \( K \to \exp(i\alpha)K \). In other words, specifying \( \alpha \) is equivalent to specifying the \( \phi \to \phi + 2\pi \) boundary condition that is to be satisfied by the
Kernel. Thus, we can obtain Kernels corresponding to GP and non-GP boundary conditions by choosing $\alpha = \pi$ and $\alpha = 0$, which gives

$$
K_G(x, x_0|t) = K_e(x, x_0|t) - K_o(x, x_0|t)
$$
$$
K_N(x, x_0|t) = K_e(x, x_0|t) + K_o(x, x_0|t)
$$

(10)

where $K_e = \sum K_n$, with the sum running over all even $n$, and $K_o$ is similarly defined for odd $n$.

To put this result in context, one should imagine a crude semi classical calculation, in which one propagates Newtonian trajectories, each of which is given a phase $\exp(iS/\hbar)$. One could implement the GP boundary condition by counting the number of loops $n$ made by each trajectory around the CI, and adding an extra $n\pi$ to the associated phase. To our knowledge, no such calculation has been reported, almost certainly because it would be difficult to disentangle genuine GP effects from errors in the approximation. However, Eq. (10) tells us that such an intuitive approach can be applied to the Feynman paths, and thus implemented rigorously, without approximation.

D. Consistency between the Symmetry and Feynman Approaches

The separation of the Feynman paths in Eq. (10) is equivalent to the splitting of the wave function into $\Psi_e$ and $\Psi_o$ in Eq. (6). To demonstrate this, we connect the Kernel to the wave function using [48],

$$
\Psi(x) = \frac{1}{A(E)} \int dx_0 \int_0^\infty dt e^{iE_t/\hbar} K(x, x_0|t) \chi(x_0)
$$

(11)

where $\chi(x_0)$ is an initial wave packet, which contains a spread of energies $A(E)$. At time $t = 0$, $\chi(x_0)$ is localized in the reagent channel, at a sufficiently large reagent separation that the interaction potential can be neglected. The function $\Psi(x)$ given by Eq. (11) is the time-independent wave function, with incoming boundary conditions in the reagent channel, as represented schematically in Fig. 3. It follows immediately from Eq. (11) that $K_e$ generates $\Psi_e(\phi)$, and $K_o$ generates $\Psi_o(\phi)$. Hence, unwinding the nuclear wave function according to Eq. (6) is equivalent to separating the even $n$ Feynman paths, which are contained in $\Psi_e(\phi)$, from the odd $n$ paths, which are contained in $\Psi_o(\phi)$.

Some care must be taken when applying the Feynman interpretation to Eq. (6), as the Feynman interpretation must be consistent with the position of the cut line (used to map from the double to the single space). For example, Fig. 6a and b shows two different choices of cut line. It is clear that the relative sign of $\Psi_e(\phi)$ and $\Psi_o(\phi)$, and hence all the GP effects, are independent of the position of the cut line. However the overall phase of $\Psi_G(\phi)$ does depend on the cut line. This phase is important, because it must cancel out a corresponding
phase in the electronic wave function $\Phi(\phi)$, to give a total wave function $\Psi_G(\phi)\Phi(\phi)$, which is independent of the position of the cut line. Because of this, one needs to define the winding number $n$ with respect to the cut line and not with respect to the $(\phi = 0)$ point at which the nuclear wave function enters the encirclement region. Thus in Fig. 6b, a path that starts at $\phi = 0$ and terminates at $\phi = \pi/2$, and has made no loops around the CI, is classified as an $n = 0$ path. However, a path that starts at $\phi = 0$ and terminates just short of the cut line at, say, $\phi = \pi/6$, and has also made no loops around the CI, is an $n = -1$ path. A path that enters at $\phi = 0$ and makes one clockwise loop around the CI will be an $n = 0$ path if it terminates just short of the cut line, and will only become an $n = 1$ path once it has passed the cut line, and so on. By classifying the Feynman paths with respect to the cut line in this way we ensure that the overall phase of $\Psi_G(\phi)$ has the correct dependence on $\phi$ needed to cancel the corresponding dependence of the phase of $\Phi(\phi)$.

In Fig. 3, we placed the cut line between $\phi = -\epsilon$ and $\phi = 2\pi - \epsilon$, where $\epsilon$ is an arbitrarily small number. This choice will often be the most convenient cut line, because $n$ then exactly describes the number of complete loops that the system has made around the CI since entering the encirclement region. Thus the paths that scatter inelastically will each have described an (internal) rotation of exactly $\phi = 2n\pi$.

However, it will not always be possible to fix the cut line at the same value of $\phi$ as the entry points, for the reason that the system does not enter the encirclement region at one unique value of $\phi$. Up till now, we have assumed (see Section II.B) that the reaction is bimolecular, that it can only encircle the CI when the nuclei are all close together, and that the reagents and products are
distinguishable. These conditions are what are required to guarantee that the system starts at one unique value of \( \phi \), which we have taken to be \( \phi = 0 \). We can now relax these conditions, and consider unimolecular reactions, and reactions that can encircle the CI at large separations of the reagents or products. We consider the case of bimolecular scattering with identical reagents and products in Section IV. A.

We can represent a unimolecular reaction using the diagram of Fig. 6c. The gray blob indicates the initial state of the system. For example, it could be the Frank–Condon region accessed in a photodissociation experiment [49]. All the Feynman paths that contribute to the nuclear wave function will originate in the initial state. Hence, the paths will have a spread of start points, distributed over the range of \( \phi \) for which the initial state is nonnegligible. Clearly, the symmetry argument of Section II.B applies immediately to this system, so we may unwind the wave function, and extract \( \Psi_e \) and \( \Psi_o \) using Eq. (6). We can then interpret these functions as containing the even \( n \) and odd \( n \) Feynman paths, respectively, where \( n \) is defined with respect to a fixed cut line. Note that, when we discuss, say, the even \( n \) Feynman paths, we are not referring to paths that all necessarily complete an even number of loops around the CI, since the paths may have started on different sides of the cut line (if the latter passes through the initial state), or on different sides of the end point. The reader may verify that, in either of these cases, the even \( n \) paths will contain a mixture of paths that have looped an even and an odd number of times around the CI.

Hence, when applied to a unimolecular reaction, Eq. (6) does not give such a neat separation into even- and odd-looping Feynman paths. However, the separation that it does give (into even and odd \( n \), each of which contains a mixture of even- and odd-looping paths) is the one that is necessary to explain the effect of the GP, since these are the two contributions to \( \Psi \) whose relative sign is changed by the GP. Clearly, if we were to compute directly the Kernels, we could then separate out the odd- and even-looping paths, because we would know the starting point of each path. In the wave function, however, we neither know the starting points of the individual paths, nor do we need to in order to explain the effect of the GP.

Similar arguments to those just given apply to bimolecular reactions in which the CI can be encircled when the reagents are still well separated from one another. For such systems, one cannot define \( \phi \) such that it tends to a unique value as the system travels out along the reagent channel. The incoming boundary condition must then be applied across a range of \( \phi \), which is analogous to the range of \( \phi \) contained in the initial state of the unimolecular reaction. Applying Eq. (6) will then separate out the even and odd \( n \) paths with respect to a fixed cut line, which paths may contain a mixture of odd- and even-looping paths (as in the unimolecular case).
III. APPLICATION TO THE HYDROGEN-EXCHANGE REACTION

The above theory is completely general, but was developed in response to a specific challenge, which was to explain the “cancellation puzzle” posed by the results of Kendrick’s calculations [20–22] on the hydrogen-exchange reaction. This application of the theory gives a very good illustration of the effect of the GP on a gas-phase reaction, and we discuss it here in some detail. Note that the only familiarity with reactive scattering assumed is a basic knowledge of quantum scattering theory (e.g., given in standard introductory texts on quantum mechanics).

A. Reaction Paths and Potential Energy Surface

The first thing to be done when applying the theory is to identify the e and o reaction paths. One can then proceed to calculate $\Psi_G$ and $\Psi_N$, and then to extract $\Psi_e$ and $\Psi_o$ using Eq. (6). In $H_2 + H_2$, the form of the potential energy surface is very well characterized [50–53], and the form of the CI is a standard example of an $E \times e$ Jahn–Teller intersection.

Figure 7 shows a schematic representation of the $H_2 + H_2$ potential energy surface [29], plotted using the hyperspherical coordinate scheme of Kuppermann [54]. In this section, we will treat the three hydrogen nuclei as distinguishable particles, ignoring the requirement that the nuclear wave function be antisymmetric under exchange of two $^1H$ nuclei. The exchange symmetry is

![Figure 7](image_url)

**Figure 7.** Schematic representation of the 1-TS (solid) and 2-TS (dashed) (where TS = transition state) reaction paths in the reaction $H_A + H_B H_C \rightarrow H_A H_C + H_B$. The $H_3$ potential energy surface is represented using the hyperspherical coordinate system of Kuppermann [54], in which the equilateral-triangle geometry of the CI is in the center ($\times$), and the linear transition states ($\dagger$) are on the perimeter of the circle; the hyperradius $\rho = 3.9$ a.u. The angle $\phi$ is the internal angular coordinate that describes motion around the CI.
easy to incorporate by taking appropriate linear combinations of the unsymmetrised (distinguishable particle) wave functions, and we explain how to do this in Section IV.A. Hence, we consider here only the effects of the GP on the unsymmetrised wave functions, since these are the effects caused by reaction paths that encircle the CI, which give rise to the cancellation puzzle.

Hence, we consider wave functions in which the reaction starts at an asymptotic separation of one uniquely specified arrangement of the atoms (A + BC), and analyze the cross-sections produced by reactive scattering into one of the product channels (AC + B). It is well known [55,56] that the dominant H + H₂ reaction path passes over one transition state (1-TS), as illustrated schematically in Fig. 7. Since GP effects are found in the reaction probabilities at sufficiently high energies [20–22, 25–27] (>1.8 eV above the potential minimum), the wave function must encircle the CI at these energies, and thus also contain reaction paths that pass over two transition states (2-TS).

For the AC + B products, the 1-TS paths make less than one full revolution, loop in a clockwise sense around the CI (see Fig. 7), and are assigned a winding number \( n = 0 \) (following the convention of Section II.C). The 2-TS paths also make less than one full revolution, but loop in an anticlockwise sense, and are assigned \( n = -1 \) [28]. This means that the 1-TS (Feynman) paths are contained in \( \Psi_e \) and the 2-TS paths in \( \Psi_o \). In principle, there are also paths with higher winding numbers present in both \( \Psi_e \) and \( \Psi_o \), but these can be ignored (since reaction paths passing over three or more TS are highly unlikely in H + H₂). In this chapter, we will therefore use the terms \( \Psi_e \) paths and 1-TS paths (and \( \Psi_o \) paths and 2-TS paths) interchangeably.

### B. Application of Topology to Reactive Scattering

Here we summarize what the reader will need to know about reactive scattering in order to understand the application of the topological ideas of Section II to the hydrogen-exchange reaction. The only thing we assume is that the reader is familiar with the concept of scattering wave functions and boundary conditions. Generalizing these concepts to reactive scattering is conceptually straightforward (although technically difficult [57–62], but we will not need to discuss the technicalities here).

We consider a nuclear wave function describing collisions of type \( A + BC(n) \rightarrow AC(n') + B \), where \( n = \{v, j, k\} \) are the vibrational \( v \) and rotational \( j \) quantum numbers of the reagents (with \( k \) the projection of \( j \) on the reagent velocity vector of the reagents), and \( n' = \{v', j', k'\} \) are similarly defined for the products. The wave function is expanded in the terms of the total angular momentum eigenfunctions \( D_{j k'}^{l'}(\theta, \eta, \chi) \) [63], and takes the form [57–61]

\[
\Psi_{n}^{[\lambda]}(R, r, \gamma; \theta, \eta, \chi) = \frac{1}{2k_{n}R} \sum_{j k'} (2J + 1) D_{j k'}^{l'}(\theta, \eta, \chi) F_{n k'}^{[\lambda]}(R, r, \gamma) \tag{12}
\]
where $\hbar k_n$ is the magnitude of the A–BC approach momentum. The label $\lambda$ will be taken to be each of $G, N, e, \text{ or o (see below), which indicate whether the wave function is } \Psi_G, \Psi_N, \Psi_e, \text{ or } \Psi_o$. The internal degrees of freedom (in which the GP boundary condition is applied in $\Psi_G$) are described by the coordinates $(R, r, \gamma)$, which can be defined in either the reagent or product arrangements; in the product arrangement, $r$ is the AC bond length, $R$ is the length of the vector joining B to the AC center of mass, and $\gamma$ is the angle between this vector and the AC bond. The external, spatial degrees of freedom are described by the Euler angles $(\theta, \eta, \chi)$.

The most important observable is the angular distribution of the scattered products with respect to the initial approach direction of the reagents, which is called the state-to-state differential cross-section (DCS). The DCS can be written [57–61]

\[
\frac{d\sigma^{[\lambda]}_{n'\rightarrow n}(\theta, E)}{d\Omega} = \frac{1}{2j+1} |f^{[\lambda]}_{n'\rightarrow n}(\theta, E)|^2
\]  

where $f^{[\lambda]}_{n'\rightarrow n}(\theta, E)$ is the scattering amplitude, obtained by taking the asymptotic limit of $\Psi^{[\lambda]}_{n}(R, r, \gamma; \theta, \eta, \chi)$ as $R \to \infty$. To obtain an expression for $f^{[\lambda]}_{n'\rightarrow n}(\theta, E)$, we take the $R \to \infty$ limits of the components $F^{[\lambda]}_{Jn'}(R, r, \gamma)$, which are

\[
F^{[\lambda]}_{Jn'}(R, r, \gamma) \to \sum_{\nu\nu'} \sqrt{\frac{k^\nu}{k_n}} \Theta_{Jn'}(\gamma) \psi_\nu(r) e^{ik_\nu R} S^{[\lambda]}_{n'\rightarrow n}(J, E)
\]

where $\Theta_{Jn'}(\gamma)$ and $\psi_\nu(r)$ are the rotational and vibrational wave functions of AC. The aim of the calculation is thus to determine the matrix of coefficients $S^{[\lambda]}_{n'\rightarrow n}(J, E)$, which is called the reactive scattering S matrix. Using this expression, we obtain

\[
f^{[\lambda]}_{n'\rightarrow n}(\theta, E) = \frac{1}{2ik_{\nu ij}} \sum_J F(J)(2J + 1)d^{[\lambda]}_{Jk}(\pi - \theta) S^{[\lambda]}_{n'\rightarrow n}(J, E)
\]

where $d^{[\lambda]}_{Jk}(\pi - \theta)$ is a reduced Wigner rotation matrix [63]. We have included a filter $F(J)$ in Eq. (15), which allows us to calculate separate DCS corresponding to different ranges of $J$. There is a rough correspondence between $J$ and the classical impact parameter $b$, such that low values of $J$ correspond to low-impact (i.e., head on) collisions of the reagents, and high values of $J$ correspond to high-impact (i.e., glancing) collisions.

In addition to the DCS, we also need to consider the state-to-state integral cross-section (ICS), which is a measure of the total amount of scattered AC product in quantum state $n'$, and is given by

\[
\sigma^{[\lambda]}_{n'\rightarrow n}(E) = \frac{2\pi}{2j+1} \int_0^{\pi} |f^{[\lambda]}_{n'\rightarrow n}(\theta, E)|^2 \sin \theta d\theta,
\]
We will also need to consider the reaction probability, which is a measure of the amount of scattered product at a given value \( J \), defined by

\[
P^{[\text{R}]}_{n'-n}(J, E) = |S^{[\text{R}]}_{n'-n}(J, E)|^2
\]

In general, it is difficult to map contributions from different reaction paths onto the DCS. However, Eq. (6) tells us that, in a reaction with a CI, one can easily map the contributions from the \( e \) and \( o \) (Feynman) paths onto the DCS. Since Eq. (6) applies to the entire wave function, we can apply it to the asymptotic limit of the wave function in Eq. (14), and thus to \( S^{[\text{G}]}(E) \) and \( S^{[\text{N}]}(E) \), to obtain

\[
S^{[e]}(E) = \frac{1}{\sqrt{2}} [S^{[\text{N}]}(E) + S^{[\text{G}]}(E)]
\]

\[
S^{[o]}(E) = \frac{1}{\sqrt{2}} [S^{[\text{N}]}(E) - S^{[\text{G}]}(E)]
\]

These equations are all that we need to explain the effect of the GP on scattering cross-sections, such as the DCS and ICS. They allow us to compute separate \( e \) and \( o \) cross-sections using Eq. (15), which show the scattering produced by the \( e \) and \( o \) reaction paths in isolation. They tell us that we can only expect GP effects if \( f^{[e]}_{n'-n}(\theta, E) \) and \( f^{[o]}_{n'-n}(\theta, E) \) overlap.

In the case of the \( \text{H} + \text{H}_2 \) reaction, Eq. (18) specializes to

\[
S^{[1-\text{TS}]}(E) = S^{[e]}(E)
\]

\[
S^{[2-\text{TS}]}(E) = S^{[o]}(E)
\]

Hence, simply by adding and subtracting the computed \( S^{[\text{G}]}(E) \) and \( S^{[\text{N}]}(E) \), we can identify the contributions from the 1-TS and 2-TS reaction paths in the DCS and ICS, and thus explain the effects of the GP on the \( \text{H} + \text{H}_2 \) reaction.

**C. Details of the Calculation: The Use of Vector Potentials**

Before discussing the results of applying Eq. (19), we explain how the GP boundary condition is implemented numerically in the calculations of \( S^{[\text{G}]}(E) \). This is the most technical part of the chapter, and the material here is not needed to understand the sections that follow.

There are three ways of implementing the GP boundary condition. These are (1) to expand the wave function in terms of basis functions that themselves satisfy the GP boundary condition [16]; (2) to use the vector-potential approach of Mead and Truhlar [6,64]; and (3) to convert to an approximately diabatic representation [3, 52, 65, 66], where the effect of the GP is included exactly through the adiabatic–diabatic mixing angle. Of these, (1) is probably the most
elegant method, but it can be implemented efficiently only if the encirclement angle \( \phi \) is a simple function of the coordinates used to represent the Hamiltonian. This is not the case for the coordinates \( (R, r, \gamma) \), which we use in our calculations. Approach (3) is in general numerically the most robust way to include nonadiabatic effects, but it requires that one carry out the calculation on two coupled surfaces, which is clearly inefficient if the system is confined to the lower adiabatic surface. Approach (2) is both numerically robust and uses just the one (lower adiabatic) surface, and is hence the approach we used in the \( \text{H} + \text{H}_2 \) calculations [26].

In the vector potential approach [6], the (real) electronic wave function \( \Phi(\phi) \) is multiplied by a complex phase factor \( f(\phi) \), defined such that

\[
f(\phi + 2\pi)\Phi(\phi + 2\pi) = f(\phi)\Phi(\phi)
\]

A simple choice of phase factor is one that has the form

\[
f(\phi) = e^{i\frac{l}{2}\phi}
\]

where \( l \) must be chosen odd to ensure that Eq. (20) is satisfied. All odd values of \( l \) will correctly incorporate the GP, and give physically equivalent wave functions, which will differ only in an overall phase factor. Similarly, all even values of \( l \) will give the non-GP wave function.

When \( f(\phi) \) takes the form of Eq. (21), the nuclear Laplacian operator is modified according to

\[
-\nabla^2 \rightarrow (-i\nabla - A) \cdot (-i\nabla - A)
\]

where the vector potential \( A \) is given by

\[
A = -\frac{l}{2}\nabla\phi
\]

Hence, the method of Mead and Truhlar [6] yields a single-valued nuclear wave function by adding a vector potential \( A \) to the kinetic energy operator. Different values of odd (or even) \( l \) yield physically equivalent results, since they yield \( \Psi(\phi) \) that are identical to within an integer number of factors of \( \exp(i\phi) \). By analogy with electromagnetic vector potentials, one can say that different odd (or even) \( l \) are related by a gauge transformation [6, 7].

To implement the vector potential in the Jacobi coordinate system \( (R, r, \gamma) \), one proceeds as follows. The Jacobi kinetic energy operator splits into three parts [61]:

\[
\hat{T} = \hat{T}_R + \hat{T}_r + \hat{T}_{\text{ang}}
\]
where each contains a derivative term in just one of the Jacobi coordinates $(R, r, \gamma)$, and hence contains one component of the vector potential,

$$A_a(R, r, \gamma) = -\frac{l}{2} \frac{\partial \phi(R, r, \gamma)}{\partial a}$$  \hspace{1cm} (25)

where $a$ denotes, respectively, $R$, $r$, and $\gamma$. Note that $\phi$ is a function of all three of the coordinates.

The method used to propagate solutions to the Schrödinger equation [61] requires $\hat{T}$ to be represented on a grid of points distributed in $(R, r, \gamma)$, which we will denote using the labels $|klm\rangle$. The first term in $\hat{T}$ is given by

$$\hat{T}_R = -\frac{\hbar^2}{2\mu_R} \frac{\partial^2}{\partial R^2}$$  \hspace{1cm} (26)

(where $\mu_R$ is the reduced mass associated with $R$). Application of Eq. (22) changes the derivative operator according to

$$-\frac{\partial^2}{\partial R^2} \rightarrow \left( -i \frac{\partial}{\partial R} - A_R \right) \left( -i \frac{\partial}{\partial R} - A_R \right)$$

$$\rightarrow -\frac{\partial^2}{\partial R^2} + A_R^2 + i \left( \frac{\partial}{\partial R} A_R + A_R \frac{\partial}{\partial R} \right)$$  \hspace{1cm} (27)

This operator is diagonal in all but the $R$ grid basis functions (denoted $|k\rangle$), and its matrix elements change according to

$$\langle k|\hat{T}_R|k'\rangle \rightarrow \langle k|\hat{T}_R|k'\rangle + \frac{\hbar^2}{2\mu_R} \left\{ \delta_{kk'} A(R_k, r_l, \gamma_m)^2 \right. $$

$$\left. + i \left( \langle k| \frac{\partial}{\partial R} |k'\rangle [A(R_k, r_l, \gamma_m) + A(R_{k'}, r_l, \gamma_m)] \right) \right\}$$  \hspace{1cm} (28)

where $R_k$ denotes the value of $R$ at the $k$th grid point; $r_l$ and $\gamma_m$ are the $r$ and $\gamma$ grid points (see below). Note that this expression was derived by keeping the operator in the symmetric form of Eq. (27), and acting outward with the first derivative operators, on the bra and the ket. This approach (as opposed to taking the second derivative of the ket [20,21]) yields a grid matrix which is exactly Hermitian.

The second term $\hat{T}_r$ has exactly the same form as $\hat{T}_R$ (with $r$ in place of $R$) and produces an exactly analogous change in the matrix elements between the $r$-grid basis functions $|l\rangle$.

The most complicated changes are those produced in the third term $\hat{T}_{\text{ang}}$. This operator can be split into three terms [67]

$$\hat{T}_{\text{ang}} = \hat{T}_{\text{ang}}^{(1)} + \hat{T}_{\text{ang}}^{(2)} + \hat{T}_{\text{ang}}^{(3)}$$  \hspace{1cm} (29)
which are given by

\[
\hat{T}_{\text{ang}}^{(1)} = \frac{\hat{J}^2 - 2\hat{J}_z^2}{2\mu_R R^2} \\
\hat{T}_{\text{ang}}^{(2)} = \left( \frac{1}{2\mu_R R^2} + \frac{1}{2\mu_r r^2} \right) \hat{J}^2 \\
\hat{T}_{\text{ang}}^{(3)} = -\frac{\hat{j} j + j\hat{j}}{2\mu_R R^2}
\]

The term \( \hat{T}_{\text{ang}}^{(1)} \) contains the total angular momentum operators \( \hat{J}^2 \) and \( \hat{J}_z^2 \). These do not operate on the internal degrees of freedom, and are thus not changed by Eq. (22). The term \( \hat{T}_{\text{ang}}^{(2)} \) contains the BC angular momentum operator \( j^2 \), which involves a \( \gamma \)-derivative operator. The change brought about in this operator by Eq. (22) is similar to Eq. (27). The matrix elements of \( \hat{T}_{\text{ang}}^{(2)} \) are diagonal in all but the \( \gamma \) grid basis functions \( |m\rangle \), and change according to

\[
\langle m|\hat{T}_{\text{ang}}^{(2)}|m'\rangle \rightarrow \langle m|\hat{T}_{\text{ang}}^{(2)}|m'\rangle + \hbar^2 \left( \frac{1}{2\mu_R R_k^2} + \frac{1}{2\mu_r r_l^2} \right) \left\{ \delta_{nmr} A_\gamma(R_k, r_l, \gamma_m) \right\}^2 \\
+i \left\{ \left. \frac{\partial}{\partial \gamma} |m'\rangle \right| A_\gamma(R_k, r_l, \gamma_m) + A_\gamma(R_k, r_l, \gamma_{m'}) \right\}
\]

The operator \( \hat{T}_{\text{ang}}^{(3)} \) contains the cross-terms that give rise to the Coriolis coupling that mixes states with different \( \Omega \) (the projection of the total angular momentum quantum number \( J \) onto the intermolecular axis). This term contains first derivative operators in \( \gamma \). On application of Eq. (22), these operators change the matrix elements over \( \hat{T}_{\text{ang}}^{(3)} \) according to

\[
\langle mJ\Omega|\hat{T}_{\text{ang}}^{(3)}|m'J'\Omega'\rangle \rightarrow \langle mJ\Omega|\hat{T}_{\text{ang}}^{(3)}|m'J'\Omega'\rangle + \frac{i\hbar^2}{2\mu_R R_k} \left\{ \delta_{\Omega\Omega'+1} \mathcal{C}_{j\Omega'}^{+} \langle m\Omega|m'\Omega'\rangle A_\gamma(R_k, r_l, \gamma_{m'}) \right\} \\
- \delta_{\Omega\Omega'+1} \mathcal{C}_{j\Omega}^{+} \langle m'\Omega'|m\Omega\rangle A_\gamma(R_k, r_l, \gamma_m) \}
\]

where

\[
\mathcal{C}_{ab}^{\pm} = \sqrt{a(a + 1) - b(b \pm 1)}
\]

To apply the above equations to \( \text{H} + \text{H}_2 \), we need an expression for the vector potential \( A(R, r, \gamma) \), which can be obtained from Eq. (25) once the angle \( \phi \) has
been specified. As mentioned above, we are free to define $\phi$ in any way, provided that $\phi = 0 \to 2\pi$ describes a closed path around the CI. We chose the form,

$$\phi(R, r, \gamma) = \tan^{-1}\left(\frac{d^2R^2 - r^2/d^2}{2Rr \cos \gamma}\right)$$

where $d$ is a dimensionless scaling factor defined in [26]. In the potential cut shown in Fig. 7, this definition of $\phi$ corresponds to the circular polar angle describing an internal revolution about the CI.

D. Solving the Cancellation Puzzle

We are now in a position to discuss the cancellation puzzle in $H + H_2$. We start by considering the state-to-state reaction probabilities $P_{n \to n}^{[3]}(J, E)$, computed according to Eq. (17), with the filter $F(J) = 1$. Representative results, taken from ref. [26], are shown in Fig. 8. These results [26] reproduce those obtained earlier by Kendrick [21], and show that there are noticeable GP effects in some of the state-to-state reaction probabilities, which indicate that a small proportion of the wave function encircles the CI. A curious feature of these

![Figure 8](image_url)

**Figure 8.** State-to-state reaction probabilities, for $H + H_2(100) \to H_2(250) + H$, computed using GP (solid lines) and non-GP (dashed lines) boundary conditions.
results is the apparent alternation in sign of the difference between the GP and non-GP probabilities,
\[
\Delta P_{n' \rightarrow n}(J, E) = P_{n' \rightarrow n}^{[G]}(J, E) - P_{n' \rightarrow n}^{[N]}(J, E)
\]  
(35)
as a function of \(J\).

One would expect that effects of similar magnitude to those shown in Fig. 8 should also appear in the corresponding state-to-state differential and integral cross-sections. However, this is not the case. As already mentioned, there is a considerable amount of cancellation of GP effects in these quantities, which we refer to as the cancellation puzzle. The unexpected cancellations appear in the state-to-state DCS at low impact parameters (i.e., low values of \(J\)), and in the state-to-state ICS (including all impact parameters). We now discuss each of these cancellations in turn.

1. **Low Impact-Parameter Cross-Sections**

The cancellation in GP effects in the state-to-state DCS are found [20–22, 26, 27, 29] at low impact parameters, when \(F(J)\) in Eq. (15) is chosen to include only contributions for which \(J \leq 9\). It is well known [55,56] that most of the reactive scattering in this regime consists of head-on collisions, in which the reaction proceeds mainly by the H atom striking the \(\text{H}_2\) diatom at geometries that are close to linear. Most of the products are then formed by direct recoil in the backward (\(\theta = 180^\circ\)) region, this being typical behavior for a hydrogen-abstraction reaction.

Figure 9 shows the low impact DCS obtained for the same initial and final states (250 \(\leftarrow\) 100), and at the same energy (2.3 eV above the potential minimum),

![Figure 9](image-url)

**Figure 9.** Low-impact parameter DCS at 2.3 eV, for \(H + \text{H}_2(100) \rightarrow \text{H}_2(250) + H\), describing the scattering of (a) \(\Psi_G\) (solid lines) and \(\Psi_N\) (dashed lines), and (b) \(\Psi_e\) (solid lines) and \(\Psi_o\) (dashed lines).
as the reaction probabilities of Fig. 8. These DCS are clearly consistent with the abstraction mechanism just mentioned. Most of the scattering is concentrated in the backward direction, but there is a tiny component in the forward \((\theta = 0)\) direction. The striking feature is that the GP and non-GP DCS are in perfect agreement. The noticeable GP effects in the reaction probabilities of Fig. 8 appear to have canceled completely.

This observation is the first part of the cancellation puzzle [20, 21, 27, 29]. We know from Section III.B that we should be able to solve it directly by applying Eq. (19), which will separate out the contributions to the DCS made by the 1-TS and 2-TS reaction paths. That this is true is shown by Fig. 9(b). It is apparent that the main backward concentration of the scattering comes entirely from the 1-TS paths. This is not a surprise, since, by definition, the direct abstraction mechanism mentioned only involves one TS. What is perhaps surprising is that the small lumps in the forward direction, which might have been mistaken for numerical noise, are in fact the products of the 2-TS paths. Since the 1-TS and 2-TS paths scatter their products into completely different regions of space, there is no interference between the amplitudes \(f_n^{[\text{1-TS}]}(\theta)\) and \(f_n^{[\text{2-TS}]}(\theta)\), and hence no GP effects.

We note that the particular \((2, 5, 0 \rightarrow 1, 0, 0)\) state-to-state DCS that we have chosen happens to have a perfectly clean separation between the 1-TS and 2-TS scattering, which is why the cancellation in GP effects is perfect. Most of the other low impact DCS, however, have a small amount of overlap between the 1-TS and 2-TS scattering, which means that the GP effects almost cancel out in the cross-sections, but that tiny, genuine GP effects remains. Examples of such cross-sections are given in [29].

The fact that the 1-TS and 2-TS paths scatter mainly in opposite directions is also the reason for the alternation in sign of \(\Delta P_{n' \rightarrow n}(J, E)\) pointed out above. It is important to realize that this alternation is not exact. For example, in Fig. 8, we see that \(\Delta P_{n' \rightarrow n}(J, E)\) has the same sign for \(J = 13\) and 14, and that, at 2.3 eV, the alternation is broken by \(\Delta P_{n' \rightarrow n}(2, E) = 0\).

We can explain the approximate alternation of \(\Delta P_{n' \rightarrow n}(J, E) = 0\) by substituting Eq. (19) into Eq. (35), to obtain

\[
\Delta P_{n' \rightarrow n}(J, E) = -2F(J)^2 \text{Re} \left[ S_{n' \rightarrow n}^{(1-TS)}(J, E) \ast S_{n' \rightarrow n}^{(2-TS)}(J, E) \right]
\]

We then invert Eq. (15) by integrating over \(\theta\), which yields the following expression for the S-matrix elements,

\[
S_{n' \rightarrow n}^{[\lambda]}(J, E) = \frac{i k_{ij}}{F(J)} \int_{0}^{\pi} f_{n' \rightarrow n}^{[\lambda]}(\theta, E) d_{j}^{l}(\pi - \theta) \sin \theta d\theta
\]
Equations (36) and (37), together with the property \( d'_{k'0}(\pi - \theta) = (-1)^J d'_{k'0}(\theta) \) [63], show that if the scattering amplitudes satisfied

\[
f'_{n'\rightarrow n}^{[1-\text{TS}]}(\theta, E) \propto f'_{n'\rightarrow n}^{[2-\text{TS}]}(\pi - \theta, E)
\]  

(38)

then the sign of \( \Delta P_{n'\rightarrow n}(J, E) \) would follow \((-1)^J\) exactly. In the DCS of Fig. 9, there is a major component in the 1-TS and 2-TS amplitudes that satisfies Eq. (38), as well as a minor component that does not. Hence, overall the sign of \( \Delta P_{n'\rightarrow n}(J, E) \) displays an approximate alternation with \( J \).

2. Full Cross-Sections

The second part of the cancellation puzzle concerns the full state-to-state DCS and ICS (i.e., including all the impact parameters). In this case, the GP effects do not cancel in the DCS [26, 27, 29], as is shown in Fig. 10. Instead, they shift the phase of the fine oscillations that are superimposed on the main DCS envelope. Following the above, this indicates that the 1-TS and 2-TS paths scatter into overlapping regions of space, so that the GP produces an effect by changing the sign of the interference between \( f'_{n'\rightarrow n}^{[1-\text{TS}]}(\theta) \) and \( f'_{n'\rightarrow n}^{[2-\text{TS}]}(\theta) \). This is confirmed by Fig. 10b, which shows that the 1-TS and 2-TS DCS do indeed overlap.

Figure 10. Full DCS (i.e., including all impact parameters), for \( \text{H} + \text{H}_2(100) \rightarrow \text{H}_2(250) + \text{H} \), describing the scattering of (a) \( \Psi_G \) (solid lines) and \( \Psi_N \) (dashed lines), and (b) \( \Psi_e \) (solid lines) and \( \Psi_o \) (dashed lines). (c) The phases of the corresponding \( e \) and \( o \) scattering amplitudes.
The surprising result is that these GP effects cancel out completely when the DCS is integrated over \( \theta \) to yield the ICS [via Eq. (16)]. This cancellation is shown in Fig. 11(a) for the (2, 5, 0 \( \leftarrow \) 1, 0, 0) ICS over a range of energies, and in Fig. 11(b) for all the nonzero state-to-state ICS at \( E = 2.3 \) eV. In general, one expects small differences in a DCS to average out on integrating over \( \theta \), but a complete cancellation of the differences, which is found to hold for all final states and collision energies tested [26], suggests that there is a systematic difference between the 1-TS and 2-TS scattering dynamics that is causing the cancellation.

This last point is the second part of the cancellation puzzle, and is soon explained by plotting the phases \( \Phi_{n' \rightarrow n}^{[e]}(\theta, E) \) and \( \Phi_{n' \rightarrow n}^{[o]}(\theta, E) \) of the scattering amplitudes \( f_{n' \rightarrow n}^{[e]}(\theta, E) \) and \( f_{n' \rightarrow n}^{[o]}(\theta, E) \) (Fig. 10c). It is clear that these phases

---

**Figure 11.** The ICS for \( \text{H} + \text{H}_2 \), computed using GP (solid lines) and non-GP (dashed lines) boundary conditions, for (a) \( \text{H} + \text{H}_2(100) \rightarrow \text{H}_2(250) + \text{H} \) over a range of energies, and (b) \( \text{H} + \text{H}_2(100) \rightarrow \text{H}_2(v, j) + \text{H} \) at 2.3 eV.
depend in opposite senses on $\theta$, and the same trend is observed for most of the other final states considered [29]. As a result the integrand in

$$\sigma_{n'\rightarrow n}^{[G]}(E) - \sigma_{n'\rightarrow n}^{[N]}(E) = -\frac{4\pi}{2j + 1} \text{Re} \int_0^{\pi} f_{n'\rightarrow n}^{[e]}(\theta, E) f_{n'\rightarrow n}^{[o]}(\theta, E)^* \sin \theta \, d\theta. \quad (39)$$

is highly oscillatory (with period $\sim 12^\circ$), and thus integrates to a very small value (although not to zero).

From semiclassical scattering theory [68,69], it is known that a negative dependence of $\Phi_{n'\rightarrow n}(\theta, E)$ on $\theta$ indicates scattering into positive deflection angles, and vice versa. The terms nearside and farside are sometimes used to describe these two types of scattering (see Fig. 12). Hence, the reason that GP effects cancel in the state-to-state ICS is that the 1-TS and 2-TS paths scatter in opposite senses (with respect to the center-of-mass). There is thus a mapping between the sense in which the reaction paths loop around the CI (clockwise for 1-TS, counterclockwise for 2-TS), and the sense in which the products scatter into space.

3. The 2-TS Mechanism

To complete the explanation of why GP effects cancel in the ICS, we need to explain why the 2-TS paths scatter into negative deflection angles. (It is well known that the 1-TS paths scatter into positive deflection angles via a direct recoil mechanism [55, 56].) We can explain this by following classical trajectories, which gives us the opportunity to illustrate a further useful consequence of the theory of Section II.

This is that, once we have separated the nuclear wave function into $\Psi_e$ and $\Psi_o$ using Eq. (6), we are free to model the dynamics of each component separately using classical trajectories, secure in the knowledge that we have removed the

![Diagram illustrating the difference between nearside scattering into positive deflection angles $\Theta$, and farside scattering into negative $\Theta$. The arrow (chains) represents the initial approach direction of the reagents in center-of-mass frame; the gray rectangle represents the spread of impact parameters in the initial plane wave. Most of the 1-TS paths scatter into positive $\Theta$, and most of the 2-TS paths into negative $\Theta$.](image-url)
effects of the GP. Hence, although we cannot use classical mechanics to predict the effect of the GP, which is entirely a quantum effect, we can use it to model the dynamics of the $e$ and $o$ paths separately. This allows us to predict the extent to which these paths will overlap, and hence estimate the likely magnitude of the GP effects. Of course, we must make allowances for the other types of quantum effects found in reaction dynamics, such as tunneling, zero-point energy, reactive resonances, and threshold effects. The best way to do this is to make detailed comparisons with quantum scattering data (e.g., the state-to-state product distributions), using the quasiclassical trajectory (QCT) approach [70–72].

Hence, in $H + H_2$ we were able to use QCT to model the dynamics of the 1-TS and 2-TS reaction paths separately [29]. The main feature of the quantum calculations that the QCT calculations must reproduce is the scattering of the 1-TS paths into positive deflection angles, and the 2-TS paths into negative deflection angles (since this is what causes the cancellation of GP effects in the ICS). Figure 13 shows that the scattering of the classical 1-TS and 2-TS paths agrees strikingly in this regard, suggesting that, at the very least the classical trajectories are able to give a good overall explanation of why the 2-TS paths scatter into negative deflection angles.

Figure 14 shows a representative 2-TS trajectory, which demonstrates that the 2-TS paths follow a direct S-bend insertion mechanism. The trajectory passes through the middle of the molecule, and avoids the CI; this forces the products to scatter into negative deflection angles. The 2-TS QCT total reaction

![Figure 13. Density plot of the correlation between the deflection angle, $\Theta$, and the total angular momentum, $J$, for 1-TS (open diamonds) and 2-TS (circles) trajectories at 2.3-eV total energy. Note, only 5000 trajectories are plotted for each type for clarity.](image-url)
Figure 14. Classical trajectories for the $H + H_2(v = 1, j = 0)$ reaction representing a 1-TS (a–d) and a 2-TS reaction path (e–h). Both trajectories lead to $H_2(v' = 2, j' = 5, k' = 0)$ products and the same scattering angle, $\theta = 50^\circ$. (a–c) 1-TS trajectory in Cartesian coordinates. The positions of the atoms ($H_A$, solid circles; $H_B$, open circles; $H_C$, dotted circles) are plotted at constant time intervals of 4.1 fs on top of snapshots of the potential energy surface in a space-fixed frame centered at the reactant $H_BH_C$ molecule. The location of the conical intersection is indicated by crosses ($\times$). (d) 1-TS trajectory in hyperspherical coordinates (cf. Fig. 1) showing the different $H + H_2$ arrangements (open diamonds) at the same time intervals as panels (a–c); the potential energy contours are for a fixed hyperradius of $\rho = 4.0$ a.u. (e–h) As above for the 2-TS trajectory. Note that the 1-TS trajectory is deflected to the nearside (deflection angle $\Theta = +50^\circ$), whereas the 2-TS trajectory proceeds via an insertion mechanism and is deflected to the farside ($\Theta = -50^\circ$).
cross-section (Fig. 15) is roughly one-half of the quantum result, suggesting that the reaction is enhanced by quantum tunnelling. This seems reasonable for such a constrained reaction path, and in [29] we use the product rotational distributions (not shown here) to argue that the insertion is facilitated by tunneling through the side of the lower cone of the CI.

IV. FURTHER ASPECTS OF TOPOLOGY

A. Including Particle-Exchange Symmetry

So far, we have treated the atoms as distinguishable particles, both in the general theory of Section II and in the application to H + H₂ in Section III. Here, we explain how to incorporate the effects of particle exchange symmetry. First, we discuss how the symmetry of the system maps from the physical onto the double space, and then explain what effect the GP has on wave functions of reactions that (like H + H₂) have identical reagents and products.

1. Symmetry in Double Space

A useful property of the double space is that it clarifies the treatment of symmetry [28]. In the single space, the symmetry of Ĥ_G can appear confusing, because it depends on the position of the cut line. One way to avoid this confusion is to consider the symmetry of the total (electronic + nuclear) wave function Ĥₚ, which is of course independent of the position of the cut line [6, 7]. Another way is to map Ĥ_G onto the double space.

In Section II, we explained that Ĥ_N and Ĥ_G are respectively symmetric and antisymmetric under the operator R_{2π} in the double space. More generally, if the

Figure 15. Quantum (lines) and QCT (points) total cross-sections for the H + H₂ 1-TS and 2-TS paths.
molecular symmetry group in the single space is \( G \), then the symmetry group in the corresponding double space is the direct product double group \( G \otimes \mathbb{R} \), where \( \mathbb{R} = \{ E, \hat{R}_{2\pi} \} \) (and \( E \) is the identity operator). The properties of direct product double groups are well known from molecular spectroscopy [73]. A double group \( G \otimes \mathbb{R} \) has exactly twice as many symmetry operations as the corresponding single group \( G \), though not necessarily twice as many classes and irreducible representations (irreps). All irreps in the double group will be either completely symmetric or antisymmetric under \( \hat{R}_{2\pi} \). Clearly, the antisymmetric (symmetric) irreps constitute all the allowed symmetries of the \( \Psi_{G}(\Psi_{N}) \) states.

For example, consider the system shown in Fig. 16, in which the (single-space) molecular symmetry group is \( \{ E, \sigma \} \), where \( \sigma \) is a mirror plane of symmetry running from top to bottom of the figure. The symmetry group in the double space is then \( \{ E, \sigma \} \otimes \{ E, \hat{R}_{2\pi} \} \), which is isomorphic with \( C_{2v} \), with a second plane of symmetry \( \sigma^{*} = \hat{R}_{2\pi} \sigma \). Clearly, there are two \( \Psi_{N} \) irreps (symmetric under \( \hat{R}_{2\pi} \)) and two \( \Psi_{G} \) irreps (antisymmetric under \( \hat{R}_{2\pi} \)). The \( \Psi_{N} \) irreps have the same symmetry under \( \sigma \) and \( \sigma^{*} \); the \( \Psi_{G} \) irreps have opposite symmetries. Of the latter, let us take the irrep that is symmetric under \( \sigma \) and antisymmetric under \( \sigma^{*} \), which describes a reaction in which the reagents are prepared in a state that is symmetric with respect to \( \sigma \). In the single space, this function is antisymmetric under \( \sigma \) when the cut line is placed at \( \phi = 0 \), symmetric when it is placed at \( \phi = -\pi \), and unsymmetric when it is placed at, say, \( \phi = \pi/4 \). Use of the double space removes this ambiguity.

2. **Identical Reagents and Products**

When the reagents and products are identical, then the system enters the encirclement region at several different values of \( \phi \), and must therefore be
treated analogously to a unimolecular system [28] (Section II.D). Figure 17a represents the nuclear wave function of the H + H₂ reaction (in the single space), which was considered earlier by Mead under the assumption that it did not encircle the CI [14]. In Fig. 17a, \( \Psi_G \) (or \( \Psi_N \), depending on which relative phase of \( \Psi_e \) and \( \Psi_o \) is assumed in the diagram) is drawn as though the reagents and products of the reaction were distinguishable (as they would be for say D + H₂). Thus all the Feynman paths that enter the encirclement region start at one unique value of \( \phi \).

To treat the reagents and products as indistinguishable, one must make the total (electronic + nuclear) wave function symmetric under a cyclic exchange of nuclei, which is equivalent to making it symmetric under rotations \( \hat{R}_{2\pi/3} \), \( \hat{R}_{4\pi/3} \), about the threefold axis of symmetry. Mead showed that, because the electronic wave function \( \Psi_e \) is antisymmetric under \( \hat{R}_{2\pi/3} \), then \( \Psi_G \) must be symmetrized according to

\[
\Psi_G^{\text{sym}}(\phi) = \frac{1}{\sqrt{3}}[\Psi_G(\phi) - \Psi_G(\phi - 2\pi/3) + \Psi_G(\phi - 4\pi/3)]
\]

By assuming that the system does not encircle the CI, Mead showed [14] that this equation implies that the GP changes the relative sign of the inelastic and reactive contributions in the scattering amplitude.

It is straightforward to combine Eq. (40) with the arguments of Section II, in order to extend Mead’s result to systems that encircle the CI. One has simply to substitute Eq. (5) into each term of Eq. (40), which yields

\[
\Psi_G^{\text{sym}}(\phi) = \frac{1}{\sqrt{6}}[\Psi_e(\phi) - \Psi_o(\phi) - \Psi_e(\phi + 2\pi/3) + \Psi_o(\phi + 2\pi/3) + \Psi_e(\phi + 4\pi/3) - \Psi_o(\phi + 4\pi/3)]
\]
We can represent this function in the single space, provided we use a *common cut line* for all three components. This is shown schematically in Fig. 17. Use of the common cut line is equivalent to taking the linear combinations in the double space, then cutting a $2\pi$-wide section out of the entire $\Psi_{G}^{\text{sym}}(\phi)$. The winding numbers $n$ of the Feynman paths that enter the three equivalent reagent channels must all be defined with respect to the common cut line, since they are analogous to paths starting at different points in the initial state of a unimolecular reaction (Section II.D).

We can now extend Mead’s argument in order to find out the relative sign of the inelastic and reactive contributions to an encircling nuclear wave function. The symmetrized wave function of Eq. (41) can be represented graphically by combining the three functions of Figs. 17a–c, using the convention that the dashed lines have the opposite sign to the solid lines. Four types of Feynman path contribute to $\Psi_{G}^{\text{sym}}$ in a given reagent–product channel. We will call these the direct inelastic, looping inelastic, direct reactive, and looping reactive. To work out whether a contribution (in a given reagent–product channel) is direct or looping, one should identify the shortest route back to the point at which the path entered the encirclement region. If this route passes by one or more exit channels then the path is looping; if it does not, then the path is direct. It is then clear that the GP changes the sign of the direct reactive with respect to the direct inelastic contribution, and that it leaves unchanged the sign of the looping reactive with respect to the direct inelastic. The first of these observations is Mead’s result [14]; the second is the required generalization of Mead’s result to an encircling nuclear wave function. Although we have considered here a reaction that has the same particle-exchange symmetry as the $\text{H} + \text{H}_2$ reaction, the arguments above can clearly be generalized to treat reactions of any symmetry.

**B. Complete Unwinding of the Nuclear Wave Function**

One can think of the mapping of the nuclear wave function onto the double space in Eq. (5) as a partial unwinding of the wave function [28]. This amount of unwinding is sufficient to explain completely the effect of the GP. However, it is interesting to consider unwinding the wave function further. If the range of $n$ that contribute significantly to $\Psi_{G}$ is finite, then Eq. (8) implies that, in principle, one can unwind $\Psi_{G}$ completely, separating the contributions from individual values of $n$.

First, let us consider a system in which $n$ is restricted to $n = 0$ and $n = -1$. The Kernels $K_n$ are therefore negligibly small for $n < -1$ or $n > 0$. In such a system, $\Psi_e$ contains only the $n = 0$ paths, and $\Psi_o$ only the $n = -1$ paths. Hence, mapping onto the double space, to generate $\Psi_e$ using Eq. (6), is sufficient to unwind completely the nuclear wave function (Figs. 18a and b). In the double space (Fig. 18b) the $n = 0$ and $n = -1$ paths are the branches of $\Psi_e$ accessed by
rotating clockwise and counterclockwise from the entry point at \( \phi = 0 \). There is a gap between these two branches, such that \( \Psi_e \) in the double space is like a nonencircling \( \Psi_g \) in the single space (Fig. 4).

When higher \( n \) Feynman paths contribute to the wave function, one has simply to apply repeatedly the single- to double-space mapping, until the nuclear wave function is completely unwound (in the sense just defined). Thus, if the wave function contains only \( n = -2, -1, 0, 1 \) paths, then we need to compute a function \( \Psi'_e \) in the double space that satisfies the boundary condition \( \Psi'_e(\phi) = -\Psi'_e(\phi + 4\pi) \). Adding this function to \( \Psi_e \) [which satisfies \( \Psi_e(\phi) = \Psi_e(\phi + 4\pi) \)] then gives a new function, \( \Psi_4(\phi) \), which occupies the quadruple space \( \phi = 0 \rightarrow 8\pi \) (see Fig. 18c). This new quadruple-space wave function will be completely unwound, such that there is a gap between its clockwise and counterclockwise branches. The \( n = -2, -1, 0, 1 \) contributions will lie in the \(-4\pi \rightarrow -2\pi, -2\pi \rightarrow 0, 0 \rightarrow 2\pi, \) and \( 2\pi \rightarrow 4\pi \) sectors, respectively. If desired, we can convert \( \Psi_4(\phi) \) back to \( \Psi_G \) by taking the combinations,

\[
\Psi_G(\phi) = \frac{1}{2} [\Psi_4(\phi) - \Psi_4(\phi + 2\pi) \\
+ \Psi_4(\phi + 4\pi) - \Psi_4(\phi + 6\pi)]
\]

and then cutting a \( 2\pi \)-wide sector out of the quadruple space, to map back onto the single space.

Figure 18. Complete unwinding of an encircling nuclear wave function \( \Psi_G \) by mapping onto higher cover spaces. (a) The function \( \Psi_G \) in the single space; (b) \( \Psi_e \) in the double space; (c) \( \Psi_4 \) in the quadruple space; (d) schematic picture of \( \Psi_h \) in a \( 2h\pi \) cover space. In each case, \( \Psi_G \) will be completely unwound if it contains contributions from Feynman paths belonging to (b) 2, (c) 4, and (d) \( h \) different winding-number classes.
Clearly, the above procedure can be continued (in principle) as many times as required. Thus, if the wave function includes $n = 4 \cdots 3$ paths, we have simply to define the function $\Psi'_n(\phi) = -\Psi'_n(\phi + 8\pi)$, and then map onto the $\phi = 0 \rightarrow 16\pi$ cover space, which will unwind the function completely. In general, if there are $h$ homotopy classes of Feynman paths that contribute to the Kernel, then one can unwind $\Psi_G$ by computing the unsymmetrised wave function $\Psi_h$ in the $0 \rightarrow 2h\pi$ cover space. The symmetry group of the latter will be a direct product of the symmetry group in the single space and the group $\{E, R_{2\pi}, R_{4\pi}, \ldots, R_{2(h-1)\pi}\}$.

This approach is applicable even to a system that supports long-lived scattering resonances that correspond classically to periodic orbits [74] looping around the CI. Clearly, such systems can support paths for which $-\infty \leq n \leq \infty$, meaning that if one wants to compute $\Psi_G$ exactly then it will never be possible to unwind it completely. However, if one wants to compute $\Psi_G$ to within a given accuracy, then the number of homotopic classes $h$ will be finite because the time-dependent wave function $\Psi_G(t)$ decays exponentially from within the encirclement region as a function of $t$. Hence, for a reactive system, there must be a value of $h$, such that mapping onto a $\phi = 0 \rightarrow 2h\pi$ cover space completely unwinds $\Psi_G$, to within a specified accuracy—meaning that there will be a region of $\phi$ in the $2h\pi$ cover space, over which $|\Psi_h(\phi)|^2$ is negligibly small. This gap region will contain contributions from Feynman paths with $|n| > h$, which could themselves be unwound (if higher accuracy were later required) by mapping $\Psi_h(\phi)$ onto a yet higher cover space.

In a numerical calculation, the number of times that one can unwind $\Psi_G$ will be limited by the maximum size of cover space that can be treated computationally. An efficient way to unwind onto an $2h\pi$ cover space will be to compute the $h$ single-space wave functions that satisfy the boundary conditions

$$\Psi_n(\phi + 2m\pi) = e^{2nm\pi/h} \Psi_n(\phi) \quad (n = 0 \ldots h - 1) \tag{43}$$

The wave function $\Psi_h$ in the $2h\pi$ cover space is then given by

$$\Psi_h(\phi) = \frac{1}{\sqrt{h}} \sum_{n=0}^{h-1} \Psi_n(\phi) \tag{44}$$

To compute each of the $\Psi_n(\phi)$, one can generalize the methods used to compute $\Psi_G$. Hence, the most elegant method would be to use basis functions that satisfy the boundary conditions of Eq. (43), if this were practical to implement. A more general method would be to extend the Mead–Truhlar vector-potential approach [6]. This approach would involve carrying out $h$ calculations, each including a
vector potential of the form

$$A = -\frac{i}{\hbar} n \nabla \phi$$

where \( n = 0 \ldots h - 1 \).

To clarify, the complete unwinding of the wave function is not required to explain the effect of the GP. The latter affects only the sign of the odd \( n \) Feynman paths with respect to the even \( n \) paths, and is thus explained completely once one has unwound these two classes of path by mapping onto the double space. The complete unwinding explains the interference within the even \( n \) and odd \( n \) contributions, by unwinding each of them further, into the contributions from individual values of \( n \).

C. Difference between Bound and Scattering Systems

This chapter has focused on reactive systems, in which the nuclear wave function satisfies scattering boundary conditions, applied at the asymptotic limits of reagent and product channels. It turns out that these boundary conditions are what make it possible to unwind the nuclear wave function from around the CI, and that it is impossible to unwind a bound-state wave function.

To see why this is so, let us attempt to apply the procedure of Section II.B to a bound-state wave function. This is illustrated schematically in Fig. 19. It is clear immediately that we cannot construct an unsymmetric \( \Psi_e \) in the double space, because each bound-state eigenfunction must be an irreducible representation of the double-space symmetry group. Thus a bound-state function in the double space is necessarily symmetric or antisymmetric under \( R_{2\pi} \), and is thus either a \( \Psi_G \) or a \( \Psi_N \) function. For a \( \Psi_G \) function, we have \( \Psi_N = 0 \) (since \( \Psi_G \) and \( \Psi_N \) cannot form a degenerate pair), which implies [from Eq. (6)] that

$$\Psi_e = -\Psi_o = \Psi_G$$

Similarly, for a \( \Psi_N \) function,

$$\Psi_e = \Psi_o = \Psi_N$$

In other words, if we map a bound-state wave function onto the double-cover space using Eq. (6), we simply duplicate the function, because the contribution from the even \( n \) Feynman paths is exactly equal to (or equal and opposite to) the contribution from the odd \( n \) paths.

If we continue mapping onto successively higher cover spaces, following the procedure of Section IV.B, then the effect is the same. Instead of completely unwinding the nuclear wave function, and producing a gap (where \( |\Psi_h(\phi)|^2 \) is
negligibly small), the mapping onto the $2\hbar\pi$ space generates a periodic repetition of the original single-space function. Hence, the technique of Section IV.B, in which one can unwind a reactive wave function completely from around the CI (to within a specified accuracy) does not work for a bound-state function. Heuristically, one may think of this as arising because, in the bound-state system, the Feynman paths (like the classical paths) can describe an infinite number of loops around the CI [28].

An encircling reactive wave function is thus topologically different from an encircling bound-state wave function. This is the reason why, in Section II.A, we said that, when the wave function encircles the CI, it is $|\Psi(\phi)|^2$ that has the form of a torus, rather than $\Psi(\phi)$. A reactive wave function $\Psi(\phi)$ is not a torus—it is essentially a coil, since it can be unwound. A bound-state function $\Psi_G(\phi)$, on the other hand, is a torus (with a twist), because if one imagines calculating it by propagating a function around the CI, then the two ends of the

![Diagram](image)

**Figure 19.** Relation between $\Psi_G$ and $\Psi_N$ for a bound-state system. The functions in the single space (a) can be mapped onto the double space (b) where they have opposite symmetries under $R_{2\pi}$, and belong to different symmetry blocks of the double-space Hamiltonian matrix (c). Unlike reactive wave functions, bound-state functions cannot be unwound from around the CI.
function must match up. In a reactive wave function, no such matching is ever required, and we may loop each end of the function around the CI as many times as we please, before allowing each end to pass out one of the entrance or exit channels (where it is then matched up to the asymptotic scattering functions). We therefore suggest that the term encirclement often used in the GP literature should be qualified as weak encirclement, when the system is reactive, and strong encirclement when it is bound.

Mapping onto the double space will therefore reveal nothing new about the effect of the GP on a bound-state system, since this is purely a boundary-condition effect. However, it does give us an alternative representation of the GP and non-GP wave functions, which may sometimes be clearer than the equivalent single-space representation (in which one deduces the symmetry $\Psi$ from the total wave function $\Phi\Psi$). For example, in the double space, it is very clear that the GP will cause all the bound states to be doubly degenerate when the (single-space) molecular symmetry group is isomorphic with $C_{2v}$ (because the double-space group is then isomorphic with $C_{4v}$). Similarly, the double-space picture is analogous to a double-well system, with periodic boundary conditions, and this may also sometimes be useful in rationalizing the effect of the GP on the spectrum. We note that use of a double space has proved very useful in the analogous field of Möbius molecules, where the electronic wave function satisfies what is in effect a GP boundary condition, on account of a twist in the nuclear structure [75].

Of course, the distinction between reactive- and bound-state wave functions becomes blurred when one considers very long-lived reactive resonances, of the sort considered in Section IV.B, which contain Feynman paths that loop many times around the CI. Such a resonance, which will have a very narrow energy width, will behave almost like a bound-state wave function when mapped onto the double space, since $|\Psi_e|$ will be almost equal to $|\Psi_o|$. The effect of the GP boundary condition would be therefore simply to shift the energies and permitted nodal structures of the resonances, as in a bound-state function. For short-lived resonances, however, $|\Psi_e|$ and $|\Psi_o|$ will differ, since they will describe the different decay dynamics produced by the even and odd $n$ Feynman paths; separating them will therefore reveal how this dynamics is changed by the GP. The same is true for resonances which are long lived, but which are trapped in a region of space that does not encircle the CI, so that the decay dynamics involves just a few Feynman loops around the CI.

V. OUTLOOK AND CONCLUSIONS

The central theme of this chapter is that the effect of the GP on the dynamics of a chemical reaction is very simple, thanks to the topological property of homotopy.
Feynman paths that loop a different number of times around the CI are decoupled from one another in the sum-over-paths, with the result that the nuclear wave function can be split into separate contributions from the even- and odd-looping paths. The sole effect of the GP is to change the relative sign of these two components.

Although the basic physics behind this idea is not new (it was applied to the Aharonov—Bohm effect many years ago [42, 43, 45]), its application to chemical reaction dynamics was made only very recently, in our application to solve the cancellation puzzle in H + H2 [27–29]. This application has been discussed here in some detail, since it illustrates how to use homotopy to explain GP effects in chemical reactions. If one can compute the GP and non-GP nuclear wave functions, then adding and subtracting these functions yields the even- and odd-looping components. If one cannot compute the wave functions, one can still estimate the likely magnitude of GP effects by modeling the dynamics of the even- and odd-looping reaction paths using classical trajectories.

Hence, the GP has a much milder effect on reactive systems than on bound-state systems. This difference has been overlooked in the past, but becomes apparent on noting that an encircling bound-state function contains Feynman paths that loop an infinite number of times around the CI [28]. Consequently, the encirclement of a bound-state wave function is much stronger than that of a reactive wave function; the bound wave function cannot be unwound from around the CI, whereas the reactive wave function can. One consequence of this is that the separation into even- and odd-looping paths yields no information about the dynamics of a bound state system, in which these two contributions are necessarily equal and opposite [28].

The H + H2 example gives clues as to whether GP effects are likely to be important in other reactions. The GP effects cancel in the integral cross-section of H + H2 because the even- and odd-looping paths scatter into different regions of angular phase space. It seems reasonable to assume that many other direct reactions will scatter in a similar manner (since looping different numbers of times around the CI necessarily entails very different types of reaction mechanism), and therefore show similar cancellations. We can also conclude that there will be no GP effects in reactions that involve phase averaging (e.g., statistical capture reactions [76]). For similar reasons, GP effects are very unlikely to affect the outcome of reactions in liquids. The best experiments to see GP effects in reactions are thus likely to be ultrafast control experiments, in which an encircling wave function is probed at short times [77].

Acknowledgments

This research was funded by a grant from the UK Engineering and Physical Sciences Research Council (EPSRC), and by the award of a Royal Society University Research Fellowship to SCA.
References